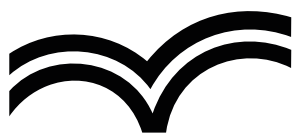


MAIÊUTICA
TECNOLOGIAS DA
INFORMAÇÃO



UNIASSELVI

CENTRO UNIVERSITÁRIO LEONARDO DA VINCI

Rodovia BR 470, Km 71, nº 1.040, Bairro Benedito

89084-405 - INDAIAL/SC

www.uniassevi.com.br

REVISTA MAIÊUTICA

Tecnologias da Informação

UNIASSELVI 2023

Presidente do Grupo UNIASSELVI

Prof. Pedro Jorge Guterres Quintans Graça

Reitor da UNIASSELVI

Prof. Janes Tomelin

Pró-Reitora de Ensino de Graduação Presencial

Prof. Antônio Roberto Rodrigues Abatepaulo

Pró-Reitora de Ensino de Graduação a Distância

Prof.^a Neuzi Schotten

Pró-Reitor Operacional de Graduação a Distância

Prof. Érico Ribeiro

Diretor de Educação Continuada

Prof. Tiago Stachon

Editor da Revista Maiêutica

Prof. Luis Augusto Ebert

Comissão Científica

Prof.^a Airan Arinê Possamai

Prof. Danice Betania Almeida

Prof. Eliomar Russi

Prof.^a Greisse Moser Badalotti

Prof. Nader Ghoddosi

Prof. Pedro Sidnei Zanchett

Prof.^a Cláudia Neli de Souza Zambon

Publicação *On-line*

Propriedade do Centro Universitário Leonardo da Vinci

Apresentação

A Revista Maiêutica dos Cursos Superiores de Tecnologia em Análise e Desenvolvimento de Sistemas, Gestão da Tecnologia da Informação, Licenciatura em Informática, Big Data e Inteligência Analítica, Jogos Digitais, Segurança da Informação, Tecnologia em Agrocomputação, Formação Pedagógica em Informática, Segunda Licenciatura em Informática, Sistemas de Informação, Sistemas para Internet e Engenharia de Software do grupo Uniasselvi, com satisfação, apresenta a você artigos específicos das áreas de Tecnologias de Informação: Aspectos Legais na Informática, Ciência da Computação, Inovação e Dispositivos Móveis, Governança e Gestão da Tecnologia da Informação, Internet das Coisas, Inteligência Artificial, Gamificação, Microserviços, Sistemas de Informação, Tecnologia aplicada na Educação, Tecnologias Emergentes, TICs, Melhoria e Qualidade de Software.

A concepção do conhecimento descrito nestes textos está associada à construção de instrumentos de forma coletiva, propiciando a interação dos atores pedagógicos que integram o nosso processo de ensino-aprendizagem. Estamos passando por uma Era em que os avanços tecnológicos são imensuráveis e possuem uma velocidade inatingível. Cada vez mais, as tecnologias substituem o trabalho humano, deixando os processos industriais mais automatizados, mais rápidos e, a longo prazo, mais baratos.

Tem-se como resultado da construção destes novos saberes a condição de contribuirmos para momentos de reflexão, de análise crítica, de constatações diante de realidades diversas, fatos ou teorias, gerando não somente respostas, mas igualmente novas perguntas que poderão nos conduzir a novas investigações.

Compreender a relevância da produção destes textos acadêmicos significa reconhecer o papel dos profissionais da Tecnologia da Informação, com habilidades e competências em suas áreas de atuação, na condição de agentes de transformação da sociedade em que se encontram inseridos.

Perceber a realidade do mundo, analisar as diversas perspectivas de condução das ações e atividades de uma comunidade, colaborar para a solução de problemas, propor inovações são iniciativas que têm como fonte a organização e a disposição do conhecimento.

Aproprie-se destes conteúdos, saiba mais, construa seu saber.

Seja bem-vindo a um pedaço de nossa história da construção do conhecimento. Convidamos você a conhecê-la. Boa leitura!

Prof^ª. Greisse Moser Badalotti
Coordenadora dos Cursos Superiores de Tecnologia



SUMÁRIO

ENRIQUECIMENTO SEMÂNTICO COM ANÁLISE DE SENTIMENTO NA ETAPA DE ETL: Estudo coletando dados eleitorais

Semantic enrichment with sentiment analysis in the etl stage: study collecting election data

Jonathan Vinicius Suter

Rodrigo Ramos Nogueira 7

ADEQUAÇÃO E IMPLEMENTAÇÃO EM SITES E APLICATIVOS. LGPD LAW IN BRAZIL

Adequacy and Implementation in websites and applications.

Eraldo Pereira de Jesus Júnior

Pedro Sidnei Zanchett 15

O CENÁRIO PÓS-PANDEMIA COVID-19 E A SEGURANÇA DA INFORMAÇÃO Information Security in a Post Pandemic Scenario

Laura Saravia da Silva

Pedro Sidnei Zanchett 27

ANÁLISE COMPARATIVA DE TECNOLOGIAS PARA IMPLEMENTAÇÃO OLAP E OLTP

Comparative analysis of technologies for olap and oltp implementation

Luis Roberto Weck

Rodrigo Ramos Nogueira 33

LEI GERAL DE PROTEÇÃO DE DADOS VERSUS MANIPULAÇÃO DE DADOS General Data Protection Law versus Data Handling

Thays Popper Ciardullo

Nader Ghoddosi 41

INTELIGÊNCIA ARTIFICIAL APLICADA AOS NON-PLAYABLE CHARACTERS (NPCS) EM JOGOS DO GÊNERO ROLE-PLAYING GAME (RPG)

Artificial intelligence applied to non-playable characters (npcs) in role-playing game (rpg) genre

Daniela Maccali

Luiz Felipe Teixeira Bauman

Elzimar José de Carvalho Neto 49

TESTES DE QUALIDADE, ESTRUTURAÇÃO E IMPLEMENTAÇÃO EM UM SISTEMA CORPORATIVO

Quality Tests, Structuring and Implementation in a Corporate System

Guilherme Jackson Lamim

Nader Ghoddosi 59

OOB! (OUT OF THE BOX): DESENVOLVIMENTO DE UM APLICATIVO EDUCACIONAL CONSTRUTIVISTA

Oob! (out of the box): development of a constructive educational application

João Vitor Maia Neves Cordeiro

Rodrigo Ramos Nogueira 67

USO DA TECNOLOGIA E INTELIGÊNCIA ARTIFICIAL NO AUXÍLIO DE APRENDIZADO DAS CRIANÇAS.

Use of Technology and Artificial Intelligence to aid Children's Learning.

Kennedy Anderson Panzenhagen

Pedro Sidnei Zanchett 73

DESENVOLVIMENTO DE UM SISTEMA DE DETECÇÃO AUTOMATIZADA DE FAKE NEWS UTILIZANDO MACHINE LEARNING

Development of an automated fake news detection system using machine learning

Roger Oliveira Monteiro

Rodrigo Ramos Nogueira 89

BANCO DE DADOS PARA ORGANIZAÇÕES

Database for organizations

Juliana Farias Pires

Cristiano Domingues Rodrigues

Telmo Muniz Pacheco Junior 95

ENRIQUECIMENTO SEMÂNTICO COM ANÁLISE DE SENTIMENTO NA ETAPA DE ETL: Estudo coletando dados eleitorais

Semantic enrichment with sentiment analysis in the etl stage:
study collecting election data

Jonathan Vinicius Suter¹

Rodrigo Ramos Nogueira²

Resumo: As redes sociais cada dia mais causam impacto no cotidiano das pessoas e organizações, neste contexto, o Twitter, no qual um usuário escreve uma expressão com até 280 caracteres e outras pessoas podem ver ou compartilhar novamente essa mesma expressão ou a sua própria. Este artigo apresenta um trabalho que tem como objetivo consumir o grande repositório de dados que é o Twitter, e, a partir dele criar um Data Warehouse no qual é possível analisar os textos e as expressões contidas. Também para enriquecer essa base com a análise de sentimento, além do projeto do banco de dados a proposta inclui um método classificador para os textos, utilizando aprendizado de máquina, que é capaz de prever um sentimento relacionado a um Tweet, seja ele positivo, neutro ou negativo. E assim foram qualificados os demais textos, 107393, totalizando os 108693 tweets, e aplicados conceitos de machine learning e data warehousing.

Palavras-chave: Educação. Aprendizado. Explicações Exploráveis. Apache Cordova.

Abstract: Social networks increasingly impact the daily lives of people and organizations, in this context, Twitter, in which a user writes an expression of up to 280 characters and other people can see or share that same expression or their own again. This article presents a work that aims to consume the large data repository that is Twitter, and from it create a Data Warehouse in which it is possible to analyze the texts and expressions contained. Also, to enrich this base with sentiment analysis, in addition to the database design, the proposal includes a classifier method for texts, using machine learning, which can predict a sentiment related to a Tweet, be it positive, neutral or negative. And so, the other texts were qualified, 107393, totaling 108693 tweets, and machine learning and data warehousing concepts were applied.

Keywords: Education. Learning. Explorable Explanations. Apache Cordova.

Introdução

Dentre diversas aplicações em um corpus linguístico baseado em textos do Twitter, se destacam as pesquisas que exploram a análise de sentimento. O processo de análise de sentimentos consiste na abordagem computacional que, com a utilização de técnicas de processamento de linguagem natural e aprendizagem de máquina, tem o objetivo de julgar textos a fim de determinar sentimentos e opiniões presentes em frases. Análise de sentimentos também é comumente conhecida por vários outros termos, tais como: extração de opinião, mineração sentimento, análise de subjetividade, análise afetiva, análise de emoções e mineração de opinião (JUNQUEIRA; FERNANDES, 2017).

1 Acadêmico de Ciência Da Computação (UFSC) – Site: www.uniasselvi.com.br

2 Docente Uniasselvi. Mestre em Ciências da Computação (UFSCAR), Especialista em Desenvolvimento Web e Mobile (Faculdade Sant'Ana), Bacharel em Informática (UEPG). Balneário Camboriú, SC. rodrigo.nogueira@uniasselvi.com.br.

As redes sociais têm grande importância para a sociedade, pois está relacionada ao fato de que possuem grande potencial de compartilhamento de informação. Sendo assim, os dados extraídos de uma rede social, podem ser utilizados para o auxílio na tomada de decisão de determinado assunto de cunho estratégico, para uma corporação ou até mesmo um indivíduo (TOMAÉL et al., 2005).

No entanto, por mais interessante que seja a aplicação de aprendizado de máquina para extração de sentimento, o grande desafio no emprego de técnicas de aprendizado de máquina é que 80% de todo o esforço computacional é gasto na etapa de pré-processamento de dados (LOSARWAR; JOSHI, 2012). O desenvolvimento de uma ferramenta que faça a coleta dos dados, realize a limpeza, os normalize e guarde-os em uma estrutura definida, além de diminuir o esforço nesta etapa, ainda facilita a utilização destes dados por terceiros, permitindo ao usuário que se foque em sua atividade principal de análise destes dados.

A partir dessa problemática, essa pesquisa tem como objetivo o desenvolvimento de um Data Warehouse alimentado com dados, em tempo real, da rede social Twitter, sob o qual foram coletados e analisados os textos sobre a eleição de 2018.

Referencial teórico

A análise de sentimentos é uma subárea da inteligência artificial em ascensão tendo diversas aplicações, principalmente no marketing de produtos e política. Sabendo da ampla utilização do Twitter para armazenar dados e expressar sentimentos, este tem sido amplamente empregado como fonte de estudo para diversos trabalhos nesta área. Junqueira e Fernandes (2018), realizaram a coleta de 988.512 textos do Twitter, os rótulos foram inseridos manualmente, posteriormente foram avaliados os métodos de aprendizado de máquina, em que o melhor método foi o SVM com uma acurácia de 95,7%.

Um estudo de Araújo et al (2016) sobre o funcionamento dos métodos de análise de sentimento no contexto das redes sociais. Utilizando duas bases com dados de redes sociais, foi feita comparação do funcionamento entre oito métodos de classificação de sentimentos e quais os resultados da análise. Utilizando a mineração de dados no Twitter, Corrêa (2017) fez a extração e análise dos sentimentos dos filmes indicados ao Oscar de 2017 utilizando o algoritmo de classificação naive Bayes, baseado no teorema de Thomas Bayes, classificando os Tweets em relação ao seu conteúdo como positivo, negativo e/ou neutro. Efetuando a extração de Tweets relacionados a três categorias de notícias, Nascimento et al. (2012) efetuaram a análise de sentimento, se o sentimento em relação a elas era positivo ou negativo. Utilizando classificadores, com aprendizado supervisionado, mediu-se qual era mais eficaz para este tipo de tarefa, para textos em português, especificamente.

Campanha, Lochter e Almeida (2014) identificaram a necessidade de qualificadores de textos mais eficazes para auxiliar na medição da polaridade destes, dada a grande quantidade de abreviações, gírias e símbolos utilizados nas redes sociais. Para isto, utilizou-se dicionários semânticos e ontologias para auxiliar na elaboração de um comitê de classificadores que detectam automaticamente os métodos de classificação de linguagem natural mais eficazes nessa tarefa. Por sua vez, Moraes et al. (2015) coletaram Tweets em português que foram postados durante a partida entre Brasil e Alemanha na Copa do Mundo FIFA de 2014 para identificar a polaridade. A classificação dos Tweets foi feita a mão e após, foram mostrados os resultados, a quantidade de Tweets positivos, negativos e/ou neutros.

Aguiar et al. (2018), mediram a capacidade de um comitê de algoritmos de aprendizado de máquina para análise de sentimento em redes sociais com a língua portuguesa, usando como estudo de caso a rede social Twitter. Concluiu-se que em alguns casos, os outros algoritmos e o

comitê obtiveram desempenho equivalente na mesma tarefa. Tavares et al. (2017) apresentam uma solução de Business Intelligence para facilitar a extração de informações da rede social Twitter por organizações, efetuando a etapa de ETL, extraindo informações por meio do reconhecimento de entidades nomeadas, a descoberta de conhecimento em texto, inserindo os dados em uma nova base e permitindo a análise gráfica dos dados.

Para minerar dados da rede social Twitter, Trevisan (2015) desenvolveu uma ferramenta para recuperação inteligente de dados para que pudesse efetuar a sumarização e posterior análise dos dados para que se possa extrair informações a partir desta base de dados, a respeito de algum evento no mundo real.

Metodologia

Uma pesquisa bibliográfica tem vários fins, tanto para aprimorar o conhecimento a respeito do assunto abordado e as tecnologias relacionadas, como também para auxiliar na definição do escopo do que será desenvolvido. Por isso a primeira etapa desta pesquisa foi dedicada ao levantamento bibliográfico para se obter a fundamentação teórica sobre o que é Data Warehouse, o que é a análise de sentimento, os métodos de classificação por aprendizado de máquina, bem como os trabalhos já desenvolvidos na mesma linha de estudo (estado da arte).

Esta pesquisa também se enquadra como pesquisa tecnológica de acordo com Freitas Júnior et al. (2014), pois o produto final é conjunto de arquitetura, software, complementado de um conjunto de dados. Para o desenvolvimento desta etapa foi realizada a extração dos dados, mediante ao emprego de um Web Crawler, que busca os Tweets por meio da API disponibilizada pelo próprio Twitter. Então, os Tweets são classificados de acordo com seu conteúdo, se eles têm conteúdo positivo, neutro ou negativo relativo ao assunto.

A condução do desenvolvimento foi realizada tendo como base a arquitetura criada com base na arquitetura de um Data Warehouse de Kimball e Caserta (2011). A Figura 1 mostra a arquitetura proposta por esta aplicação de Data Warehouse. Inicialmente efetuada a coleta dos textos assim como o pré-processamento, compondo a etapa de ETL. Finalmente, após os dados pré-processados e limpos podem ser realizadas consultas OLAP para explorar o cubo de dados. As etapas da arquitetura são descritas em maior nível de detalhamento na sequência.

Figura 1 – Arquitetura utilizada para coleta e Data Warehousing



Fonte: os autores.

Acoplado à etapa de extração da ETL, a coleta é feita por um Web Crawler, desenvolvido utilizando a linguagem de programação Python, na versão 3.6. As requisições ocorrem por meio do uso da biblioteca “TwitterSearch 1.0.210”. Uma vez optando-se por coletar textos em língua portuguesa, sobre as eleições de 2018.

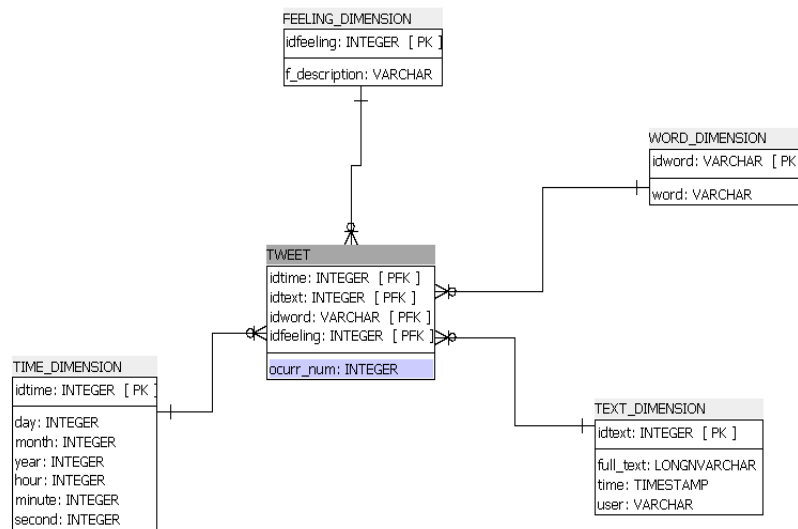
A etapa da limpeza de dados é essencial para o armazenamento de textos, pois é nela que são removidos os dados desnecessários, que além de ocupar espaço em disco, podem atrapalhar o desempenho dos métodos computacionais que utilizam os dados armazenados. Na etapa de limpeza desenvolvida durante a arquitetura Data Warehouse deste projeto foram considerados os seguintes fatores que foram removidos dos textos coletados:

- a) Existência de imagens, bitmaps, gifs etc. no meio dos textos, sendo necessária a retirada deles para que possam ser inseridos na base.
- b) Retweets: devido às limitações que a API do Twitter impõe, não há como desconsiderá-los entre as requisições, diminuindo a variação dos textos e criando a necessidade de tratar os textos com essa marcação.
- c) Links no meio dos Tweets. Exemplo:
- d) Sequência de caracteres que são reconhecidos como de “escape” pelo compilador ou que prejudicam a construção do SQL para inserção do texto na base (sequências do tipo “\n” e aspas simples no meio da cadeia de texto)
- e) Espaços vazios em “excesso”. Exemplo:
- f) Remoção de Stop-words.

Com os textos já limpos, seleciona-se a data do registro e é efetuada sua formatação para que possa ser inserida na base. A partir disso, os dados do Tweet estão preparados para que seja “quebrado” e se efetue a Bag of Words. Com os dados do Tweet, as palavras são quebradas pelo script e inseridas na base de dados multidimensional. Caso a palavra já exista na base, é apenas atualizada sua frequência. E assim, tem-se um documento com os termos e sua frequência em cada Tweet e com uma consulta, sua frequência na base como um todo.

O banco de dados multidimensional armazena os textos dos Tweets que foram padronizados e limpos. O modelo multidimensional descrito na Figura 2 representa os dados armazenados neste artigo. No qual a tabela fato são os textos curtos (tweets) que são analisados por suas dimensões (sentimento, tempo, palavra).

Figura 2 – Modelo Multidimensional para armazenamento de tweets



Fonte: os autores.

Foram coletados 108893 Tweets entre os meses de julho e outubro, referentes à hashtag “eleições 2018”. Após as etapas de coleta, preparação dos textos e enriquecimento semântico e, ao efetuar o treinamento do algoritmo de classificação, usando o conjunto de dados para treinamento com 1300 tweets classificados manualmente.

Resultados e discussão

Esta seção tem como objetivo ilustrar alguns dos resultados obtidos que podem ser explorados a partir do modelo multidimensional desenvolvido. Para a execução das consultas, foi gerada uma base com dados coletados entre setembro e outubro de 2018 utilizando como termo de busca “eleições 2018”. Vale ressaltar que a API não confere acesso total à base de dados da rede social e há limite de coleta por dia. A começar pela palavra de maior menção, de acordo com os parâmetros e período utilizados.

Então, após o teste, foram qualificados os demais tweets da base e assim, explorando as dimensões do Data Warehouse, pode-se obter os resultados de palavras e ocorrências mostradas pelo Quadro 1.

Tabela 1 – Ocorrência de termos

| Palavra | Quantidade |
|----------------|-------------------|
| Eleições 2018 | 51458 |
| Bolsonaro | 24424 |
| Candidato | 10559 |
| Haddad | 9726 |
| Diz | 8184 |
| Presidente | 7188 |
| Contra | 7160 |
| Eleições | 7125 |
| Sobre | 6443 |

Fonte: os autores.

Pode-se observar que naturalmente, o termo usado para a pesquisa dos Tweets é o que tem mais ocorrências, este pode ser desconsiderado no momento, entretanto a segunda palavra mais citada entre os textos é “Bolsonaro”. O segundo termo mais citado é “candidato” e o terceiro é “Haddad”, indicando primariamente que estes foram os candidatos mais citados.

Ao analisar as menções diretas por candidato, os valores foram Bolsonaro: 24424, Haddad: 9726, Ciro: 4510, Alckmin: 1897, Daciolo: 1819, Marina: 1817, Boulos: 1098, Meirelles: 588, Amoêdo: 192, Álvaro: 139, Goulart: 78, Vera: 61, Eymael: 13.

No primeiro momento, é possível observar que nenhum candidato obteve mais citações boas que ruins, refletindo que o sentimento geral entre os tweets foi ruim, e que nenhum candidato conseguiu obter uma grande aprovação dos eleitores. Com base nos textos. O candidato que obteve a maior quantidade de citações com sentimento “bom” foi o Bolsonaro, entretanto também foi o candidato que obteve a maior quantidade de citações classificadas como “ruim”.

Ao comparar os resultados com os da eleição (RESULTADO..., 2018), obtém-se certa equivalência entre os resultados extraídos do Data Warehouse e as intenções de voto, apesar de muitas divergências, há de se considerar ainda que a base possui muitas citações qualificadas como neutras, podendo ocorrer maior distribuição para as citações com sentimento “ruim” e “bom”.

Considerações finais

O tratamento e análise de textos escritos por pessoas, que possuem pouca ou nenhuma revisão, ainda mais em um espaço de informalidade como o Twitter, podem trazer desafios, tanto com os dados em si quanto com o sentido que eles possuem. Assim, a inserção de uma etapa para classificação dos textos como parte da ETL se tornou essencial para automatizar essa tarefa, que pode ser bastante morosa para um humano. Desta forma, o pré-processamento dos textos para que possam entrar na base de dados já limpos e qualificados permite ao usuário se preocupar apenas com o processo analítico dos dados, e desta forma, extrair informações e relatórios, como proposto. Apesar das limitações que a API do Twitter impõe, ainda é possível criar aplicações interessantes, usando os métodos corretos para a estrutura e análise dos dados. As consultas efetuadas para explorar o modelo multidimensional do Data Warehouse são só alguns exemplos do que pode ser feito.

As consultas efetuadas e os dados extraídos, foram capazes de demonstrar bem o sentimento dos eleitores a respeito das eleições como um todo e dos candidatos. Muita indiferença dos eleitores em relação às eleições; grande parte das pessoas que possuíam algum sentimento em relação aos candidatos, levaram para o Twitter o sentimento geral sobre os políticos: desaprovação, seja por ações ou ideologias de cada. O fato é que, a amostra deste estudo e sua análise é coerente até certo ponto com os fatos verificados no mundo real, gerando a necessidade de melhorias na aplicação com um todo.

Referências

AGUIAR, Erickson *et al.* Análise de Sentimento em Redes Sociais para a Língua Portuguesa Utilizando Algoritmos de Classificação. *In: SIMPÓSIO BRASILEIRO DE REDES DE COMPUTADORES E SISTEMAS DISTRIBUÍDOS (SBRC)*, 36., 2018, Campos do Jordão. **Anais [...]**. Porto Alegre: Sociedade Brasileira de Computação, 2018. p. 393-406.

ARAÚJO, Mateus *et al.* **Métodos para análise de sentimentos no Twitter**. Belo Horizonte: Universidade Federal de Minas Gerais, 2016.

CAMPANHA, Jhony M.; LOCHTER, Johannes V.; ALMEIDA, Tiago A. Detecção automática de spammers em redes sociais. *In: ENCONTRO NACIONAL DE INTELIGÊNCIA ARTIFICIAL E COMPUTACIONAL*, 11., São Carlos. **Anais** [...]. Sao Carlos: USP, 2014.

CORRÊA, Igor. **Análise de sentimentos expressos na rede social Twitter em relação aos filmes indicados ao Oscar 2017**. 2017. 72 f. Trabalho de conclusão de curso (Bacharelado em Sistemas de Informação) – Universidade Federal de Uberlândia, Uberlândia, 2017.

FREITAS JÚNIOR, Vanderlei *et al.* A pesquisa científica e tecnológica. **Espacios**, Caracas, v. 35, n. 9, p. 1-10, 2014.

JUNQUEIRA, Kássio T. C.; DA RFERNANDES, Anita Maria R. Análise de sentimento em redes sociais no idioma português com base em mensagens do Twitter. *In: COMPUTER ON THE BEACH*, 9., Itajaí. **Anais** [...]. Itajaí: UNIVALI, 2018. p. 681-690.

KIMBALL, Ralph; CASERTA, Joe. **The data warehouse ETL toolkit: practical techniques for extracting, cleaning, conforming, and delivering data**. Hoboken: John Wiley & Sons, 2011.

LOSARWAR, V.; JOSHI, D. M. Data preprocessing in web usage mining. *In: INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND EMBEDDED SYSTEMS*, 2012, Singapore. **Proceedings** [...]. Singapore: Planetary Scientific Research Centre, 2012. p. 15-16.

MORAES, Silvia *et al.* 7x1PT: um corpus extraído do Twitter para análise de sentimentos em língua portuguesa. *In: Brazilian SYMPOSIUM IN INFORMATION AND HUMAN LANGUAGE TECHNOLOGY*, 10., 2015, Natal. **Proceedings** [...]. Natal: Sociedade Brasileira de Computação, 2015.

NASCIMENTO, Paula *et al.* Análise de sentimento de tweets com foco em notícias. **Revista Eletrônica de Sistemas de Informação**, Curitiba, v. 14, n. 2, p. 1-14, 2012.

RESULTADO da eleição para presidente no Brasil no 1º turno. **Gazeta do Povo**, Eleições 2018, Curitiba, 2018.

TAVARES, Jonatas *et al.* Soluções de BI 2.0 para Análise de Dados a partir do Twitter®: Eleições 2014. **Revista Tecnologia da Informação e Comunicação: Teoria e Prática**, Florianópolis, v. 1 n. 2, p. 69-79, 2017.

TREVISAN, Allan. Mineração de textos no Twitter. 2015. 59 f. Trabalho de Conclusão de Curso (Bacharelado em Sistemas de Informação) – Universidade Tecnológica Federal do Paraná, Curitiba, 2015. **LEI LGPD NO BRASIL**

ADEQUAÇÃO E IMPLEMENTAÇÃO EM SITES E APLICATIVOS.

LGPD LAW IN BRAZIL

Adequacy and Implementation in websites and applications.

Eraldo Pereira de Jesus Júnior¹

Pedro Sidnei Zanchett

Resumo: Com as redes sociais, aplicativos e sites crescendo cada vez mais na sociedade brasileira e relacionamentos econômicos do Brasil com outros países, em 2018 foi instaurada a lei LGPD, que significa Lei Geral de Proteção de Dados e tem como objetivo principal proteger os dados coletados dos usuários e deixá-los cientes da coleta de informações em sites e aplicativos, portanto, neste trabalho será abordado o assunto da lei, visando apresentar a adequação das empresas, implementação necessária, tecnologias utilizadas e dificuldades encontradas durante ou antes mesmo da adequação. Como o desenvolvimento é realizado em três etapas, a introdução tem como objetivo apresentar a lei, são informados no meio do desenvolvimento alguns casos de vazamento que ocorreram no Brasil e após isso é formulado a adequação necessária, tecnologias que podem ser utilizadas, as dificuldades encontradas e implementações que podem ser realizadas, visando redirecionar para sites e aplicativos. A apresentação de softwares e programas que podem ser utilizados têm o intuito de mostrar as possibilidades de trabalho e ferramentas que os profissionais têm à disposição para contornar e ajustar seus produtos para com os usuários. Informo também alguns métodos e formas que são utilizadas hoje como os plugins nos sites e avisos em aplicativos para atender a lei e dar o consentimento ao usuário que consome e fornece informações.

Palavras-chave: Vazamento; Software; Aplicativos.

Abstract: With social networks, applications and websites growing more and more in Brazilian society and Brazil's economic relationships with other countries, in 2018 the LGPD. users and make them aware of the collection of information on websites and applications, therefore, in this work the subject of the law will be approached, aiming to present the adequacy of the companies, necessary implementation, technologies used and difficulties encountered during or even before the adequacy. As the development is carried out in three stages, the introduction aims to present the law, some cases of leakage that occurred in Brazil are informed in the middle of the development and after that the necessary adequacy is formulated, technologies that can be used, the difficulties encountered and implementations that can be performed, aiming to redirect to websites and applications. The presentation of software and programs that can be used are intended to show the possibilities of work and tools that professionals have at their disposal to work around and adjust their products to users. I also inform some methods and forms that are used today such as plugins on websites and notices in applications to comply with the law and give consent to the user who consumes and provides information.

Keywords: Leak; Software; Apps.

1 Introdução

A LGPD (Lei Geral de Proteção de Dados) tem como objetivo, desde 2018, alterar a legislação brasileira com relação à proteção de dados pessoais dos brasileiros. Foi aplicada no Brasil por meio da Lei nº 13.709/2018 e persiste até os dias atuais.

Essa lei estabelece punições perante o descumprimento em situações de vazamento de informações e outras irregularidades em comparação ao que é exigido, são afetados desde aplicativos de celular de pequenas empresas até grandes sites que coletam informações dos usuários, como CPF, data de nascimento, nome e assim por diante...

¹ Centro Universitário Leonardo Da Vinci – UNIASSELVI – Rodovia BR 470 – Km 71 – nº 1.040 – Bairro Benedito – Caixa Postal 191 – 89130-000 – Indaial/SC Fone (47) 3281-9000 – Fax (47) 3281-9090 – Site: www.uniasselvi.com.br.

Se a Lei foi gerada e aplicada, era necessário uma demanda e origem para que surgisse a criação, como relatado por Cunha *et al.* (2021, p. 40):

A LGPD foi inspirada na lei europeia de proteção de dados, General Data Protection Regulation (GDPR), e foi criada, majoritariamente, em razão de questões econômicas, dado que a LGPD foi criada para acompanhar o modelo europeu que estabeleceu que todos os países que estivessem fora dos padrões internacionais exigidos pelos países da Europa e dos EUA não poderiam manter relações comerciais com eles.

Com isso o interesse além de proteger os usuários seria relacionado às relações internacionais que o Brasil poderia deixar de ter com países europeus e norte-americanos por suas diretrizes caso não fosse criada a lei.

Os desafios das empresas que utilizam sites e aplicativos com foco e objetivo no usuário, é de certa forma aplicar métodos para conseguir com que os usuários aceitem a coleta de informações e tenham consentimento. Este aceite que é coletado do usuário muitas vezes é ignorado, portanto temos que.

A LGPD entende que o consentimento é uma manifestação livre, informada e inequívoca pela qual o titular concorda como tratamento de seus dados pessoais para uma finalidade determinada. No caso das redes sociais por exemplo, o usuário está consentindo com a utilização, a captação de seus dados e informações sempre que está navegando e curtindo as publicações da plataforma, ele também concede o direito de utilizar o seu nome e quaisquer conteúdos, incluindo, imagens, fotos, sons por ele publicados no seu perfil, ou seja, não basta apenas em clicar em “aceito” ou “li e concordo”, é preciso que esse consentimento e sua manifestação de vontade sejam livres. Em relação ao titular dos dados, também será necessária uma importante mudança de comportamento, pois ele terá de passar a ler o conteúdo desses termos, pois o fato é que, de cada dez usuários da rede, menos de um lê os termos de uso para os quais deu o "aceito" (CUNHA *et al.*, 2021, p. 44).

Com isso podemos validar que mesmo que os conteúdos passados nos termos sejam de fácil leitura e captação para que o usuário faça a consumo dessa informação e valide se deseja ou não aceitar, é muito comum que nem sejam verificados pelos usuários.

Contudo este trabalho tem o objetivo fazer uma revisão baseada em informações de sites da web, livros e artigos, elaborado utilizando como metodologia a pesquisa bibliográfica e estudo de casos.

2 Objetivo do TCC

Mostrar ao leitor um pouco da parte teórica que implica a Lei no Brasil e dessa forma instigar o que deve ser realizado para que uma empresa que utiliza sites e aplicativos que coletam informações de usuários não recebam penalidades.

Adequar os sites e aplicativos que utilizam como base os usuários e conseqüentemente os seus dados para aplicações internas e externas na internet, como uma consultoria aplicaria essas métricas para as empresas se enquadrarem de forma geral ao LGPD.

3 Lei aplicada no Brasil

Com o avanço da Lei no Brasil, há a obrigação de defender os brasileiros e seus dados, mais especificamente as pessoas físicas, Mulholland (2020, p. 33) abordou que a lei se aplica a somente as pessoas físicas, de acordo com o art. 5º da lei:

Contudo, a LGPD se destina a proteger apenas dados de pessoas físicas. É esta a conclusão que se chega pela leitura do inciso I do art. 5º, que determina que dado pessoal para os termos dessa Lei será “informação relacionada a pessoa natural identificada ou identificável”. Pessoa natural é pessoa física, de modo que as pessoas jurídicas não se encontram amparadas pelos termos da LGPD.

Com isso podemos validar que as reais vítimas de atendados que ocorrem como vazamento de dados e venda de informações são as pessoas físicas, já as empresas mais conhecidas como pessoas jurídicas não têm a aplicação sobre a LGPD.

Na leitura da Lei, em BRASIL (2018), retirado dos incisos I, II e III do art. 3º, é informado que a lei será aplicada caso tenha relação com o território nacional, caso serviços, bens ou tratamento de dados sejam coletados dentro de território nacional, ou seja, mesmo um site do exterior, caso tenha alguma relação com pessoas físicas dentro de território nacional, deverá ter aplicada as métricas da LGPD em suas aplicações.

Confiança sobre a liderança de dados é o que maioria dos agentes que prestam serviços e aplicações precisam fornecer aos seus usuários, desta forma, Mulholland (2020, p. 11) defende que:

A lei busca, de forma geral, garantir que o titular de dados seja capaz de compreender os procedimentos atinentes ao tratamento de dados, para que ele possa se proteger de eventuais abusos e usos ilegítimos. Ao mesmo tempo, a norma impõe limites aos agentes econômicos que realizam a operação de tratamento de dados, regulando o mercado. Assim, o que uma lei geral de proteção de dados busca é o restabelecimento da confiança entre o titular de dados e os agentes econômicos, de forma a diminuir a assimetria de informação entre ambos, para que a coleta e o uso de dados sejam feitos de maneira transparente, garantindo ao titular o controle sobre os seus dados e criando para os agentes de tratamento um ambiente de segurança e conformidade das suas atividades.

Com essa citação fica ainda mais claro que uma das principais barreiras dos agentes com relação aos usuários é a confiança que deverá ser adquirida para que as informações pessoais desses usuários sejam compartilhadas, e que assim o usuário tenha ciência e esteja confortável para aceitar a coleta de dados.

Resumindo de forma geral a aplicação da lei no Brasil, podemos separar os tópicos dentre o assunto principal, primeiramente se trata de uma regra que se aplica em todo território nacional, após isso temos que é aplicada para pessoas jurídicas beneficiando primeiramente e exclusivamente pessoas físicas, exige o consentimento da pessoa física para coleta de dados, deve mostrar claramente a finalidade e necessidade da coleta dos dados, por fim temos Agentes responsáveis e as penalidades.

A Autoridade Nacional de Proteção de Dados Pessoais (ANPD) é o agente que fica responsável em fazer a fiscalização e penalização caso a lei não seja cumprida. Como um agente de cobrança, já existem outros agentes responsáveis pela certificação de empresas para se adequarem à Lei, assim prestando consultoria para a situação.

A penalidade no Brasil se estende e pode ter vários graus diferentes dependendo do fato que fez certa empresa transgredir a lei, portanto de forma reduzida pode se aplicar as multas para as empresas da seguinte forma, antes de uma sanção ser aplicada a empresa, a ANPD avisa e emite alertas com orientações, caso ignorada a multa é aplicada a 2% do faturamento anual da empresa e cada multa de infração pode ter até no máximo o valor de 50 milhões de reais, dessa forma, dependendo da gravidade também pode ser aplicada a retirada do serviço ou aplicação do ar em situações mais agravantes.

Com os tópicos citados, de certa forma o cidadão ganha uma garantia e segurança com relação aos seus dados, portanto são atribuídos direitos como a solicitação de exclusão de dados, transferência de dados entre empresas, quando é adquirido um serviço com outro fornecedor por exemplo.

4 Caso de uso – Dados vazados

4.1 Vazamento de dados do Banco Inter

Em 2018 um caso ocorreu com um banco conhecido no Brasil, Higa (2018) comentou que o banco Inter, conhecido nacionalmente, teve seus dados vazados por um hacker, que seria uma pessoa que se caracteriza por ataques virtuais, esses dados seriam de correntistas. Logo o banco se pronunciou informando que não houve vazamento de dados, porém, mais tarde, o MPDFT (Ministério Público do Distrito Federal e dos Territórios) confirmou que os arquivos vazados pelo hacker seriam do banco Inter.

Com isso o MPDFT pediu julgamento do banco por se tratar de danos morais coletivos causados aos correntistas, uma multa avaliada em 10 milhões de reais foi aplicada na ação e o valor seria destinado ao Fundo de Defesa de Direitos Difusos (FDD).

4.2 Vazamento de dados Netshoes

Acerca do vazamento ocorrido pela Netshoes, citado por Mindsecblog (2019), temos que a empresa foi multada em 500 mil reais em 2018 após o vazamento de dados de mais de 2 milhões de usuários. Dados como nome, histórico de compras, CPF, data de nascimento e e-mail foram vazados, além da multa aplicada o MPDFT recomendou que a Netshoes entrasse em contato com os correntistas e informasse do ocorrido.

5 Sites da internet

Hoje, na internet, temos vários sites onde podemos ter vários tipos de funções para o usuário que consome informações, temos desde a área de mídias de vídeo, leitura e áudio até sites que buscam trazer serviços que são disponibilizados para os usuários.

Com a diversidade de informações que podem ser passadas pelos sites, muitas vezes é necessário informação do usuário para que consiga aplicar um determinado serviço, ou até mesmo amplificar o método de perfil do usuário da forma mais adequada para o site, com isso os sites buscam se adequar a nova lei da LGPD para que os autores e empresas relacionadas não tenham imprevistos com eventuais multas.

Os sites como um grande meio de comunicação direta e indireta na internet global recebe então essa normatização e as empresas precisam passar por alguns passos para entrar nessa normatização, são esses a adequação ao que é proposto pela lei, visualização das tecnologias que podem ser aplicadas, dificuldades encontradas sobre este assunto e por fim como e por quem a implementação pode ser realizada.

5.1 Adequação

Neste caso, para tratar da adequação dos sites os criadores precisam se enquadrar e atentar em alguns pontos específicos para seu tipo de nicho na internet. Sabendo que os sites hoje capturam informações, a adição dos plugins existem para facilitar que o usuário aceite de certa forma a captura de dados.

Por exemplo os sites de grandes empresas como Netshoes e Centauro, que tem como objetivo o mercado varejista de artigos esportivos, tem em seus sites os seus devidos plugins

para que o usuário aceite que as informações sejam compartilhadas, como mostram as imagens a seguir respectivamente:

Figura 1 - *Cookies* Netshoes.



Fonte: <https://www.netshoes.com.br/>. Acesso em: 27 set. 2022.

Figura 2 - *Cookies* Centauro.



Utilizamos cookies para mostrar ofertas que combinam com você. Entenda como funciona em nossa [política de cookies](#)

Fonte: <https://www.centauro.com.br/>. Acesso em: 27 set. 2022.

Como pode ser observado nas Figuras 1 e 2, ambas as empresas optaram por informar o usuário indiretamente sobre a conformidade da coleta de dados na parte inferior do site, assim repassando com mais segurança que os seus dados serão coletados para um estudo de performance no qual a sua experiência de navegação e de busca serão otimizados e melhorados de forma geral.

Em pesquisa, foi apurado, em Bioni (2021, p. 142), que alguns usuários passaram por uma entrevista. Nesta, apresentou que 17% dos usuários deletam os *cookies* e 60% não deletam, contudo os usuários que fazem a exclusão informaram não saber o que são *cookies*, somente 30% dos usuários que fazem a exclusão realmente sabem que os *cookies* têm relação com questões de segurança de informações.

A palavra *cookies* é muito utilizada, aplicado nesse contexto, os *cookies* são fragmentos de informação que são trocados entre os usuários e computador e que melhoram significativamente a experiência de navegação.

Dessa forma podemos ter *cookies* de dois tipos basicamente, *cookies* primários que tem o objetivo de atender o usuário de forma rasa, buscando coletar informações de utilização e exibição do site, assim quando o usuário acessar o site novamente, terá as informações das sessões anteriores. De modo geral essa otimização na experiência dos usuários ocorre de tal forma que ao fazer o login em um site por exemplo e aceitar os *cookies*, caso seja feita uma pesquisa por um artigo em específico, o site pode apresentar artigos do mesmo nicho pelo tipo de pesquisa, ou seja, conforme as informações trocadas pelos *cookies* o site otimiza as pesquisas para o usuário que está logado como já comentado.

Já os *cookies* de terceiro são informações mais sigilosas dos usuários, as empresas precisam ter mais cuidado ao manipular este tipo de informação para evitar vazamentos e outros casos como já citados anteriormente, geralmente esses dados são históricos de navegação, informações demográficas e pessoais.

5.2 Tecnologias

Atualmente existem várias tecnologias que são capazes de gerar um site para que as empresas, por exemplo empresas internacionais como BBC América e o grande jornal The New York Times utilizam de tecnologias como WordPress para sustentar seus sites até hoje, esse sistema é utilizado pela fácil aplicação e integrações com várias linguagens de programação, por ser de código aberto isso chama a atenção da comunidade de programadores e logo o WordPress tomou conta da WEB.

Segundo Sanches (2022), quase 30% de todos os sites da internet foram criados usando o WordPress? Este percentual faz do WordPress a plataforma de criação de sites mais popular da internet. O que indica que ainda em 2022 o WordPress é muito utilizado no mundo todo.

O WordPress é um sistema de gestão de conteúdo para internet, possui código livre e aberto e é baseado na linguagem PHP com banco de dados em MySQL, portanto se trata de uma tecnologia muito utilizada e de acordo com o próprio contador de downloads, temo que:

Figura 3 – Contador de downloads do site Wordpress.

WordPress 6.0 has been downloaded

91,107,875

times

[← Back to WordPress.org](#)

Fonte: <https://wordpress.org/download/counter/>. Acesso em: 27 set. 2022.

Esse sistema embarcado tem a opção para utilização de plugins pagos e gratuitos para implementação da lei LGPD por exemplo assim aplicando em sites do Brasil que utilizam essa tecnologia. A maior parte dos sites usa o Wordpress, apenas mudando as linguagens que são utilizadas para programar.

Empresas que se especializam em criação de sites podem gerá-los e administrá-los a partir do Wordpress, mudando apenas as linguagens que podem ser trabalhadas, por exemplo as linguagens PHP, XML, HTML, CSS, JavaScript, dentre outras conhecidas no mercado.

Hoje para lojas virtuais temos também a possibilidade de utilização do Magento, que seria outra plataforma visando o e-commerce na internet, uma plataforma *de código aberto* que tem suas vantagens. De acordo com Hermano (2021), a utilização de Magento apresenta vantagens como código aberto, personalização de sua loja, SEO avançado, fácil integração de ferramentas, disponibilidade de recursos, dentre outros.

Porém é uma plataforma que pode ser complexa em sua estrutura, mas como já informado, é de código aberto e a comunidade por trás dessa plataforma também cria plugins para atender a necessidade das empresas com relação ao LGPD.

5.3 Dificuldades encontradas

As dificuldades encontradas pelas empresas que precisam aderir à lei LGPD são muitas, visando que deve se estabelecer uma parametrização de norma diferente do que era comum para se adequar corretamente.

Com isso, muito do que é comentado acerca das dificuldades encontradas são com relação a fazer com que o usuário permita a coleta de informações de forma simples e ágil, porém os usuários, muitas vezes leigos sobre o assunto, deixam de aceitar a coleta de informações, o que antes era coletado com certeza, hoje tem uma queda na apuração de informações para usuários.

Complementado, além do ponto de coleta de dados com relação a site e usuário, as empresas possuem carência em profissionais qualificados para adequação da LGPD em sites por exemplo, esse serviço pode ser prestado por um profissional contratado ou até mesmo por um terceiro, que presta consultoria para tirar dúvidas e adequar a empresa. Outra questão é a falta de capital das empresas para investimento nessa nova área, quando se trata de um novo

processo, as empresas podem ter problemas para saber no que investir e se realmente o que está sendo contratado vai resolver seus problemas com relação à adequação.

De acordo com Gomes Júnior (2022, on-line, grifo nosso), “o mercado percebeu que ter um diferencial na segurança oferecida ao cliente é uma grande vantagem, sobretudo em ramos com alta competição. Hoje existem escritórios especializados para *full service*, ou seja, para realizarem a adaptação total da empresa aos termos da LGPD”. Como citado, a demanda gera busca por solução, o que empresas especializadas encontram e em meio ao desespero muitos sites que precisam de ajuste, as procuram para realizar consultorias.

5.4 Implementação

Com relação a implementação, temos alguns tópicos que podem ser tratados para que um empresário possa realizar a normatização da sua empresa e site. Donda (2020) cita em seu e-book que os passos que podem ser seguidos seriam a criação de um comitê para análise e tomada de decisão dentro da empresa, além disso contratar um DPO qualificado que seria o oficial de proteção de dados, validar o entendimento de ciclo de vida dos dados, aplicar regulamentações de segurança da informação, monitorar o ambiente constantemente, criação de relatórios para gestão e por último, a criação de um plano de emergência em situações graves.

Entretanto não existe uma fórmula básica para se adequar à lei, nos sites que vemos hoje a aplicação da tecnologia de plugins já é feita e consegue suprir a necessidade das empresas de médio e grande porte.

Os DPO dessas empresas têm a função de criar uma página de política de privacidade da empresa, na qual o usuário pode acessar e ler os termos, para após aceitá-los e assim os sites já estão protegidos contra as penalidades que a lei impõe em território nacional. Na visualização do usuário é apenas isso que é feito, porém como já citado são necessárias ações antes, para decisões, durante para criação e gestão e após a implementação para monitoramento dos sites (DONDA, 2020).

As empresas, então, podem contratar os profissionais capacitados para essa função ou contratar apenas os serviços de um profissional apenas ou, em casos de empresas de grande porte, uma empresa de consultoria que apresenta as soluções de forma simples.

6 Aplicativos de smartphone

Os Aplicativos dos smartphones são alvos da adequação ao LGPD, hoje basicamente temos duas grandes fornecedoras de aplicativos em meio mobile, seriam estas a Google, disponibilizando o Google Play e a Apple, disponibilizando a Apple Store.

Portanto os criadores/autores de conteúdo para celular hoje necessitam atender às leis de LGPD dentro de território nacional para que possam ter disponibilizado nas grandes lojas de aplicativos os seus produtos e serviços.

Contudo, temos toda a parte de adequação que um aplicativo pode ter para atender as necessidades da lei LGPD, as tecnologias que podem ser utilizadas em meio Android e IOS. Dificuldade encontrada pelos criadores para que sejam disponibilizados certos tipos de dados e, por fim, a implementação que é realizada.

6.1 Adequação

Para adequação de um aplicativo na sua loja de apps, são levados em consideração alguns tópicos. Seguindo a linha de raciocínio da adequação dos sites, dependendo da plataforma em que um site está utilizando, muda a forma de plugins que podem ser aplicados para que o site se ajuste à lei do LGPD. No caso dos aplicativos, como já mencionado, as lojas de aplicativos das empresas Apple e Google precisaram ter uma proposta mais transparente para com os usuários que utilizam

as plataformas, as quais apresentam novas métricas de transparência, nas quais o provedor do aplicativo pode informar aos usuários sobre a coleta de dados (GOOGLE..., 2021)

Por exemplo temos o aplicativo FGTS (Fundo de Garantia do Tempo de Serviço), esse aplicativo do governo está disponível em todas as lojas de aplicativos para smartphone em território nacional, nas duas já mencionadas temos as seguintes figuras:

Figura 4 - Aplicativo FGTS - Apple Store.

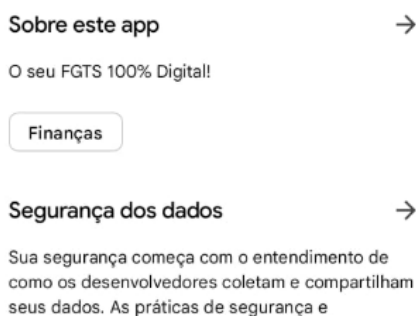


Fonte: <https://apps.apple.com/br/app/fgts/id1038441027>. Acesso em: 28 set. 2022.

Na Figura 4, podemos observar que na própria aba de aplicativos da loja Apple temos os termos de privacidade do aplicativo do FGTS, além disso também temos brevemente os dados que são vinculados ao usuário e os dados que podem ser coletados pelo aplicativo. Esses dados podem alternar de acordo com a função do aplicativo e os serviços que são disponibilizados.

Além de avisar o usuário antes da baixa do aplicativo, os usuários da Apple ainda são questionados durante a usabilidade do app e são avisados, tendo que permitir ou não operações que possam coletar qualquer tipo de dado, um exemplo, ainda com o APP do FGTS é:

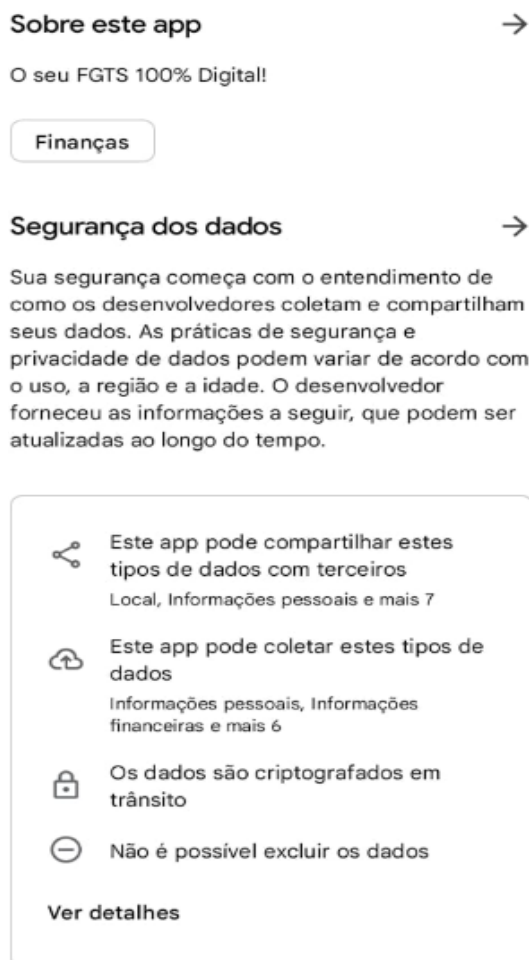
Figura 5 - Permissão do usuário ao entrar no app FGTS.



Fonte: https://play.google.com/store/apps/details?id=br.gov.caixa.fgts.trabalhador&hl=pt_BR&gl=US&pli=1. Acesso em: 28 set. 2022.

No caso dos aplicativos da Apple, podemos validar que mesmo mostrando ao usuário antes de baixar o app, os dados que podem ser coletados, sempre que uma informação for utilizada é necessário a confirmação do usuário, essa seria uma outra ferramenta de adequação. Outro exemplo para adequação, porém ainda com o aplicativo do FGTS é:

Figura 6 - Aplicativo FGTS - Google Play Store.



Fonte: https://play.google.com/store/apps/details?id=br.gov.caixa.fgts.trabalhador&hl=pt_BR&gl=US&pli=1. Acesso em: 28 set. 2022.

Na Figura 6, podemos avaliar que também acontece na loja de aplicativos da Google, o usuário tem uma relação de informações que podem ser coletadas pelo app mencionado e assim a lei de LGPD é tratada nos aplicativos dessa loja.

A adequação se resume em encontrar formas e ferramentas dentro da utilização atual, quando as empresas poderão se ajustar perante a lei, assim reestruturando a forma com que os dados dos usuários eram coletados, com isso temos ainda as tecnologias, dificuldades encontradas e a implementação que pode ser realizada.

6.2 TECNOLOGIAS

Os smartphones são compostos basicamente por hardware, que seria a parte física, a qual pode ser tocada, e o software, onde se encaixa os sistemas operacionais, de forma breve os sistemas operacionais são responsáveis por fazer a gestão de dados dentro do aparelho, essa gestão abrange inclusive outros softwares fornecidos por terceiros e que podem ser chamados de aplicativos.

Com isso as duas maiores lojas de aplicativos do mundo, Apple e Google oferecem aplicativos para sistemas operacionais, dentre todos os sistemas operacionais disponíveis para o mercado hoje, os mais utilizados mundialmente são o Android e em seguida o IOS da Apple.

Em julho de 2018, o sistema operacional móvel do Google figurava com **77,32%** de *market share*, enquanto o da Apple tinha **19,4%**. Já em janeiro de 2022, o primeiro caiu para **69,74%**, com o segundo indo para **25,49%**. A mudança nos números é atribuída à crescente competição por espaço na área (ZAIDAN, 2022, on-line, grifos nossos).

Podemos validar que a briga entre essas tecnologias é grande no mercado internacional e nacional, porém mesmo assim o Android ainda ganha essa disputa entre os brasileiros, assim cita Zaidan (2022, on-line) “Quando se observam estatísticas regionalizadas, a situação não é muito diferente, exceto na América do Norte e na Oceania, onde o sistema da Apple figura em primeiro lugar, com 54%, contra 45% do outro. Na África, por exemplo, 84% dos smartphones usam Android. Na América do Sul o robô vai ainda mais além, com 90% (contra apenas 10% da maçã)”. Com isso pode-se observar a disputa na América do Sul.

Tendo em vista a utilização geral dessas tecnologias, as quais têm por obrigação prestar o suporte para que os criadores de aplicativos mobile, como geralmente os aplicativos são criados dentro dos sistemas operacionais citados, existe uma junção com as lojas de aplicativos para aplicar a normatização, porém o órgão responsável por fazer o levantamento de dados que devem ser coletados, aplicação da lei LGPD são as empresas que estão disponibilizando os aplicativos na loja de aplicativos e as próprias lojas de apps.

Portanto as tecnologias que podem ser utilizadas nos sistemas operacionais, são diversas, linguagens como JAVA, KOTLIN, C++, BASIC podem atender as empresas para criação de aplicativos, porém as conformidades que devem ser cumpridas para normatização do LGPD geralmente já são realizadas pelas empresas como Apple e Play Store para que os aplicativos sejam disponibilizados para download em suas plataformas. Ou seja, é necessário que a empresa entre em consenso e faça um levantamento de requisitos e informe o usuário na página do seu aplicativo antes mesmo que o aplicativo esteja disponível na loja de apps.

6.3 Dificuldades encontradas

Os riscos que a empresa que precisa de adequação sofre são diversos, por esse motivo existem empresas especializadas no ramo de jurisdição, com isso temos citado por Cunha *et al.* (2021, p. 43).

Uma das principais medidas, se não a principal, para implementar a LGPD em empresas e sites é, portanto, ter uma assessoria jurídica especializada para isso, sem está e a interdependência é praticamente improvável que consigam estar em conformidade adequadamente com a Lei Geral de Proteção de Dados. Pode ser que existam outros departamentos que acompanhem a situação e tentem adequar as necessidades, mas trazendo de certa forma grandes riscos para a empresa.

Além das empresas especializadas, com a lei em vigência, as lojas de aplicativos também ficam mais burocráticas para empresas que desejam apresentar o seu produto na banca de aplicativos, agora é necessário informar os requisitos de coleta de dados para que o usuário, antes mesmo do download, possa validar o que o aplicativo está coletando.

Alguns softwares como o IOS, que já foi mencionado, podem possuir mais validações ao utilizar um aplicativo, essas validações, que dependendo do software, não são necessárias em outros sistemas operacionais causam um atraso maior na implementação dentro desse sistema operacional, e por sua vez é mais caro e demorado para o empreendedor ou empresário pois geralmente o investimento para adequação e implementação será maior.

6.4 Implementação

Quando se fala em implementação, pode-se citar sempre um passo a passo com relação a LGPD para os aplicativos mobile, os passos não são necessariamente divididos por tópicos, mas sim um roteiro que pode ser seguido e pode ter variações.

Todo aplicativo precisa de Termos de Uso e Políticas de Privacidade. Bruna Souza, recomenda que todos os dados que foram mapeados e definidos suas necessidades, devem estar explicados e claros nesses documentos. Além disso, salienta que todos os usuários devem saber se seus dados serão enviados para algum e-mail marketing, anúncios personalizados, venda de dados e outros. E ainda deixar que o próprio usuário escolha em qualquer momento a exclusão de seus dados do aplicativo/sistema. Sempre pensando em fornecer mais segurança e transparência aos usuários (ARAÚJO, 2022).

Com o que foi mencionado, o que é necessário como base é o consentimento do usuário, com isso dificilmente as sanções serão aplicadas, porém os passos podem ser dividir em, primeiramente os relacionar e agrupar os dados que serão solicitados aos usuários no seu aplicativo, pode ser desde o nome para um futuro login, como até mesmo a sua localização, como ocorre em aplicativos com intuito de locomoção.

Além disso, também é necessário que os documentos de termos de privacidade sejam explicativos e claros ao usuário, visando buscar todos os tipos de usuário, desde os mais leigos até mesmo os mais astutos.

Por fim, validar realmente a segurança do seu próprio aplicativo para evitar vazamento de dados como já ocorreram em grandes empresas, assim independente dos dados coletados, é garantido ao usuário a proteção de forma geral. De acordo com Farinha (2021, on-line) “Os colaboradores responsáveis pelo tratamento dos dados devem estar treinados e inseridos na cultura de proteção de dados. Uma vez que eles são uma ponta de vulnerabilidade, eles devem ter conhecimentos e boas práticas para evitar que vazamentos ocorram. Isso será um diferencial para o seu aplicativo”. O que indica que o aplicativo é seguro, é informar o usuário sobre a coleta de dados e garantir a proteção desses dados.

Após a implementação da lei LGPD no seu aplicativo, as empresas podem optar por contratação de serviços, que podem ser tanto de uma pessoa física, que seria um profissional DPO, ou também pode ser gerida por uma pessoa jurídica, que se trata de uma outra empresa, porém que faz a gestão dos dados coletados e a parte de implementação.

7 Considerações finais

Este trabalho como um todo, permitiu a melhor compreensão do impacto que a lei geral de proteção de dados pode causar nos usuários de aplicativos e sites. Hoje com a análise realizada, pode-se observar que as tecnologias e ferramentas que são utilizadas para adequação da lei em sites e smartphones são principalmente realizadas por software, então a adequação de um site ou aplicativo acaba sendo finalizada de forma mais rápida por causa dessas tecnologias. Certamente não é concluída somente por software, por isso devemos lembrar dos profissionais que realizam a gestão e aplicam os ajustes para finalizar a implementação da lei, contudo também foram validadas as dificuldades encontradas para implementação da LGPD em sites e aplicativos e as implementações que podem e devem ser realizadas para finalizar o processo de adequação. Por fim, o trabalho tem o intuito de gerar alerta ao usuário, auxílio e compreensão do que se trata a lei e como pode ser adequada e implementada e como a tecnologia pode ajudar e influenciar para a segurança dos usuários e o oposto.

8 Referências

- ARAÚJO, Camila. **LGPD, o que é e como se aplica no seu aplicativo?** Disponível em: <https://www.devmarker.com.br/blog/lgpd-aplicativos>. Acesso em: 05 out. 2022.
- BIONI, Bruno Ricardo. **Proteção de dados pessoais: a função e os limites do consentimento**. 2. ed. Rio de Janeiro: Forense, 2021.
- BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Presidência da República, 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm. Acesso em: 13 set. 2022.
- CUNHA, Blenda Eduarda de M. *et al.* AS DIFICULDADES DA IMPLEMENTAÇÃO DA LGPD NO BRASIL. **Revista Projetos Extensionistas**, Pará de Minas, v. 1, n. 2, p. 39-47, 2021.
- DONDA, Daniel. **Guia prático de implementação da LGPD**. São Paulo: Editora Labrador, 2020.
- FARINHA, Aline. LGPD para aplicativos: lei garante controle de dados de usuários. **Clickweb**, São Paulo, 9 abr. 2021. Disponível em: <https://www.clickweb.com.br/post/80/como-funciona-a-lgpd-para-aplicativos->. Acesso em: 5 out. 2022.
- GOMES JÚNIOR, Francisco. As dificuldades na implementação da LGPD. **ABEINFO**, São Paulo, 22 jun. 2022. Disponível em: <https://abeinfo brasil.com.br/as-dificuldades-na-implementacao-da-lgpd/>. Acesso em: 27 set. 2022.
- GOOGLE vai exigir que apps da Play Store informem como coletam e usam seus dados. **LGPD Brasil**, São Paulo, 19 maio 2021. Disponível em: <https://www.lgpdbrasil.com.br/google-vai-exigir-que-apps-da-play-store-informem-como-coletam-e-usam-seus-dados/>. Acesso em: 26 out. 2022.
- HERMANO, Pedro. Quais são as principais vantagens em utilizar o Magento? **Attri**, [s. l.], 10 mar. 2021. Disponível em: <https://blog.attri.com.br/quais-sao-as-principais-vantagens-em-utilizar-o-magento/>. Acesso em: 27 set. 2022.
- HIGA, P. Banco Inter vazou dados de quase 20 mil clientes, diz investigação do MP. **Tecnoblog**, [s. l.], 31 jul. 2018. Disponível em: <https://bit.ly/2O39mZd> . Acesso em: 20 set. 2022.
- MULHOLLAND, Caitlin (ed.). **A LGPD e o novo marco normativo no Brasil**. Porto Alegre: Arquipélago Editorial, 2020.
- MINDSECBLOG. Netshoes é multada em R\$500 mil por vazamento de informações de 2 milhões de clientes. **Minuto da Segurança**, [s. l.], 14 fev. 2019. Disponível em: <https://minutodaseguranca.blog.br/netshoes-e-multada-em-r500-mil-por-vazamento-de-informacoes-de-2-milhoes-de-clientes/>. Acesso em: 20 set. 2022.
- SANCHES, Ricardo. Ainda vale a pena usar o Wordpress em 2022? **Ricardo Sanches**, [s. l.], 7 mar. 2022. Disponível em: <https://rvsanches.com.br/2022/03/07/ainda-vale-a-pena-usar-o-wordpress-em-2022/>. Acesso em: 21 set. 2022.
- ZAIDAN, João Vítor Sales. Android vem perdendo espaço para o iOS, aponta pesquisa. **Mac Magazine**, Recife, 22 abr. 2022. Disponível em: <https://macmagazine.com.br/post/2022/04/22/android-vem-perdendo-espaco-para-o-ios-aponta-pesquisa/>. Acesso em: 28 set. 2022.

O CENÁRIO PÓS-PANDEMIA COVID-19 E A SEGURANÇA DA INFORMAÇÃO

Information Security in a Post Pandemic Scenario

Laura Saravia da Silva¹

Pedro Sidnei Zanchett¹

Resumo: A pandemia Covid-19 impactou diretamente a modalidade de trabalho conhecida até então, dando espaço ao home office, modalidade de trabalho fora das estruturas físicas das companhias. Em virtude do isolamento social vivido mundialmente, oriundo, naquele momento, de ações visando o controle do contágio do Covid-19, as empresas e por consequência seus trabalhadores, adotaram essa nova perspectiva de trabalho, que é caracterizada por trabalhadores desempenharem suas atividades de trabalho dentro de suas próprias casas. Considerando que as informações são um dos bens mais valiosos dentro das organizações e com o intuito de mantê-las seguras, muitas empresas necessitaram repensar e adequar seus processos e políticas no que tange à segurança da informação para esse novo cenário. Para isso, a segurança cibernética assumiu um papel fundamental dentro das organizações, passando a fazer parte do planejamento estratégico das gestões. Este estudo tem como objetivo trazer reflexões acerca do cenário pós-pandemia Covid-19 na área da tecnologia da informação, bem como os impactos ocasionados na segurança das informações. Assim como, de qual forma as empresas conseguiram adaptar-se a esse novo panorama. Para este estudo foi utilizado o tipo de pesquisa descritiva, baseada em referências bibliográficas disponíveis, como livros, artigos, normas e leis. Os resultados trouxeram ilustrações de como as empresas se basearam em normativas vigentes para se manterem seguras no cenário pós-pandemia, bem como quais os benefícios em se manterem em compliance. Com um olhar para a segurança dos dados, diversas empresas puderam seguir agregando valor ao seu negócio, inclusive em meio ao cenário incerto de pós-pandemia.

Palavras-chave: Informação; Segurança da Informação; Pandemia Covid-19; Políticas de Segurança da Informação

Abstract: The pandemic of Covid-19 has reshaped the way that labor is traditionally known. The working from home era starts, where the job happens away from the physical building. Due to the social isolation experienced worldwide at that specific moment, actions aimed at controlling the contagion of Covid-19, companies, and consequently their workforces, have adopted this new practice of work, which is characterized by workers performing their work activities within their own homes. Considering that information is one of the most valuable assets within organizations and to keep it safe, many companies need to rethink and adapt their processes and policies about information security for this new scenario. For that reason, cybersecurity has become a top priority within organizations, being part of their strategic planning and management. This study aims to provide reflections on the Covid-19 post-pandemic scenario focused on information technology, as well as the effects on information security and how companies have been able to adapt to this new era. For this study, a type of descriptive research was used based on available bibliographic references, such as books, articles, norms, and laws. The results showed how companies relied on current regulations to stay cyber secure in the post-pandemic scenario, as well as the benefits of compliance. By protecting the data, even considering the uncertain post-pandemic scenario, many companies were able to grow their businesses.

Keywords: Information; Security Information; Covid-19 Pandemic; Information Security Polices

¹ Centro Universitário Leonardo da Vinci – UNIASSELVI – Rodovia BR 470 - Km 71 - no 1.040 – Bairro Benedito – Caixa Postal 191 – 89130-000 – Indaial/SC Fone (47) 3281-9000 – Fax (47) 3281-9090 – Site: www.uniasselvi.com.br

Introdução

Atualmente, com todas as tecnologias disponíveis e os recursos de internet ilimitados, temos acesso às informações muito mais fácil do que anos atrás, bem como o compartilhamento dessas informações. Todas essas evoluções contribuíram para que a informação se tornasse um dos ativos mais preciosos, representando um bem estratégico para as organizações.

Em paralelo a isso, o advento da pandemia Covid-19 trouxe impactos gigantescos à rotina de trabalho conhecida até então. De um dia para o outro, inúmeras empresas tiveram que adotar o home office, como modalidade de trabalho. Nesse sentido, entende-se por home office, toda atividade de trabalho profissional desempenhada fora do ambiente empresarial de trabalho, habitualmente estando relativo ao trabalhador exercer as suas atividades profissionais dentro da sua própria casa.

Considerando o valor das informações e as mudanças vividas no período pandêmico, irrefutavelmente as organizações necessitaram rever seus procedimentos internos, suas políticas, bem como seu planejamento estratégico para garantir a segurança das suas informações.

Segundo Coelho, Araújo e Bezerra (2014, p. 2):

A informação pode existir em diversos formatos: impressa, armazenada eletronicamente, falada, transmitida pelo correio convencional de voz ou eletrônico etc. Seja qual for o formato ou meio de armazenamento ou transmissão, recomenda-se que ela seja protegida adequadamente, sendo assim é de responsabilidade da segurança da informação protegê-la de vários tipos de ameaça.

Assim, muitas foram as mudanças vividas nesse período, principalmente no âmbito da segurança das informações.

Referencial teórico

A informação é recurso de valor inestimável. Com a contribuição dela o mundo evoluiu em diversos setores, seja tecnológico, social ou científico. Este ativo é composto por um conjunto de dados, que organizados geram valor. A informação é um recurso que move o mundo, além nos embasar para identificar como o universo está caminhando. (FONTES, 2006).

Desde que nascemos somos expostos a informações e comunicações. O bebê, o qual ainda não sabe falar, comunica-se por meio do choro e de sons ainda não traduzíveis, no entanto já é uma forma de comunicação. Durante o nosso desenvolvimento, vamos evoluindo e criando habilidades em diversos âmbitos, inclusive de comunicação.

De acordo com Fontes (2006, p. 3), “Informação é muito mais que um conjunto de dados. Transformar esses dados em informação é transformar algo de pouco significado em um recurso de valor para a nossa vida pessoal e profissional.”

Pensando no valor da informação e o quanto ela pode impactar o futuro de uma geração, protegê-la se torna essencial. Este bem se não for devidamente registrado, arquivado de forma segura e livre de qualquer alteração, pode ser perdido e com ele todos o valor que ele representa também será extinto. Além disso, se utilizado por pessoas indevidas pode gerar impactos negativos nas vidas das pessoas, grupo de pessoas ou até mesmo para os negócios de determinada empresa.

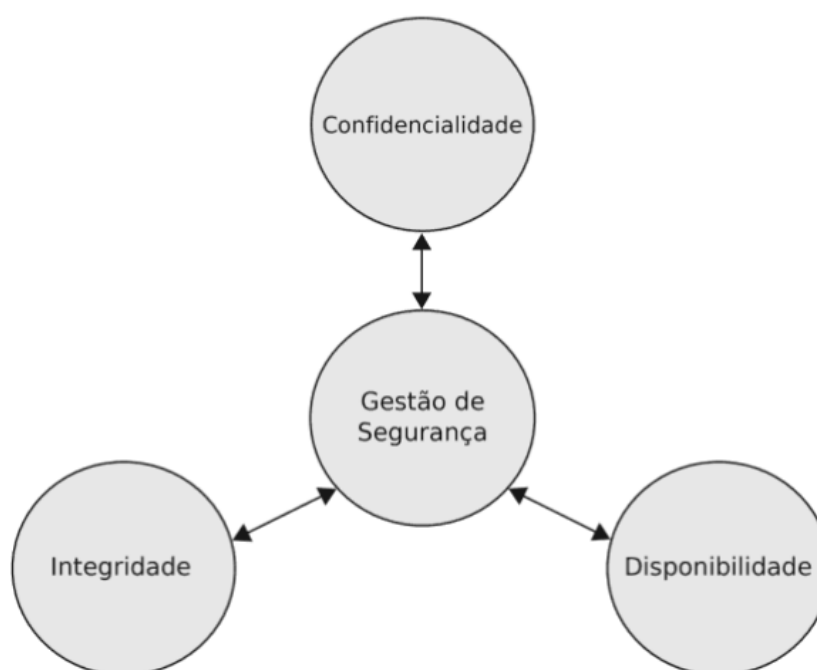
Um exemplo disso, são os registros médicos, esse tipo de informação deve ser confidencial e de acesso somente de pessoas autorizadas. Tais informações trazem o histórico de problemas de saúde do paciente, procedimentos realizados e muitas vezes segredos deles. Esse tipo de informação se disponível a agentes mal-intencionados, podem ser utilizados para impactar a imagem de certa pessoa em detrimento à outra, por exemplo.

Atualmente a Internet – rede mundial de computadores – representa o maior meio de distribuição de informações disponível no mundo. Cada vez mais desenvolvida, com acessibilidade universal, em constante expansão e ao alcance de todas as classes sociais, faz dessa rede protagonista no acesso à informação e aos avanços em benefício de toda sociedade. Em contrapartida, inúmeras vezes é pivô para ações ilegais e inapropriadas com uso das informações.

Segundo Oliveira e Filgueiras (2022, p. 441), “E é por isso que a segurança da informação se tornou tão necessária, onde atualmente existem leis que determinam os direitos e deveres de todos o que estão nessa enorme rede que engloba o mundo todo.”

Com o intuito de proteger as informações, existem os princípios fundamentais da segurança da informação, conhecidos também como a *tríade da segurança da informação*. Essa tríade abrange três principais pilares: confidencialidade, integridade e disponibilidade.

Figura 1: Tríade de Segurança da Informação.



Fonte: Hintzbergen *et al.* (2015, p. 33).

Conforme Hintzbergen *et al.* (2015, p. 33):

Confidencialidade, integridade e disponibilidade são princípios críticos de segurança. Você deve compreender o seu significado, como eles são providos por diferentes mecanismos e como a sua ausência pode afetar negativamente um ambiente. Tudo isso o ajuda a identificar melhor os problemas e fornecer soluções adequadas.

O princípio da confidencialidade tem como objetivo garantir que somente pessoas autorizadas tenham acesso às informações, mantendo assim a sua confidencialidade. No que tange à integridade, esse pilar garante que as informações mantenham a sua exatidão, ou seja, que elas não sofram alteração alguma após a sua criação, seja durante a transferência ou ao recebimento dela. Por fim, o terceiro pilar visa garantir a disponibilidade das informações, isto é, que elas estejam acessíveis e disponíveis sempre que preciso for.

Para ilustrar os conceitos da tríade da Segurança da Informação, podemos imaginar um baú com uma fechadura, lançado ao fundo do mar, onde nele contenham informações sensíveis. Essas informações estarão seguras, pois somente quem tiver a chave desse baú poderá abri-lo, garantindo a confidencialidade das informações. Essas informações se manterão intactas, mantendo a sua integridade, no entanto o princípio da disponibilidade será ferido, pois estando no fundo do mar dificilmente esse baú será encontrado quando houver a necessidade de acessar essas informações. Sendo assim, de nada adianta ter uma informação segura, íntegra, mas indisponível ao seu devido fim.

Em 1995, instituições internacionais ISO – *International Organization for Standardization* e IEC – *International Electrotechnical Commission* originaram um grupo de normas que regem e padronizam diretrizes relacionadas à Segurança das Informações, representadas pelo grupo de normas 27000. Especificamente as normas ISO/IEC 27001 e 27002 estão relacionadas respectivamente aos pré-requisitos para um Sistema de Gestão de Segurança da Informação (SGSI) e aos controles que auxiliam na aplicabilidade do SGSI.

Este grupo de ISO auxilia na implementação de um sistema gerencial seguro, desde a avaliação e mitigação de riscos, ameaças, bem como a conscientização e o contínuo treinamento de colaboradores e parceiros. É visível que as pessoas são o elo mais fraco no âmbito de segurança da informação, de nada adianta investir orçamentos gigantescos em tecnologia e arquiteturas computacionais complexas, se as pessoas não vislumbram o real valor da segurança das informações.

Segundo Hintzbergen *et al.* (2015, p. 40):

Em um mundo ideal, a segurança da informação é parte das operações diárias. Todos os funcionários estão cientes da segurança e reconhecem as falhas de segurança. A segurança é implementada em todos os sistemas e um alto nível de maturidade é alcançado.

A Lei Geral de Proteção de Dados Pessoais – LGPD tornou-se lei no Brasil no ano de 2020. A sua concepção foi inspirada na lei europeia de proteção de dados chamada *General Data Protection Regulation* (GDPR).

A LGPD garante a proteção de dados pessoais, seja quanto à coleta, tratamento e armazenamento desses dados. Essa lei causa certas dúvidas quanto à sua aplicabilidade, considerando que as empresas, ou seja, pessoas jurídicas devem seguir a referida lei, no entanto, a proteção de dados prevista em lei refere-se aos dados de pessoas naturais e não aos dados empresariais.

Assim a proteção dos dados prevista nesta lei, visa resguardar a privacidade e proteção dos dados que as empresas possuem de seus clientes, colaboradores, terceiros, dentre outros, em virtude da natureza do seu negócio.

Metodologia

Este estudo foi baseado em uma pesquisa do tipo descritiva, por meio de pesquisas bibliográficas acerca dos conceitos de Segurança da Informação, bem como o cenário pós-Covid na área de tecnologia da informação.

O presente trabalho foi realizado com base em referências bibliográficas de livros, normas técnicas e artigos científicos. Bem como, a aplicabilidade das normas ISO/IEC 27001 e LGPD.

Os resultados deste estudo serão ilustrados qualitativamente, demonstrando assim os benefícios em aderir políticas de segurança sólidas e confiáveis, baseadas em normas vigentes.

Resultados e discussão

De acordo com as pesquisas realizadas observou-se que a pandemia Covid-19 trouxe consigo mudanças circunstanciais às rotinas de trabalho, impactando diretamente à segurança das informações.

Com o objetivo de manter os negócios das empresas ativos, mesmo em um cenário de isolamento social, proveniente de normativas e orientações de instituições internacionais responsáveis pela saúde pública mundial, inúmeras empresas aderiram à modalidade de trabalho home office. Muitos colaboradores, ao trabalhar de suas casas, acabaram utilizando seus equipamentos particulares para o desempenho das suas atividades profissionais, aumentando assim, a superfície de exposição das organizações. Com isso, enfrentaram o desafio de manter seus ativos seguros, íntegros e disponíveis.

Frente ao panorama apresentado, as organizações necessitaram repensar os seus procedimentos de Segurança da Informação, e passaram a tratar o tema como um assunto estratégico à sobrevivência de seus negócios. Para tal, empresas de sucesso basearam suas ações e planejamentos em normas regulamentadoras como ISO/IEC 27001 e 27002, assim como a Lei Geral de Proteção de Dados – LGPD, por exemplo.

O trabalho fora das estruturas empresariais trouxe muitos desafios às organizações, uma delas foi manter seus dados seguros, assim como, igualmente importante, manter os dados de seus clientes e colaboradores preservados.

Nesse sentido a Lei Geral de Proteção de Dados, LGPD, serviu como base para ações no quesito segurança de dados de pessoas privadas. Esta lei objetiva resguardar os usuários de uso indevido e abusivo de suas informações, bem com o vazamento delas. Segundo Pinheiro (2021):

[...] a LGPD é um novo marco legal brasileiro de grande impacto, tanto para as instituições privadas como para as públicas, por tratar da proteção dos dados pessoais dos indivíduos em qualquer relação que envolva o tratamento de informações classificadas como dados pessoais, por qualquer meio, seja por pessoa natural, seja por pessoa jurídica. É uma regulamentação que traz princípios, direitos e obrigações relacionadas ao uso de um dos ativos mais valiosos da sociedade digital, que são as bases de dados relacionados às pessoas.

O vazamento de dados através da ineficiência dos controles de segurança está suscetível a diversas sanções previstas em lei, que variam desde advertências, multas até a suspensão de atividades relacionadas ao tratamento de dados pessoais. No Capítulo VII da Lei Geral de Proteção de Dados, especificamente no Art. 46 encontra-se o seguinte detalhamento:

Art. 46. Os agentes de tratamento devem adotar medidas de segurança, técnicas e administrativas aptas a proteger os dados pessoais de acessos não autorizados e de situações acidentais ou ilícitas de destruição, perda, alteração, comunicação ou qualquer forma de tratamento inadequado ou ilícito (BRASIL, 2018).

Nesse sentido, com o intuito que manter os dados de seus clientes seguros, considerando todas as particularidades do trabalho home office, empresas aderiram a ferramentas e softwares que lhes auxiliassem a se manterem *em compliance* à lei vigente. Um exemplo disso, são as tecnologias de DLP, *data loss prevention* ou prevenção à fuga da informação. Atualmente, ferramentas de DLP são ofertadas por diversos *players* no mercado e estão presentes nos ambientes corporativos como aliadas à adequação das empresas à LGPD. Com o auxílio dessa tecnologia, políticas e regras são configuradas para barrar o vazamento de informações confidenciais, sejam elas informações de identificação pessoal (PII), informações de saúde protegidas (PHI), dados de cartões de pagamento, bem como informações intelectuais.

Outro ponto importante, foi a preocupação das empresas em manter seus processos relacionados a políticas de segurança da informação atualizados e orientados em normas regulamentadoras. Nessa ótica, o *framework* ISO 27001 popular na segurança das informações e ISO 27002 com práticas detalhadas nortearam as ações das corporações durante à adequação no cenário pandêmico.

Em consequência do modelo de trabalho adotado nesse período, o número de reuniões on-line cresceu consideravelmente, fazendo com que novas posturas fossem adotadas ou até mesmo orientações já divulgadas pelas empresas anteriormente, passassem a ser vividas de fato diariamente pelos seus colaboradores. Exemplo disso, é a política de mesa e tela limpa, o qual preza pela não exposição de informações sensíveis, seja através de notas autoadesivas na tela do computador, bilhetes com usuário e senha, bem como anotações em agendas expostas na bancada de trabalho.

Pensando em potencializar seu poder de proteção das informações, muitas empresas optam por certificarem-se na ISO 27001. Essa certificação traz consigo a garantia e a credibilidade de que empresas certificadas possuem seus processos em conformidade no que diz respeito à gestão da segurança da informação, assim, agregando valor ao negócio.

Em linhas gerais, as empresas que almejam ser certificadas, deverão passar um processo de auditoria. Esta realizada por auditores devidamente credenciados pela certificadora. A implementação da ISO 27001 passa por diversos passos como: estudo do escopo, avaliação de risco, análise de lacunas, aplicabilidade, melhorias de segurança, auditoria interna, implementação, documento final, revisão, melhoria contínua e ações corretivas.

Considerações finais

A importância de manter as informações seguras ficou evidente com todas as mudanças vividas no cenário pós-pandemia Covid-19.

Investir em atualização, certificações, tecnologias, assim como na conscientização e capacitação dos ativos vivos das organizações foram algumas das estratégias utilizadas pelas empresas para manterem-se reconhecidas no competitivo mundo empresarial. Para isso, as normativas e leis guiaram e continuarão embasando o planejamento e as ações dos gestores da área de tecnologia da informação.

Por fim, todos os avanços tecnológicos alcançados nesse período, aliados aos diversos desafios oriundos desse progresso, é um caminho sem volta. Neste novo cenário, o qual vivemos hoje, ainda em decorrência da aceleração tecnológica vivida durante e posteriormente à pandemia, é preciso pensar, mais do que nunca, a segurança da informação como um conjunto de pessoas, processos e tecnologia.

Referências

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018.** Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Presidência da República, 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm. Acesso em: 13 set. 2022.

COELHO, Flávia Estélio Silva; ARAÚJO, Luiz Geraldo Segadas; BEZERRA, Edson Kowask. **Gestão da segurança da informação NBR 27001 e NBR 27002.** Rio de Janeiro: RNP/ESR, 2001.

FONTES, Edison. **Segurança da informação: o usuário faz a diferença.** São Paulo: Saraiva, 2006.

HINTZBERGEN, Jule, HINTZBERGEN, Kees, SMULDERS, André, BAARS, Hans. **Fundamentos da segurança da informação com base na ISO 27001 e na ISO 27002.** Van Haren Publishing, 2010, 2015.

OLIVEIRA, Eliane Vendramini, FILGUEIRAS, Rodrigo. A importância da segurança da informação para as organizações. **Revista Alomorfia**, Presidente Prudente, v. 6, n. 1, p. 438-447, 2022.

ANÁLISE COMPARATIVA DE TECNOLOGIAS PARA IMPLEMENTAÇÃO OLAP E OLTP

COMPARATIVE ANALYSIS OF TECHNOLOGIES FOR OLAP AND OLTP IMPLEMENTATION

Luis Roberto Weck¹

Rodrigo Ramos Nogueira²

Resumo: A escolha de um banco de dados é um dos grandes passos a serem tomados durante a análise e desenvolvimento de um sistema. A necessidade de um banco de dados relacional pode ser dada por meio dos modelos do software, sendo que pode majoritariamente possuir características de um ambiente transacional, quanto de um ambiente analítico. Neste trabalho, será realizada uma comparação de performance entre dois dos mais populares bancos de dados objeto-relacionais do mercado, por meio da utilização do software de benchmarks HammerDB. Foram testados os modelos operacionais OLTP e OLAP dos bancos PostgreSQL e MySQL, operando sobre ambientes containerizados, para reduzir a interferência de processos não relacionados aos SGBDs e aumentar sua reprodutibilidade e confiança nos resultados. Os SGBDs contam com parametrizações que fazem com que o ambiente simulado seja similar a ambientes de produção. O objetivo é identificar qual dos dois SGBDs possui o melhor desempenho em cada ambiente de trabalho.

Palavras-chave: Bancos de dados, Comparação, Desempenho, Container

Abstract: The choice of a database is one of the big steps to be taken during the analysis and development of a system. The need for a relational database can be seen through the software models, which can mostly have characteristics of a transactional environment, as well as of an analytical environment. In this work, a performance comparison will be made between two of the most popular object-relational databases on the market, using the HammerDB benchmarking software. The operational models OLTP and OLAP of PostgreSQL and MySQL databases, operating in containerized environments, were tested to reduce the interference of processes not related to the DBMS and increase their reproducibility and confidence in the results. DBMSs have parameterizations that make the simulated environment like production environments. The objective is to identify which of the two DBMSs has the best performance in each work environment.

Keywords: Databases, Compare, Performance, Container

1. Introdução

Um banco de dados relacional é definido tradicionalmente como um conjunto de dados relacionados entre si. Kanellakis (1990, p. 1076) define um banco de dados relacional como “um conjunto finito de relações (entidades) interpretando um conjunto de símbolos de relação”. Por isso, é comum para desenvolvedores pensarem em bancos de dados apenas como um depósito de informações, no entanto um banco de dados é muito mais do que isso. Atualmente, os bancos de dados mais populares permitem realizar buscas textuais (utilizando frases complexas como critério de busca), cálculos de distância entre coordenadas, operações com polígonos (sobreposição, contenção, intersecção), além de agregações complexas (LU; LIU, 2000, p. 6).

Ao utilizar toda arquitetura proporcionada pelos sistemas gerenciadores de bancos de dados relacionais, é possível explorar diversos recursos como processamento de textos, arma-

1 Bacharel em Engenharia de Software – Universidade do Estado de Santa Catarina (UDESC). www.uniasselvi.com.br

2 Docente UNIASSELVI. Mestre em Ciências da Computação (UFSCAR), Especialista em Desenvolvimento Web e Mobile (Faculdade Sant’ana), Bacharel em Informática (UEPG). Balneário Camboriú, SC. rodrigo.nogueira@uniasselvi.com.br.

zenamento e processamento de dados não estruturados como JSON, tipos de dados de localização (latitudes e longitudes), formas geométricas, dentre outros. Ao utilizar esse mapa mental, um erro muito comum é não utilizar o poder da expressividade da linguagem SQL (*Structured Query Language*), em favor de adaptar as solicitações de dados e informações pela aplicação, utilizando modelos procedurais.

A motivação deste trabalho se dá a partir da necessidade de escolher o melhor banco de dados para uma ou mais tarefas de aplicações atuais. Serão realizadas operações OLTP (*Online Transaction Processing*) e OLAP (*Online Analytical Processing*). Dentre os diversos aprendizados adquiridos ao desenvolver este trabalho, destacam-se as horas de pesquisa sobre ambientes containerizados, que somam mais de cinquenta horas, além do período de testes das consultas OLAP e OLTP, que somam mais de 100 horas. Parte destas horas se devem ao fato de os testes só poderem ser realizados de um a um, por falta de disponibilidade de um computador exclusivo para testes. Além disto, deve-se registrar as dificuldades em encontrar ferramentas de benchmark que entregam resultados comparáveis entre diferentes SGBDs.

Com toda esta pesquisa, é possível observar que benchmarks são ainda o modelo preferido de comparação de bancos de dados, e existe muito conteúdo na internet a respeito, mas não existe um grande foco acadêmico na área, e sim um comercial. Empresas continuam disputando suas soluções para venda de seus produtos, e o TPC (*Transaction Processing Performance Council*) proporciona o palco para a apresentação destes produtos.

Para atingir o objetivo geral proposto nesse trabalho, apresentam-se os seguintes objetivos específicos: verificar o desempenho de dois dos mais populares SGBDs do mercado; obter as configurações de benchmarking para o PostgreSQL e MySQL; preparar um ambiente de testes que permita igualdade nas duas ferramentas; comparar o desempenho do PostgreSQL e MySQL em modo de operação OLTP; comparar o desempenho do PostgreSQL e MySQL em modo de operação OLAP.

Para a conclusão dos objetivos, faz-se necessário o fornecimento de dados comparativos entre os SGBDs, bem como gráficos, de forma que exista uma clara visualização e entendimento entre o desempenho de ambos

2. Metodologia

Este trabalho caracteriza-se como uma pesquisa aplicada (MARCONI; LAKATOS, 2010) que aborda diferentes cenários criados para a comparação entre o PostgreSQL e o MySQL: OLAP, OLTP, e processamento de texto. Cada um foi executado três vezes, tendo sua média extraída e comparada com o concorrente.

Benchmarks são, de acordo com o dicionário Merriam-Webster, “um problema ou teste padronizado que serve como base para avaliação ou comparação (a partir do desempenho do sistema de computador)”. Utilizou-se o software HammerDB para a execução de benchmarks. HammerDB é, segundo os autores: “o software líder de benchmarking e teste de carga para os bancos de dados mais populares do mundo, oferecendo suporte a banco de dados Oracle, SQL Server, IBM Db2, MySQL e PostgreSQL.”. Este software foi escolhido pois implementa a execução de cargas baseadas nos benchmarks padrões de mercado TPC-C e TPC-H.

O benchmark TPC-C é voltado para a execução de um operações transacionais de pedidos, baseando-se em um cenário de um fornecedor de peças de atacado. O benchmark TPC-H foi desenvolvido pensando em tomada de decisões. Para isso, ele executa consultas de agregação sobre grandes conjuntos de dados com o objetivo de testar o processamento de consultas complexas. O HammerDB não é, no entanto, um substituto, mas sim uma alternativa fiel aos benchmarks TPC. Inclusive, a distribuição do software é feita pelo próprio conselho TPC.

A implementação do HammerDB difere, de forma sutil, da implementação oficial dos benchmarks. A principal diferença é a rigidez dos testes oficiais, nos quais o TPC-C só pode ser executado com tamanhos pré-determinados de base de dados, o HammerDB aceita qualquer configuração de tamanho. Outra diferença é a métrica final de desempenho, que para os benchmarks TPC oficiais, precisam levar em conta qual o hardware utilizado, enquanto para o HammerDB existe uma métrica mais simples e universal, que é o número de “Pedidos” por minuto.

Nos ambientes OLTP, as métricas são o número de transações realizadas por minuto (TPM – *Transactions Per Minute*) e NOPM (*New Orders Per Minute*). Para comparação direta entre os bancos de dados, será utilizada exclusivamente a métrica NOPM, conforme recomendado pelo software utilizado para a realização dos benchmarks. Foram realizados experimentos com a ferramenta HammerDB, no qual as execuções variaram em número de usuários por execução. Foram realizadas execuções com 1, 2, 4, 8, 12, e 16 usuários concorrentemente.

A execução do experimento com diferentes números de usuários foi realizada para que os bancos de dados sejam testados sobre um ambiente de alta demanda, exercitando suas capacidades de lidar com *locks* e concorrência. Quanto maior o número de usuários, maior a concorrência e maior a demanda por recursos compartilhados, de acordo com o tamanho da base de testes. Em bases pequenas, o efeito dos *locks* é mais visível, pois o banco de dados precisará acessar e modificar dados que possivelmente já estejam em utilização por outros usuários. Já em bancos de dados maiores, o efeito acaba não sendo tão perceptível. Isto se deve ao fato que com mais registros no banco, são menores as chances de colisões de dados. Para bancos maiores, o teste será sobre a eficiência dos bancos em utilizar seus respectivos caches, já que operações de I/O de disco são mais recorrentes.

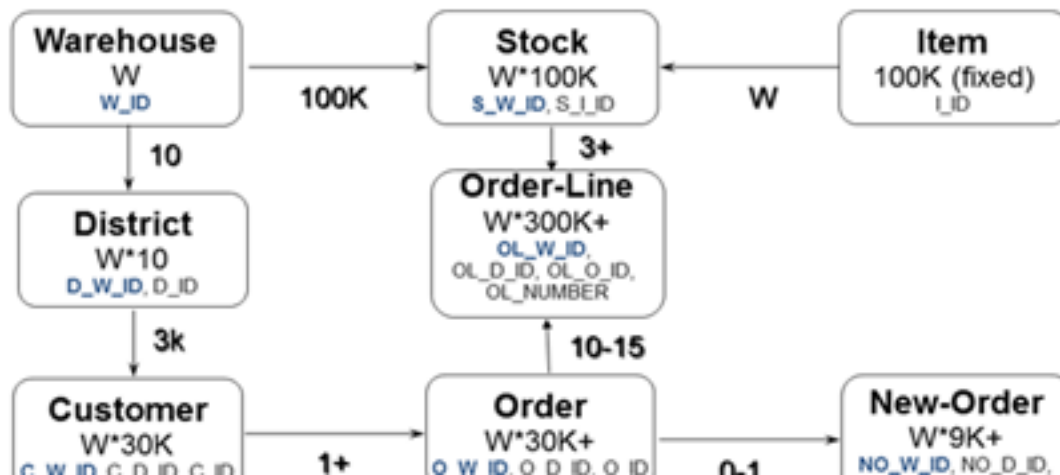
Para a comparação OLAP, existem duas métricas: tempo para alimentação da base (criação do esquema, ingestão dos dados, e criação dos índices) e o tempo para consulta, que será medido para cada query. Os testes foram realizados com apenas um usuário.

A importância da execução dos testes em ambiente containerizados vem da facilidade em reproduzir consistentemente os resultados obtidos por meio dos testes. Caso fossem executados sob um ambiente comum, os processos do sistema operacional hospedeiro poderiam interferir com o processo de benchmark, causando uma variabilidade indesejada. Além disto, a utilização de containers é uma ótima prática quando se tem a necessidade de executar operações sob um ambiente constante. Também é possível a geração de uma imagem customizada, para que esta possa ser usada como a imagem dos testes. Para este trabalho, no entanto não houve customização das imagens, pois isto diminuiria a facilidade de acesso ao mesmo tipo de ambiente deste trabalho.

O HammerDB foi executado pela máquina hospedeira, conectando através de uma network com os SGBDs dentro dos contêineres. Dentre as execuções, o container Docker é parado, de forma que valores de cache não interfiram no resultado dos testes. Além disso, a memória *swap* do *host* é zerada após a finalização de cada benchmark, para que esta métrica não interfira nos testes.

As imagens estão limitadas a utilizar 12 GB de RAM, visto que o sistema *host* possui 16 GB. Para executar os testes, apenas o aplicativo HammerDB estava aberto. Nenhum outro processo que possa representar uma utilização de recursos foi utilizado. O sistema hospedeiro conta com as seguintes especificações técnicas de hardware: 16 GB RAM; 512 GB NVME SSD; Intel Core i7-1165G7 @2.80 GHz x8; sistema operacional *host* Ubuntu 20.04.

Figura 1 – Fluxo de experimentação.



Fonte: os autores (2022).

Os SGBDs contaram com configurações de tuning básico, de forma que qualquer pessoa que precise configurar uma destas instalações consiga encontrar na internet quais parâmetros precisam ser alterados para extrair mais recursos do hardware e melhorar a performance do banco de dados.

No ambiente transacional, OLTP, sabendo da importância das propriedades ACID (Atomicidade, Consistência, Isolamento e Durabilidade) para a garantia da integridade dos dados, nenhum parâmetro que afete a segurança dos dados ou comprometa as propriedades ACID foi alterado.

Em ambientes OLTP é importante levarmos em consideração uma característica chamada “*transaction isolation*” (isolamento de transação). Hollins (2000, p. 1, tradução nossa) diz que “Um sistema ACID deve suportar o isolamento completo de uma transação em execução de outras transações em execução simultânea”. Neste contexto, *isolation* significa que uma transação não deve visualizar ou interferir em outra transação em execução paralela simultaneamente. Na teoria, isto significa que alterações realizadas em uma transação “A” são capazes de consultar e modificar dados sem que a transação em execução “B” tenha ciência das alterações de “B”. O autor complementa dizendo que “os resultados das transações em execução simultaneamente devem ser iguais a um resultado obtido pela execução dessas transações, uma de cada vez, em série (em qualquer ordem)”. Na prática, existem várias nuances para o problema de isolamento.

3. Resultados e Discussões

Como o sistema hospedeiro possui quatro núcleos físicos, com a tecnologia *Hyper-Threading*, o valor mínimo de *warehouses* seria em torno de 27. Utilizando um *scale-factor* de 100, é, portanto, um teste mais justo para ambos. *Scale-factor* é um número multiplicador de escala. O multiplicador padrão é 1. Isto significa que com este valor o tamanho do banco é conhecido e predeterminado. Segundo o site do HammerDB, no teste TPROC-C o *scale-factor* simboliza o número de *warehouses*. Este valor influencia o tamanho das outras tabelas (com exceção da tabela “item”, que tem tamanho fixo). Para diminuir a variação, os testes serão executados três vezes, e feita a média aritmética dos resultados para cada execução. O esquema criado pelo HammerDB para os testes OLTP está representado na Figura 2:

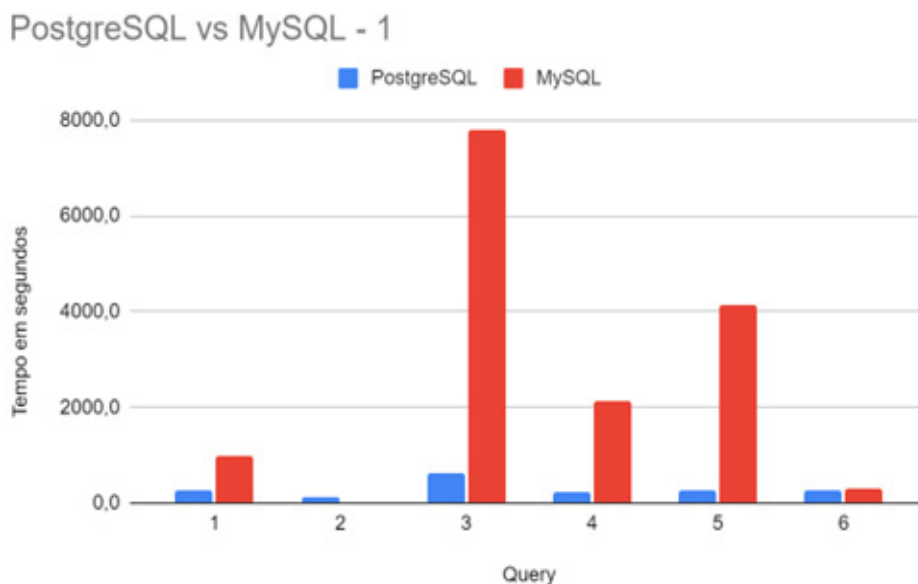
| | | |
|----|----------|----------|
| 16 | 57,398 | 447,235 |
| 17 | 469,424 | 323,641 |
| 18 | 1018,703 | 400,158 |
| 19 | 60,269 | 319,771 |
| 20 | 4641,662 | 1407,984 |
| 21 | 309,608 | 1631,297 |
| 22 | 16,986 | 53,514 |

Fonte: os autores (2022).

Os resultados das execuções mostram que o PostgreSQL é mais performático em 18 das 22 consultas. Das consultas que o PostgreSQL é mais performático, a média aritmética do percentual é de 1139,99%, e a mediana é 572,94%. Se for desconsiderada a consulta com maior variação entre os dois SGBDs (consulta 13), a média cai para 728,83%, e a mediana 545,39%. Nas 4 consultas que o MySQL é mais performático, a média da razão PostgreSQL/MySQL é de 300,47% e a mediana, 199,81%.

O Gráfico 1 está separado em grupos de 6 consultas, para facilitar a leitura. Este gráfico mostra o tempo de execução das consultas em segundos no eixo Y, portanto quanto menor a barra, mais rápida a consulta e melhor o resultado, qualitativamente.

Gráfico 1 - Tempo de execução das consultas 1 a 6.



Fonte: os autores (2022).

4. Considerações Finais

A conclusão geral é que o PostgreSQL possui um melhor desempenho quando comparado por meio do HammerDB, executando as cargas de teste TPROC-C e TPROC-H. Analisando o comportamento dos SGBDs, junto aos parâmetros que foram alterados para cada um, pode-se observar que no ambiente OLAP um dos maiores fatores contribuintes para a vantagem do PostgreSQL sobre o MySQL é a possibilidade de execução das consultas utilizando mais recursos de CPU, por meio do parâmetro “*max_parallel_workers_per_gather*”. Com a alteração deste parâmetro para o valor “8”, o PostgreSQL pode passar a utilizar até oito núcleos da CPU para processamento paralelo. Na prática, o número de núcleos utilizados depende de outros fatores, como o tamanho da tabela e o custo estimado da consulta. Do lado do MySQL,

existe funcionalidade semelhante, mas é restrita a consultas simples, sem nenhuma ligação relacional, o que limita as possibilidades de otimização de recursos. Este fator, no entanto não impacta o processamento OLTP. Nesse modo de operação as consultas são sobre pequenas quantidades de dados, e muitas das buscas são por chaves primárias. A diferença de desempenho pode ser explicada pela maturidade do motor de execução das *stored procedures* dos SGBDs. No ambiente OLAP, foi percebido algumas discrepâncias na metodologia dos testes do próprio HammerDB para o carregamento dos dados. Seria interessante que estas diferenças fossem desfeitas para que o teste seja o mais similar possível: carregamento dos dados apenas com a estrutura básica do esquema, sem índices ou chaves primárias ou secundárias. A criação das chaves e índices devem ser realizadas após a finalização do carregamento dos dados. Além disto, seria interessante obter dados de desvio padrão das execuções das consultas. A consistência dos resultados pode ser uma métrica importante para alguns leitores

Referências

- HOLLINS, Michael. Transaction isolation levels and object oriented data structures. *In: IDEA WORKSHOP*, 7th. 2000, Victor Harbor. **Proceedings** [...]. Victor Harbor: University of Adelaide, 2000.
- KANELLAKIS, Paris C. Elements of relational database theory. *In: VAN LEEUWEN, J. (ed.). Handbook of theoretical computer science: formal models and semantics – volume B*. Amsterdam: Elsevier, 1990. p. 1073-1156.
- LU, Hongjun; LIU, Hongyan. Decision tables: scalable classification exploring RDBMS capabilities. *In: EL ABBADI, A.; BRODIE, M. L.; CHAKRAVARTHY, S. (ed.). VLDB'00: proceedings of the 26th international conference on very large data bases*. San Francisco: Morgan Kaufmann Publishers Inc., 2000. p. 373.
- MARCONI, M. A.; LAKATOS, E. M. **Técnicas de pesquisa**: planejamento e execução de pesquisas, amostragens e técnicas de pesquisas, elaboração e interpretação de dados. São Paulo: Atlas. 2010.

LEI GERAL DE PROTEÇÃO DE DADOS VERSUS MANIPULAÇÃO DE DADOS

General Data Protection Law versus Data Handling

Thays Popper Ciardullo¹

Nader Ghoddosi¹

Resumo: O objetivo desse trabalho é apresentar de forma clara e sucinta a Lei Geral de Proteção de Dados Pessoais, que entrou em vigor somente em setembro de 2020 e como ocorre a manipulação dos dados desses usuários, trazendo um contexto histórico até a constituição da lei e os principais fatores que levaram a sua criação. Trazendo casos que ocorreram antes da lei ser criada, para entender a importância dela atualmente e como a lei descreve caso a empresa ou instituições forem penalizadas.

Palavras-Chave: Lei; Manipulação; Dados; Tecnologia.

Abstract: The objective of this work is to present clearly and succinctly the General Law for the Protection of Personal Data, which came into force only in September 2020 and how the manipulation of these users' data occurs, bringing a historical context to the constitution of the law and the main factors that led to its creation. Bringing cases that occurred before the law was created, to understand its importance today and how the law describes if the company or institutions are penalized.

Keywords: Law; Manipulation; Data; Technology.

Introdução

Com a chegada da Era da Informação, em meados de 1950, iniciou-se o marco em que parte crescente da vida começou a mudar para o ‘mundo digital’. Com a criação da internet, possibilitou-se gerar uma infraestrutura global que deu origem a uma vasta gama de serviços de comunicações eletrônicas e novas estruturas de comunicação. Em 1989, dois engenheiros da Organização Europeia para a Pesquisa Nuclear (CERN), Robert Cailliau e Tim Berners-Lee, desenvolveram um novo sistema de informação baseado na internet, designado de World Wide Web (WWW), auxiliando posteriormente na expansão da utilização da internet. Por meio da WWW, tornou-se possível criar servidores de informação, onde se incluem textos e imagens, gerando uma verdadeira teia de informação (GOETHALS; AGUIAR; ALMEIDA, 2000). Com a criação do navegador Mosaic, em 1993, foi possível tornar a web pública, sendo a infraestrutura utilizada pela internet a rede mundial de telecomunicações.

Não apenas novas formas de comunicação foram estabelecidas, como o e-mail, serviços de mensagens ou plataformas de mídia social, mas agora incluindo atividades cotidianas, como recuperação de informações bancárias ou compras, podem ser gerenciadas rapidamente pela internet. De acordo com Albers e Sarlet (2022), o mundo online vem de mãos dadas com a digitalização e a “datificação”. Esses desenvolvimentos são, portanto, acompanhados por um potencial crescente de geração e aquisição de dados.

Dados, por sua vez, no âmbito computacional ou informacional, segundo Sawaya (1999, p. 111), entende-se aos elementos básicos que são fornecidos, processados ou produzidos pela máquina. Algumas vezes, considera-se como dados somente as formas numéricas, porém qualquer informação deve ser vista como um dado. Dados digitais, por conseguinte, referem-se

¹ Centro Universitário Leonardo da Vinci – UNIASSSELVI – Rodovia BR 470 - Km 71 - no 1.040 – Bairro Benedito – Caixa Postal 191 – 89130-000 – Indaial/SC Fone (47) 3281-9000 – Fax (47) 3281-9090 – Site: www.uniasselvi.com.br.

a quaisquer dados que sejam criados, armazenados, usados, transferidos e/ou manipulados por meio de tecnologia de computador, neste caso, dados sensíveis e de natureza pessoal. Estes dados digitais descrevem informações coletadas sobre indivíduos, incluindo suas preferências e hábitos pessoais na internet, bem como informações financeiras ou pessoais privadas, coletadas enquanto usuários usam dispositivos ou computadores conectados à internet (DOWD, 2022).

Com a área emergente da inteligência artificial e a expansão das bases de dados nos últimos anos, notou-se uma crescente geração de dados por parte dos usuários da internet, fornecendo ‘matéria-prima’ para que entidades tanto privadas quanto públicas vissem a oportunidade lucrativa que a aquisição e comércio dessas informações geram, tornando assim os dados pessoais uma valiosa commodity atualmente (DOWD, 2022; ALBERS; SARLET, 2022), porém essa prática entrou em questionamento a partir do momento que se infringiu princípios de privacidade das informações ou dos metadados adquiridos, levando assim, a criação de leis que visam a proteção e privacidade dos usuários bem como consensualizar boas práticas de manipulação de dados.

Este trabalho, portanto, visa considerar o que advém com os dados dos usuários na internet e os aspectos relacionados à vigilância, controle, privacidade e manipulação com relação ao cenário brasileiro e entender como as empresas manipulam nossos dados quando baixamos aplicativos ou acessamos sites. A metodologia usada será a análise da Lei Geral de Proteção de Dados (LGPD), passando por suas principais características e princípios para enfim chegar-se à conclusão deste.

2 Lei geral de proteção de dados

Nesta seção é apresentada a Lei Geral de Proteção de Dados, desde sua criação e implementação, como também um caso recente de mau uso dos dados, o que levou também em consideração a criação da Lei.

2.1 Apresentando a lei

A Lei Geral de Proteção de Dados Pessoais (LGPD) é a Lei nº 13.709, publicada em agosto de 2018, sancionada pelo presidente da república, mas entrou em vigor apenas no dia 18 de setembro de 2020 e, por conta das consequências da pandemia da Covid-19, foi então acordado que as penalizações passariam a ser aplicadas a partir de agosto de 2021.

A lei regula o tratamento de dados de pessoas físicas e jurídicas no Brasil, tanto no âmbito online quanto offline, nos setores privados e públicos, tendo como objetivo proteger a privacidade e autonomia dos consumidores em relação aos seus dados pessoais. Com isso, a LGPD veio para devolver aos usuários o controle sobre seus dados pessoais.

Os dados pessoais são caracterizados como, por exemplo, o número do PIS, CPF, RG, além de endereço. Também a Lei da Proteção de Dados leva em consideração dados ditos ‘sensíveis’, que podem ser citados: filiação a organizações políticas, religião, informações genéticas e de biometria, como também orientação sexual, gênero ou origem étnica.

A lei ainda define os direitos dos dados e obriga as empresas pedirem o consentimento das pessoas para utilizarem suas informações. Um ponto importante a ser levantado, é que o país já utilizava de mais de 40 normas jurídicas que, direta e indiretamente, tratavam da privacidade aos dados pessoais. Com a LGPD, o Brasil entra para o rol de mais de 100 países que hoje podem ser considerados apropriados para proteger a privacidade e o uso dos dados.

De acordo com Sá (2019), os principais fatores que motivaram a criação de uma lei de proteção de dados foram:

-
- Uso cada vez maior de dados pessoais.
 - O recurso mais valioso do mundo não é mais petróleo, mas dados (*The Economist*).
 - Era digital, redes sociais, “*analytics*”.
 - Possibilidade de impactos nas vidas das pessoas, negócios e até eleições.
 - Ausência de um marco regulatório nacional quanto à proteção de dados 16.
 - Acontecimentos recentes e outros fatores.
 - Cambridge Analytica.
 - GDPR (*General Data Protection Regulation* – União Europeia).
 - OCDE (Organização para a Cooperação e Desenvolvimento Econômico).
 - Lei do Cadastro Positivo.
 - Ano Eleitoral.

As sanções das quais a lei dispõe, nos casos de falha na proteção dos dados, estão previstas nos artigos 52, 53 e 54. Nesse sentido, destaca, Mariana Louback Lopes (2019):

O capítulo VIII da LGPD é voltado ao processo de fiscalização do cumprimento das regras de tratamento e proteção de dados pessoais, conforme dispostas pelo texto legal. A competência fiscalizatória e sancionatória incumbe majoritariamente à Autoridade Nacional de Proteção de Dados, porém sem excluir o eventual cabimento de demais ações administrativas, civis ou penais aplicáveis a cada caso.

De acordo com o art. 52 da LGPD, as possíveis sanções são as seguintes:

- I - Advertência, com indicação de prazo para adoção de medidas corretivas.
- II - Multa simples, de até 2% do faturamento no seu último exercício, excluídos os tributos, limitada a R\$ 50 milhões por infração.
- III - Multa diária.
- IV - Publicação da infração após devidamente apurada e confirmada a sua ocorrência.
- V - Bloqueio dos dados pessoais a que se refere a infração até a sua regularização.
- VI - Bloqueio dos dados pessoais a que se refere a infração até a sua regularização.
- VII - Eliminação dos dados pessoais a que se refere a infração (BRASIL, 2018).

O parágrafo primeiro do artigo explica que as sanções são aplicadas após instaurado um procedimento administrativo, que deve prever a possibilidade de ampla defesa ao acusado.

2.2 Caso Cambridge Analytica

O escândalo mais atual do uso indevido de dados de usuários da internet se tomou conta no ano de 2018, quando foi a conhecimento público que as informações de milhares de usuários da rede social Facebook foram coletadas por diversas empresas, a mais citada foi a Cambridge Analytica, a qual foi responsável posteriormente por usar tais dados para traçar e prever personalidades de usuários de uma determinada região, a fim de manipular e influenciar na corrida presidencial a favor do ex-presidente norte-americano Donald Trump.

Este acontecimento começou com o uso de um aplicativo de teste denominado *myPersonality* criado por David Stillwell, em 2007, no Centro de Psicometria da Universidade de Cambridge. Os dados coletados por meio do Facebook pelo aplicativo foram distribuídos por acadêmicos da Universidade de Cambridge a centenas de pesquisadores por meio de um site que não possuía medidas de segurança, o que levou a ficar aberto para acesso ilegal por quatro anos (REHMAN, 2019). O aplicativo em questão fazia perguntas em forma de quiz levando em consideração cinco traços de personalidade conhecidos como ‘*Big Five*’, que são: abertura (o quanto você está aberto a novas experiências?), consciência (o quanto você perfeccionista?), extroversão (o quão sociável é você?), amabilidade (o quão atencioso e cooperativo você é?), neuroticismo

(você fica facilmente chateado?). Com essas informações, fica relativamente fácil determinar o tipo de personalidade de um indivíduo. De acordo com Rehman (2019), Stillwell e colaboradores provaram que com uma média de 68 curtidas de um usuário do Facebook, era possível prever questões pessoais desse indivíduo como cor de pele, orientação sexual e entre outros.

Em 2014, Christopher Wylie, um cientista de dados canadense, foi trabalhar para Alexander Nix em uma empresa chamada SCL Elections Ltd. Posteriormente, Nix, um dos iniciadores da Cambridge Analytica (CA) na América, introduziu Wyle para Steve Bannon (ex-vice-presidente da CA, além de ter sido estrategista-chefe da Casa Branca e ex-chefe da campanha eleitoral de Donald Trump em 2016). Ambos Wylie e Nix discutiram com Bannon a ideia sobre como a opinião dos eleitores poderia ser influenciada durante as eleições, porém para que esse projeto desse partida, haveria necessidade de recursos financeiros, portanto Banner apresentou a ideia para Robert Mercer, o qual é um cientista da computação e bilionário, usou sua fortuna para se tornar um dos homens mais influentes na política dos EUA como um dos principais doadores republicanos.

A ideia proposta por Nix e Wylie era de combinar *microtargeting* (customização de mensagens de marketing para um público-alvo) da qual existia na política e então visar indivíduos não apenas como eleitores, mas também como personalidade, e eventualmente, criar um perfil psicológico de cada eleitor em uma determinada região, ou, neste caso, em todo os Estados Unidos.

A ideia foi baseada em um artigo publicado em 2014 na PNAS, uma das revistas científicas multidisciplinares mais citadas no mundo, pela Universidade de Cambridge intitulado: ‘*Computer-based personality judgments are more accurate than those made by humans*’, que na tradução livre ficaria: ‘Julgamentos de personalidade baseados em computador são mais precisos do que aqueles feitos por humanos’. Artigo este que foi criado por meio dos dados coletados por Stillwell e colaboradores com o aplicativo *myPersonality*. Assim, Christopher Wylie precisaria de dados e primeiro abordou Michal Konsinski, um dos autores do artigo. As negociações falharam, então outro psicólogo e um dos colegas de Konsinski, Aleksandr Kogan, ofereceu-lhe uma solução que poderia replicar a pesquisa original de Stillwell e Konsinski (REHMAN, 2019). Mercer ficou impressionado com a ideia e investiu na época quinze milhões de dólares na Cambridge Analytica para executar o projeto.

Aleksandr Kogan desenvolveu um aplicativo chamado *thisisyourdigitallife* que apresentava um quiz de personalidade e questões políticas. Sua empresa, a *Global Science Research* (GSR), em parceria com a CA pagou entre \$2,00 e \$5,00 dólares para aproximadamente 320 mil eleitores norte-americanos para realizarem o quiz do aplicativo por meio de suas contas do Facebook. O aplicativo também coletava dados como curtidas e informações pessoais dos usuários, assim como também coletava informações dos amigos destes usuários, tendo um montante de mais de 50 milhões de dados brutos do Facebook.

Algoritmos combinaram dados com outras fontes, como registros dos eleitores, para criar um conjunto superior de registros. Os participantes deste aplicativo concordaram em aceitar seus dados coletados para propósitos acadêmicos, porém nenhum deles concordou que seus dados fossem vendidos para empresas como a Cambridge Analytica que posteriormente os transformou em uma ferramenta política lucrativa (REHMAN, 2019).

A campanha de Donald Trump havia contratado a Cambridge Analytica durante a eleição de 2016, para executar operação de dados, identificar eleitores para atingi-los com publicidade altamente personalizada com base em seus dados de personalidade, também onde fazer paradas de campanha e ajudar na comunicação estratégica, como o que dizer em discursos. A ideia também era atingir os eleitores com mensagens negativas contra o candidato rival na eleição, com isso, Trump conseguiu ganhar 77.000 votos em três estados-chave (Michigan, Wisconsin e Pennsylvania), levando a sua vitória como presidente eleito em 2016.

Esse “ataque digital”, de acordo com Granville (2018), levou à violação dos termos e condições do Facebook, que proíbem o compartilhamento ou venda de dados “para qualquer rede de anúncios, corretor de dados ou outro serviço relacionado a publicidade e monetização”. O Facebook acabou tomando conta desta situação em 2015 e imediatamente banuiu o aplicativo de Kogan de sua plataforma, e exigiu que Kogan e Cambridge Analytica destruíssem todos os dados que haviam coletados ilegalmente, porém, em 2018, o Facebook soube que os dados não foram excluídos depois que o denunciante, Christopher Wylie, revelou que os dados coletados do aplicativo de Kogan foram usados para criar perfis “psicográficos” de pessoas e fornecer material pró-Trump online, como citado anteriormente.

3 Manipulação de dados pessoais presente no dia a dia

Será que os nossos dados pessoais estão sendo protegidos conforme a LGPD que entrou em vigor em setembro de 2020? Com o avanço tecnológico crescendo dia após dia, muitos sites e aplicativos surgem com muita facilidade aos seus usuários, como podendo ser utilizado para realizar compras com cartão de crédito/débito, pagamentos de contas/boletos, listas de mercado, realizar pesquisas, registrar funcionário pelo eSocial e outros, tudo de forma online.

Conforme os exemplos citados anteriormente, podemos perceber que os dados circulam livremente na internet. De acordo com a revista britânica *The Economist*, na era da Quarta Revolução Industrial em que estamos entrando, os dados são o nosso “novo petróleo”. Isso nos mostra o quão valioso são os dados pessoais e que com isso podem surgir problemas ao titular desses dados, como por exemplo: manipulações que afetam o direito de escolha, que é o caso do escândalo envolvendo a Cambridge Analytica citado anteriormente, roubo de informações, uso não autorizado de dados para fins ilícitos e entre outros.

É comum navegar na internet e acessar vários sites, baixar aplicativos (aplicativos de delivery, por exemplo), redes sociais e realizar cadastros, e, muitas vezes, nem percebemos que os dados estão sendo captados. Passamos constantemente pelos termos de utilização sem lê-los, abrindo brechas para ilegalidades, o que acontece também em celulares, quando os usuários são obrigados a consentir aos termos, caso contrário, será impedido de utilizar o aplicativo.

Outro ponto que ocorre comumente por quem navega na internet é a grande aparição de anúncios de produtos em suas redes sociais, logo após de realizarem pesquisas na internet, como por exemplo no site da Google. Isso ocorre porque os smartphones ficam conectados a uma conta pessoal, a qual possui todos os dados pessoais do usuário e utiliza vários recursos para mostrar anúncios de seu interesse. Por esse motivo, a Lei Geral de Proteção de Dados é de suma importância, uma vez que traz a necessidade de tratamento dos dados dos titulares.

3.1 Cookies

De acordo com França (2015, p. 96) o *cookie* é um “pequeno arquivo de texto armazenado pelo navegador (web browser), funcionando como uma ‘carteira de identidade’ do usuário, permitindo a memorização de dados e o reconhecimento de hábitos de navegação”, de modo que podem ser convertidos em importantes informações para a manutenção de sites e a veiculação de publicidade personalizada mediante mecanismo de vigilância da atividade do

usuário em rede. Os cookies são arquivos que são depositados no computador do indivíduo pelo site acessado e que possibilitam a identificação deste usuário com o intuito de facilitar a funcionalidade da página e o monitoramento da navegação, de forma que podem ser oriundos das páginas visitadas, de outras entidades (*cookies* de terceiros), apagáveis (*cookies* de sessão) ou persistentes (*cookies* permanentes) (CASTELLUCCIA, 2012). Os principais tipos de *cookies* são: (i) os do próprio domínio (primários ou *first party cookies*); (ii) os definidos por terceiros (denominados *third party cookies*), os que somente funcionam enquanto estiver aberta determinada página (*cookies* de sessão) e os que atuam mesmo depois que o site ou página for fechada (*cookies* permanentes) (ALDEIAS, 2012 *apud* FRANÇA, 2015, p. 98).

Permitir os cookies faz com que os sites possam receber e processar informações importantes para melhor experiência de uso e remover os cookies é fácil, porém torna a navegação mais complexa. Sem eles, cada visita a uma determinada página será como a primeira vez, nenhuma informação será guardada.

Considerações finais

Conforme visto no presente trabalho, a tecnologia está em constante evolução, e presente cada vez mais em nossas vidas. Este trabalho analisou o impacto da Lei Geral de Proteção de Dados (LGPD), tendo por objetivo aferir a forma da responsabilização civil dos fornecedores de serviços pelo tratamento inadequado dos dados pessoais dos usuários, trazendo casos que ocorreram antes da lei ser criada, para entender a importância dela atualmente e como a lei é aplicada caso a empresa ou instituições forem penalizadas.

Ressalta-se que como a lei foi implementada recentemente e tem como base a lei previamente criada na Europa (*General Data Protection Regulation – GDPR*) em 2018, ainda há necessidade de se entender como a LGPD se comportará no cenário brasileiro. Muitas empresas e instituições já começaram a se adequar, gerando termos para que os usuários online tenham ciência de como seus dados são usados, armazenados e se serão ou não compartilhados com terceiros, mas não se sabe ainda se possui uma abrangência de amplo efeito que proteja de fato, em todos os níveis, os dados compartilhados. Existem poucos estudos na literatura que aborde a LGPD e seus impactos, sendo necessários para questionar sua efetividade. É importante também que esta lei seja regularmente revista e atualizada, pois o ambiente online está em constantemente evolução e aprimoramento.

Referências

ALBERS, M.; SARLET, I. W. (ed.). **Personality and data protection rights on the internet: brazilian and german approaches**. Cham: Springer International Publishing, 2022. v. 96

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Presidência da República, 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm. Acesso em: 13 set. 2022.

CASTELLUCCIA, Claude. Behavioural tracking on the internet: a technical perspective. *In*: GUTWIRTH, Serge *et al.* (eds). **European data protection: good health?** Dordrecht: Springer, 2012.

DOWD, R. **The birth of digital human rights**: digitized data governance as a human rights issue in the EU. Cham: Springer International Publishing, 2022.

FRANÇA, Lilian Cristina Monteiro. Vigilância e políticas de privacidade na sociedade pós-cookie: o caso do The Guardian. **Revista Eco Pós**, Rio de Janeiro, v. 18, n. 2, p. 95-105, 2015. Disponível em: https://revistaecopos.eco.ufrj.br/eco_pos/about/contact. Acesso em: abr. 2022.

GRANVILLE, Kevin. Facebook and Cambridge Analytica: what you need to know as fallout widens. **The New York Times**, New York, mar. 2018. Disponível em: <https://www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html>. Acesso em: 9 maio 2022.

GOETHALS, K.; AGUIAR, A.; ALMEIDA, E. **História da internet**. 1999. 41 f. (Mestrado em Gestão da Informação) – Universidade do Porto, Porto, 1999.

LOPES, Mariana Louback. Capítulo VIII: fiscalização. *In*: SIQUEIRA, Antônio Henrique Albani; FEIGELSON, Bruno (coord.). **Comentários à lei geral de proteção de dados**. 2. ed, São Paulo: Thomson Reuters Brasil, 2019.

SÁ, Marcelo Dias de. **Análise do impacto da nova lei de proteção de dados pessoais nas aplicações de internet das coisas**: aplicações mobile do governo. 2019. 39 p. Monografia (Especialização em Informática) – Universidade Federal de Minas Gerais, Brasília, DF, 2019.

REHMAN, Ikhlâq Ur. Facebook-Cambridge Analytica data harvesting: what you need to know. **Library Philosophy and Practice**, Lincoln, v. 1, p. 1-13. jan. 2019.

SAWAYA, M. R. **Dicionário de informática e internet**: inglês-português. São Paulo: Nobel, 1999.

INTELIGÊNCIA ARTIFICIAL APLICADA AOS NON-PLAYABLE CHARACTERS (NPCS) EM JOGOS DO GÊNERO ROLE-PLAYING GAME (RPG)

Artificial intelligence applied to non-playable characters (npcs)
in role-playing game (rpg) genre

Daniela Maccali¹

Luiz Felipe Teixeira Bauman¹

Elzimar José de Carvalho Neto²

Resumo: Este trabalho aborda o que é a Inteligência Artificial aplicada aos jogos digitais, com ênfase nos NPCs em jogos de RPG. Foi utilizada a pesquisa explicativa para entender o que é a IA, em artigos científicos, revistas e em sites da área de jogos para entendimento do que é a IA e como ela é aplicada nos NPCs do gênero RPG. Também foi utilizada a pesquisa exploratória, onde realizou-se uma pesquisa de campo, com a aplicação de um questionário com o público consumidor deste gênero. A pesquisa é do tipo quantitativa, com gráficos comparativos, além da pesquisa qualitativa, que foi a coleta de informações diretamente com os jogadores que responderam o questionário, para complementar e corroborar os resultados obtidos. Foi constatado que os jogadores preferem que a IA dos NPCs inimigos seja a mais adaptativa, pois isso traz um jogo mais imersivo, emocionante, que prende a atenção, fazendo com que a vitória seja mais gratificante.

Palavras-chave: Inteligência Artificial. IA. Jogos Digitais. NPC. RPG.

Abstract: This paper aims to show what Artificial Intelligence is when applied to digital games, with emphasis on NPCs in RPG games. Explanatory research was used to understand what AI is, in scientific articles, magazines and gaming websites to understand how it is applied to NPCs in the RPG genre. Exploratory research was also used, and a field research was carried out, with the application of a questionnaire to consumers of this genre. The research is of the quantitative type, with comparative graphics, in addition to the qualitative research, which was the collection of information directly from the players who answered the questionnaire, to complement and corroborate the results obtained. It has been found that players prefer the AI of enemy NPCs to be the most adaptive, as this makes for a more immersive, exciting, attention-grabbing game, making victory more rewarding.

Keywords: Artificial Intelligence. AI. Digital Games. NPC. RPG.

Introdução

Neste artigo, abordaremos a Inteligência Artificial – IA aplicada aos jogos digitais, com ênfase na aplicação nos NPCs (do inglês – Non Playable Characters) em jogos do gênero RPG (do inglês Role-Playing Game). Embora a IA seja um assunto comumente falado hoje em dia, pouco ainda se fala a respeito da aplicação em jogos.

O primeiro passo é entender o que se espera de uma IA, que é a capacidade de imitar o comportamento humano e não apenas representar sua forma de falar e modismos. Quando falamos sobre NPC, referimo-nos aos personagens que não são jogáveis, como a própria sigla sugere. Direcionado aos RPG's, esses NPCs têm um certo grau de importância, tanto na narrativa quanto ao gameplay de determinado jogo. É bom lembrar que a IA nem sempre é inteligente o suficiente (no caso de inimigos), e isso pode fazer com que a experiência do jogador não seja tão boa quanto deveria, devendo-se testar e variar as escolhas dos NPC's, visando a melhor experiência possível, tornando o jogo agradável nas mãos e controles dos jogadores.

Queremos criar jogos imersivos, que façam o jogador estar na pele do seu personagem, vivenciando uma experiência enriquecedora. Para tanto, existem vários fatores para que isso ocorra: o enredo do jogo, a narrativa, o gameplay, personagens, entre outros. Os NPCs fazem parte do jogo de RPG, pois eles interagem com o personagem de diversas formas.

Para compensar a falta de IA nos RPGs, desenvolvedores de jogos independentes e as grandes empresas procuram criar inimigos com muito mais vida e bem mais fortes que o jogador para balancear a falta de raciocínio do mesmo, sendo um exemplo visto em jogos como Hack n' Slash e vários outros jogos em que o adversário simplesmente solta golpes aleatórios torcendo que algum acerte.

A IA entra nos jogos digitais para dar mais vida aos NPCs, tanto aliados, e principalmente inimigos, pois o que acontece geralmente é que as ações são limitadas e previsíveis.

Buscando entender as expectativas dos jogadores, foi realizada uma pesquisa de campo para termos dados sobre o que é esperado de uma IA de um jogo digital. O objetivo desta amostragem é ter parâmetros de análise quanto à importância da IA, as preferências quanto à IA aplicada em aliados ou inimigos, assim como outros que serão demonstrados ao decorrer deste artigo, considerando que existem dois tipos principais na criação de IA, a primeira é a tomada de decisão, e a segunda é a aprendizagem (RAÍZEN, 2015).

Fundamentação teórica

A busca de um jogo imersivo. Quando se fala em imersão em um jogo digital, há muitas coisas envolvidas, narrativa, sons, ambiente, personagens etc. Além de um bom enredo, personagens marcantes, cenário e elementos de acordo com o contexto, e claro NPCs, temos o elemento IA.

A IA entra como um elemento importante para criar imersão, trazendo um comportamento mais humano para os NPCs. A imprevisibilidade faz parte do comportamento do ser humano, bem como o ambiente à volta também tem imprevisibilidades. Segundo Tenorio, Alves e Reis (2018, p. 19):

O objetivo da IA para jogos é fornecer um comportamento para objetos ou entidades, cujo comportamento simule um movimento de inteligência, por exemplo, no caso de um jogo de RPG (Role-Playing Game), pode ser um movimento de ataque de um NPC (Non Player Characters, personagens do jogo que não são controlados pelo humano). Por exemplo, se o jogador passar perto de um NPC, o movimento de ataque pode ser realizado com a utilização de técnicas de Inteligência Artificial.

Ainda conforme Tenorio, Alves e Reis (2018), existem várias técnicas de IA que são utilizadas no desenvolvimento de jogos, segundo Pathfinding (2013, s.p.):

Máquinas de Estado Finito, é a técnica mais usada para expressar o comportamento dos personagens de um jogo, onde ocorrem as mudanças de ações dos personagens; Pathfinding, é a técnica que verifica o melhor caminho que o personagem vai percorrer; Padrões de Movimento, definem os movimentos de pernas, braços, gestos, enfim, animação do personagem e ações de ataque e defesa; Sistemas Baseados em Regras, permite a modelagem de comportamentos complexos, mas exige um poder de processamento alto; Lógica Fuzzy, complementa a tradicional lógica booleana, representando valores que estão entre o verdadeiro e falso, por exemplo, expressões do tipo pouca vida; Redes Neurais Artificiais, técnica responsável pelo aprendizado adquirido com personagem humano, como resultado gera um comportamento diferente para surpreender o jogador; Algoritmos Genéticos, esta técnica desenvolve diferentes objetos ou indivíduos a partir de um DNA virtual.

A função dos NPCs é aparentar ser um jogador real, com comportamentos, racionalidade, inteligência e personalidade, com pouca ou nenhuma influência humana. Além disso, eles devem proporcionar diferentes dificuldades, desafiando o jogador para que este não se frustre por não poder agir como protagonista.

Principalmente em jogos do gênero RPG, os NPCs são muito importantes, pois sem eles é um mundo sem história e sem vida, às vezes não está implícito, mas grande parte das vezes

é o que salva um jogo. O cenário e a história também são partes importantes, o jogo *Skyrim* mostra muito bem isso, você pode fazer o que você quiser no jogo, ser um mago, um ladino ou até mesmo um herói de guerra, o jogador é quem faz a escolha.

A IA aplicada a jogos digitais possibilita trabalhar num jogo mais imersivo, trazendo a imprevisibilidade, tirando a monotonia de ações previsíveis. Tratando de IA aplicada a NPCs, é o que faz o jogo ser divertido, cativante, que o jogador tenha prazer em jogar. O personagem principal, controlado pelo jogador, é elemento fundamental, mas os NPCs que sejam bem construídos, que tragam um bom complemento, certamente fará o jogo ser bem-sucedido.

A IA de um NPC se baseia em um conjunto de regras pré-estabelecidas que busca reproduzir o comportamento mais adequado possível conforme os dados extraídos do ambiente. As escolhas das execuções são influenciadas pela percepção das alterações no espaço em que está inserido no ambiente virtual. O comportamento também pode ser definido pelo uso do agente inteligente, o qual é baseado em métodos e atributos que são responsáveis pela monitoração das execuções do personagem que o jogador está interpretando e pela tomada de decisão.

No caso do jogo *The Elder Scrolls V: Skyrim*, a IA antiga era previsível, tanto que se fosse explorar o reino, saberíamos que teriam algumas frases bordões e os diálogos eram limitados. Segundo Eler (2017, s.p.), referente ao jogo *Skyrim*, “A nova inteligência artificial, chamada de “CIF-CK”, é capaz de aumentar o jogo de cintura dos NPCs do jogo. Ela possibilita aos coadjuvantes um leque muito maior de ações – que os permite reagir e se adaptar ao comportamento de quem joga”.

A IA CIF-CK (CIF vem de *Comme il-Faut*) foi criada por James McCoy e Michel Mateas e o primeiro jogo a ter essa inteligência artificial implementada foi o jogo *Prom Week* (Figura 1), que era um jogo simples feito para os jogadores se socializarem uns com os outros (Fatima Toolkit, 2022, tradução nossa), muito parecidos com jogos como *Habbo Hotel* e *Club Penguin*, que são jogos extremamente famosos no mundo da internet. Essa IA foi construída pensando no aprendizado com as interações sociais.

Figura 1 – Jogo *Prom Week*



Fonte: Disponível em: <<https://promweek.soe.ucsc.edu/>>. Acesso em: 1 dez. 2022.

Voltando ao *Skyrim*, ele se torna o jogo que é por causa de seus NPCs, mesmo eles não tendo a IA mais inteligente já criada e, em algumas vezes, fazendo muitas coisas óbvias, como exemplo, uma frase específica sendo dita repetidamente em um curto espaço de tempo; são eles que fazem o mundo em si ganhar vida. Eles criam a ilusão de que o jogador não está sozinho no vasto mundo da história, diferente de RPGs on-line, que o NPC tem um único trabalho que é passar a missão para o jogador.

Outros jogos ganham destaque quanto ao NPC, *The Witcher*, *Fallout* e até mesmo *Cyberpunk* fazem um bom trabalho na hora de criar vida para esses personagens. Porém, existe um jogo em si que é pouco reconhecido e tem uma ótima IA, que é o jogo *Dragon's Dogma* (Figura 2). Lançado em 2012, desenvolvido e publicado pela CapCom, é um jogo Online Single Player (você não joga com os jogadores e sim com os *pawns* (peões, do inglês) – um NPC criado pelo jogador), do gênero de RPG, de mundo aberto.

A IA é extremamente interessante, pois ela aprende e se desenvolve baseada nas ações do jogador. Quando é criado um personagem, juntamente é criado um *pawn*, em que podem ser escolhidas certas inclinações: a maneira que seu NPC se comporta, se ele vai te ajudar, se ele vai te proteger, buscar por recursos etc. O que torna o *pawn* um NPC interessante não são essas inclinações em específico, mas sim o passar do tempo, mesmo com as inclinações escolhidas, dependendo dos atos do jogador, eles podem aprender a fazer outras coisas, por exemplo: se você dar muito um comando de ataque, ele entende que você está mandando ele ir para cima do inimigo. Depois de um certo tempo, mesmo o *pawn* sendo um mago, ele vai lutar quando identificar um inimigo, pois foi algo que o jogador o acostumou a fazer.

Figura 2 – Capa do jogo *Dragon's Dogma*



Fonte: Disponível em: <<https://www.dragonsdogma.com/pt-br/>>. Acesso em: 1 dez. 2022.

Outro fator interessante é o comportamento dos *pawns*. O jogador geralmente aprende qual é a fraqueza de seus inimigos ou o caminho que deve fazer para terminar uma missão. O *pawn* não tem tal conhecimento, porém, depois de matar uma quantidade de monstros, ele aprende as fraquezas deles, ou, depois de passar algumas vezes em uma área, ele acaba aprendendo o caminho da missão que você está fazendo e, com isso, se outros jogadores pegarem seu *pawn*, ele vai dizer "você deve seguir esse caminho" ou "há um tesouro por aqui", pois você já explorou e ensinou o caminho para ele.

Por ser um jogo mais antigo, existem diversos problemas com a IA, o sistema de inclinações não é muito prático, pois não tem um tutorial no jogo explicando como funciona e, com isso, logo o jogador acaba dando comandos para o *pawn* sem saber que está mudando suas inclinações, tendo a chance deles serem um grande motivo para ver na tela um Game Over. De todo modo, ainda é muito interessante como um jogo de 2012 pode ter criado uma IA tão interativa e divertida.

Quando se trata inimigos, podemos citar o Sistema *Nemesis*, criado pela Monolith Productions, patentado pela Warner Bros. Games até o ano de 2035, que se trata de uma IA que vai se adaptando conforme as ações do jogador sobre o mundo afetando o NPC inimigo. Conforme o que vai acontecendo, eles vão evoluindo e se adaptando, ou seja, em estratégia ou em batalhas, lembrando-se da progressão do jogador no modo história e dos acontecimentos, reutilizando posteriormente esse conhecimento.

Aplicado no mundo do jogo, esse sistema gera proceduralmente os inimigos com vantagens e desvantagens e que possa ter uma conexão com outro inimigo, seja como rival ou amigo (também tem como irmão de sangue). Cada vez que o inimigo sobrevive a um confronto direto contra o player, tanto de um golpe mortal ou porque o jogador fugiu, ele ganha bônus para subir de nível e ter armas, o mesmo valendo para os NPCs que estão ao seu lado.

Praticar os mesmos atos na luta gera uma adaptação no inimigo, pois ele aprendeu com os conflitos anteriores. Ele fica mais esperto, podendo levar consigo sentimentos como raiva ou medo, e, conforme eles sobrevivem, ficam mais fortalecidos e com características únicas. Dependendo do que acontece no mundo, eles adquirem cicatrizes e algum sentimento, dando, assim, mais personalidade para os inimigos (se ele for atacado por um animal, talvez na próxima vez que reencontre terá medo do tipo de animal que quase matou ele).

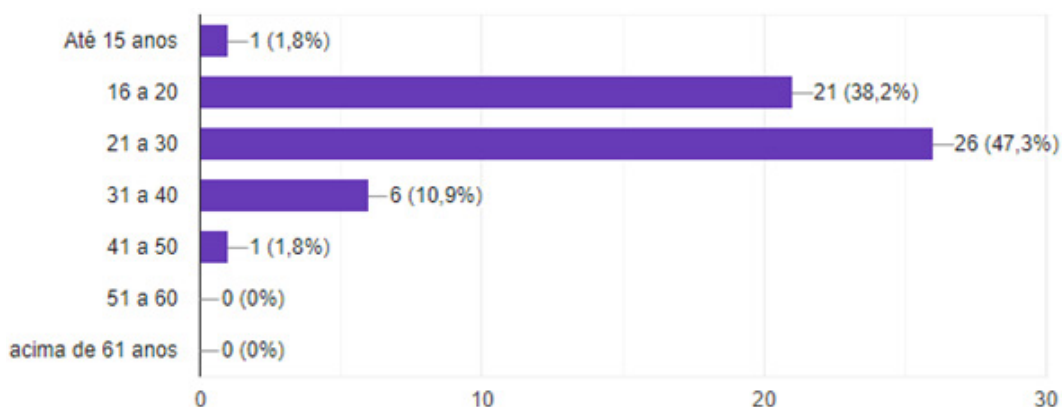
Essa IA de aprendizado se estende aos aliados, pois as consequências também afetam eles (se trair um aliado provavelmente ele se tornará seu inimigo e lhe emboscará). Inimigos e aliados podem posteriormente mudar de lado. Os jogos que possuem o sistema *Nemesis*: *Middle Earth – Shadows of Mordor* e *Sombras da Guerra*, *Gotham Knights*.

Pesquisa de campo

Foi realizada uma pesquisa de campo, aplicando-se um questionário sobre alguns aspectos esperados em um jogo, focando no tema IA, totalizando 55 participantes. Foi utilizado o Google Forms para coletar as respostas, sendo aplicado de forma on-line.

Conforme o Gráfico 1, a representação maior dos entrevistados está na faixa etária dos 21 aos 30 anos.

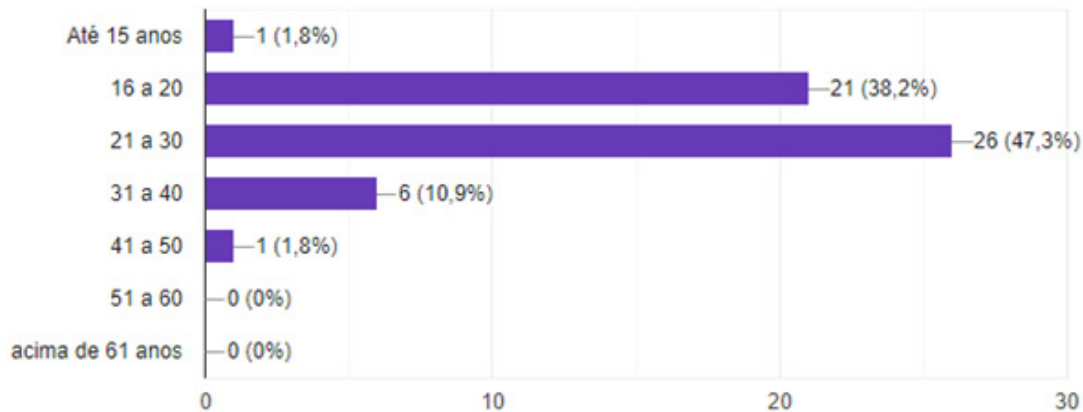
Gráfico 1 – Faixa etária



Fonte: o autor

No Gráfico 2 está a representação dos jogos votados quanto à IA inimiga que o jogador considera a mais inteligente, sendo que os 2 jogos que tiveram mais votação e empataram foram *Shadows of Mordor* e *The Witcher*, com 18,2% cada. Na sequência, temos *Horizon Zero Dawn*, com 16,4% e *Skyrim*, com 10,9%. Outros jogos votados com pequena representação são: *Dragon Age Inquisition*, *CyberPunk*, *Dragon's Dogma*, *Ifrit*, *The Division 2*, *Red dead Redemption 2*, *Alien Isolation*, *Middle Earth: Shadow of War*, *Metal Gear Solid V*, *F.E.A.R.*, *The Last of us 2*, *Star Renegade*, *Dark Souls*, *Shin Megami Tensei Digital Devil Saga 2*, *Dishonored*.

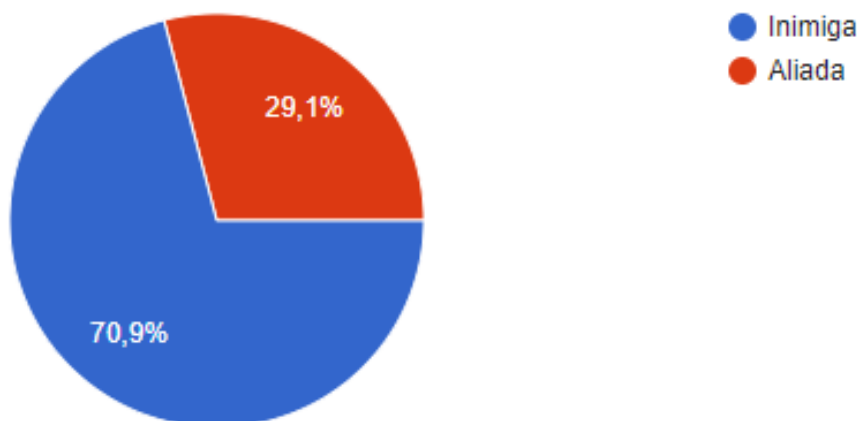
Gráfico 2 – Jogos com a IA inimiga considerada mais inteligente



Fonte: o autor.

A preferência pela IA mais adaptativa, se inimiga ou aliada, conforme o Gráfico 3, podemos observamos que os jogadores preferem disparadamente que a IA inimiga seja mais adaptativa que a IA aliada.

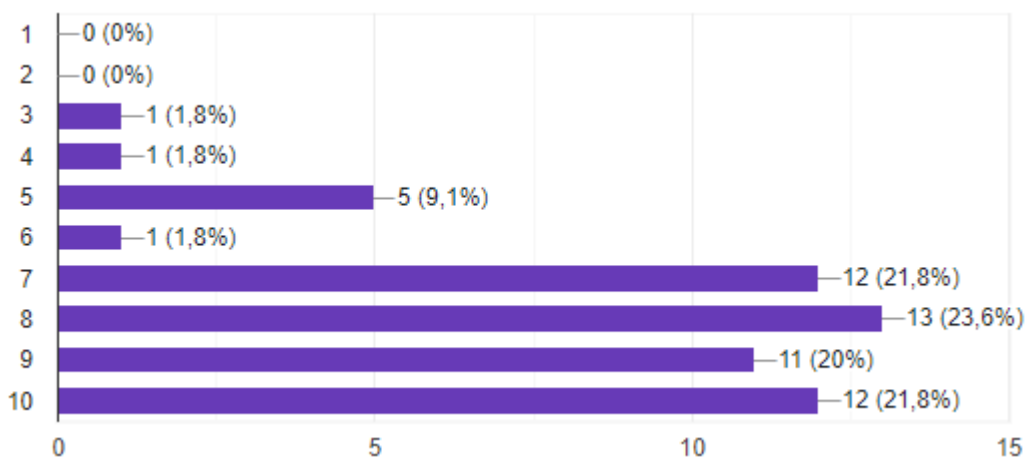
Gráfico 3 – Preferência pela IA mais adaptativa



Fonte: o autor.

Perguntamos aos jogadores qual a importância da IA, numa escala de 1 a 10, comparado com todos os outros aspectos em um jogo de RPG (gráficos, história, música, mecânica). Observamos, pelo Gráfico 4, que 87,2% da votação está entre as notas 7 e 10.

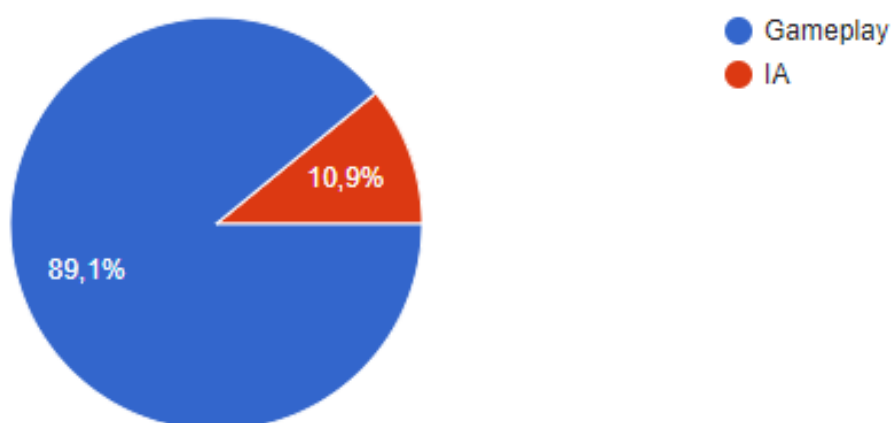
Gráfico 4 – Importância da IA no jogo



Fonte: o autor.

Podemos perceber pela representação do Gráfico 5, quando feita a pergunta para o jogador sobre que ele acha mais importante, a IA ou o Gameplay do personagem principal, ainda o personagem principal tem muito peso num jogo, com 89,1% dos votos.

Gráfico 5 – Preferência entre IA ou gameplay do personagem principal



Fonte: o autor.

Metodologia

Foi utilizada pesquisa exploratória, realizando uma pesquisa de campo e buscando entender as preferências dos jogadores. Para descobrir que tipo de IA gera mais interesse, aplicou-se um questionário, além da coleta de opiniões diretamente com os jogadores para entender quais as expectativas diante de alguns itens dos jogos digitais.

Também foi utilizada a pesquisa explicativa, buscando entender os conceitos de IA, sendo pesquisado em artigos científicos, sites e blogs da área de jogos de RPG.

Os resultados apresentados são quantitativos, pois utilizamos de gráficos para representar o que jogadores preferem em uma IA em jogos de RPG. No entanto, a pesquisa também é qualitativa, pois ouvimos e coletamos diversos comentários entre os jogadores que participaram da pesquisa, e com essas respostas descobrimos o porquê a IA inimiga teve uma maior porcentagem de votos.

Resultados e discussões

De acordo com os resultados da pesquisa de campo aplicada, verificamos que os jogadores preferem que a IA dos NPCs inimigos seja melhor que a IA dos NPCs aliados.

Baseado nos resultados desta pesquisa, entendemos que os jogadores querem jogos imersivos, e, para que isso ocorra, o jogo precisa ter desafios e ser envolvente.

Facilmente percebemos isso pela votação entre os jogos. *The Witcher* (a terceira edição é a mais agressiva), *Shadow of Mordor* e *Horizon* foram os mais votados, pois a IA inimiga tem qualidades que trazem desafios. *The Witcher* não tem uma IA excepcional, porém a maioria dos jogadores sempre acaba dizendo que a IA dele é extremamente agressiva, os inimigos te pegam de surpresa, vem para cima do jogador em grupos.

A partir desse questionário, acabamos perguntando para os jogadores “e se a IA fosse um pouco menos agressiva e mais convidativa na hora do combate você acha que seria algo ruim?” e a maioria dos jogadores responderam “sim”.

Shadow of Mordor já passa algo um pouco mais único, muitos dos jogadores reconhecem que não é um jogo bom, sua história é fraca, a música não se destaca, nem mesmo o cenário é tão incrível e seu estilo de combate é bem parecido com *Batman*, porém, o que faz as pessoas jogarem é a IA dos chefes, eles se adaptam baseado no combate anterior que eles tiveram com o jogador. Podemos concluir que muitos dos que votaram no jogo não votaram porque o jogo é bom, e sim porque a IA é única e interessante.

Horizon tem uma IA que se adapta bem nos combates, segundo um jogador anônimo “*Horizon* tem uma IA bem parecida com *Monster Hunter*, você consegue perceber que os monstros são bem espertos, eles atacam nos momentos certos, te cercam, aprendem um pouco seu estilo de luta”.

Segundo um outro jogador anônimo “A maioria dos RPGs que joguei os NPCs eram muito “fracos” e era fácil saber um padrão deles”, corroborando com o ponto crucial que um jogo vai ser mais emocionante e marcante se não for previsível.

A pesquisa mostrou que a IA inimiga sendo mais adaptativa faz com que “você tenha que pensar antes de agir e bolar uma estratégia para matar algum inimigo em específico”, conforme outro jogador.

Dark Souls é tão adorado no mundo dos jogos, por exemplo, às vezes o jogo é extremamente cruel, mas depois que o jogador aprender e executar essas ações, será bem recompensado e, aquilo que era difícil e talvez frustrante, torna-se divertido.

Após discutirmos e analisamos os dados coletados, entendemos o porquê dos jogadores fazerem tanta questão de confrontarem uma IA inteligente. Eles preferem um jogo desafiador, pois mesmo que você morra várias e várias vezes, a vitória se torna algo mais satisfatória, uma conquista que em jogos simples você não consegue adquirir.

Conclusão

Este artigo começou com a o intuito de pesquisar sobre a IA aplicada aos NPCs em jogos de RPG, mas percebemos que só buscando fundamentos teóricos não seria o suficiente, então buscamos conhecimento de pessoas jogadoras do gênero mencionado. Para tal, criamos uma pesquisa de campo para saber o que essas pessoas esperavam da IA.

Para esta pesquisa de campo, foi utilizado um questionário, direto e simples, o suficiente para entender as preferências desse público.

Verificamos, no decorrer do desenvolvimento deste artigo, o que é a IA através de pesquisas bibliográficas e em sites especializados da área de jogos, para entender o que ela representa nos jogos digitais.

Não tínhamos ideia do que os jogadores esperavam, mas com a pesquisa de campo compreendemos o desejo da maioria deles.

De início nós fomos surpreendidos como IA dos NPCs inimigos faz a diferença, pois o que o jogador quer é desafios, superar obstáculos inteligentes e que sejam imprevisíveis, buscando a recompensa pelo tempo dispendido nas inúmeras tentativas de aniquilar a força inimiga.

Podemos concluir que conhecer o público consumidor dos jogos é muito importante para termos informações, entender expectativas e desejos, tendo assim boas oportunidades de criação de jogos cada vez mais imersivos e que façam sucesso com esses jogadores.

Referências

BUTCHER, B. **A patente do Nemesis System da Warner Bros. Games é validada até o ano de 2035.** 2021. Disponível em: <https://www.gamevicio.com/noticias/2021/02/a-patente-do-nemesis-system-da-warner-bros-games-e-validada-ate-o-ano-de-2035/>. Acesso em: 27 out. 2022.

CAPCOM. **Dragon's Dogma.** 2022. Disponível em: <https://www.dragonsdogma.com/pt-br/>. Acesso em: 18 set. 2022.

ELER, G. Ferramenta torna os figurantes de Skyrim superinteligentes. **Super Interessante**, São Paulo, 4 ago. 2017. Disponível em: <https://super.abril.com.br/tecnologia/ferramenta-torna-os-figurantes-de-skyrim-superinteligentes/>. Acesso em: 25 ago. 2022.

FATIMA TOOLKIT. **The Comme il Faut Asset.** Disponível em: <https://fatima-toolkit.eu/cif-ck/>. Acesso em: 10 nov. 2022.

LIMA, F. U. R. de. **Evoluindo Redes Neurais Artificiais para geração de NPCs verossímeis utilizando Visão Artificial.** 2018. 61 f. TCC (Graduação) - Curso de Ciência da Computação, Universidade Federal do Ceará, Campus de Crateús, Cratéus/CE, 2018. Disponível em: https://repositorio.ufc.br/bitstream/riufc/38760/1/2018_tcc_%20FU%C3%A1lison.pdf. Acesso em: 23 ago. 2022.

LUIZ OLIVEIRA. **O que é NPC? O essencial que você precisa saber.** 2021. Disponível em: <https://rpgtips.com.br/o-que-e-npc-o-essencial-que-voce-precisa-saber/>. Acesso em: 25 set. 2022.

RAIZEN. **Programando #3 - Inteligência Artificial.** 2015. Disponível em: <https://centrorpg.com/index.php?topic=11460.0>. Acesso em: 31 ago. 2022.

SOONER. **O que é que Dragon's Dogma tem.** 2019. Disponível em: <https://regames.wordpress.com/2019/05/27/o-que-e-que-dragons-dogma-tem/>. Acesso em: 10 set. 2022.

SOUZA, G. L. de. **Análise de medidas biométricas em jogos de realidade virtual.** 2017. 61 f. Monografia (Bacharel) - Curso de Ciência da Computação, Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2017. Disponível em: <https://www.lume.ufrgs.br/bitstream/handle/10183/174997/001065156.pdf?sequence=1>. Acesso em: 27 ago. 2022.

TENORIO, M. B.; ALVES, E. N.; REIS, Y. C. Principais técnicas da Inteligência Artificial para jogos. **Revista Alomorfia**, Presidente Prudente, v. 2, n. 11, p. 18-29, 28 nov. 2018. Disponível em: <https://fatecpp.edu.br/alomorfia/index.php/alomorfia/article/view/4>. Acesso em: 28 ago. 2022.

VIJAYARA. **Prom Week Selected for E3 Indiecade Showcase!** 2012. Disponível em: <https://promweek.soe.ucsc.edu/>. Acesso em: 11 nov. 2022.

WP INFO. **A revolução da IA chegando às cenas mod de Skyrim e The Witcher 3.** 2021. Disponível em: <https://br.atsit.in/archives/72202>. Acesso em: 25 ago. 2022.

Testes de Qualidade, Estruturação e Implementação em um Sistema Corporativo

Quality Tests, Structuring and Implementation in a Corporate System

Guilherme Jackson Lamim¹

Nader Ghoddosi

Resumo: A realização de testes no desenvolvimento de softwares é fundamental para a obtenção da qualidade, visto que os clientes se tornaram cada vez mais exigentes devido à complexidade de seus interesses na área tecnológica. O objetivo deste estudo envolve a explanação de como a atividade de teste de qualidade no ciclo de vida do sistema, pode auxiliar na obtenção da qualidade do produto. Para isso, foi realizada a estruturação e implementação de uma rotina de teste de qualidade em uma empresa de tecnologia. Este artigo apresenta considerações sobre teste no contexto da qualidade de software, a metodologia desenvolvida, uma aplicação prática e conclusões sobre a metodologia.

Palavras-chave: Teste de software. Qualidade de software. Estruturação de testes.

Abstract: Testing in software development is essential to obtain quality, as customers have become increasingly demanding due to the complexity of their interests in the technological area. The objective of this study involves the explanation of how the activity of quality testing in the life cycle of the system, can help to obtain the quality of the product. The structuring and implementation of a quality test routine was carried out in a technology company. This article presents considerations about testing in the context of software quality, the methodology developed, a practical application and conclusions about the methodology

Keywords: Software testing. Software quality. Test structuring.

Introdução

Segundo o *Testing Company*, o controle de qualidade e os testes para o desenvolvimento de softwares, sistemas e aplicativos são, muitas vezes, vistos como dispensáveis pelas organizações, que acabam tendo prejuízos ao perceberem as consequências de falhas ignoradas desde o início do projeto. Empresas de pequeno e médio porte fazem, no geral, um controle de qualidade por meios informais, o que motiva a adoção de um processo qualificado para elevar o nível de seus produtos, visto que os testes são uma das técnicas que assegura a qualidade dos produtos a partir de identificação de defeitos.

Ao longo dos anos, muitas tentativas foram realizadas para o aperfeiçoamento das técnicas no desenvolvimento de sistemas e em especial, na fase de teste de software. E, é nesta fase que se inicia o encarecimento do projeto, visto que se o software não funciona corretamente pode levar a muitos problemas, dentre eles podemos citar perda de dinheiro com retrabalho desnecessário ou até mesmo ferimentos ou morte. É comum que ainda hoje tenhamos algum tipo de problema na usabilidade dos sistemas como reservas de avião, ônibus, hotéis ou operações financeiras por pouco ou nenhum tipo de testes. Nesses casos, a vida de milhares de pessoas pode ser impactada negativamente, por isso a atenção com os testes é essencial. Segundo Paula (2005), é incorreto substituir prazo e/ ou custo por qualidade. Qualidade é consequência das pessoas, processos e tecnologia. Os defeitos são inseridos em todas as fases do desenvolvimento do sistema e acontecem por causa das limitações humanas, erros de inter-

¹ Centro Universitário Leonardo Da Vinci – UNIASSELVI – Rodovia BR 470 – Km 71 – nº 1.040 – Bairro Benedito – Caixa Postal 191 – 89130-000 – Indaial/SC Fone (47) 3281-9000 – Fax (47) 3281-9090 – Site: www.uniasselvi.com.br

pretação e lógica, desconhecimento de técnicas, falta de atenção ou ainda falta de motivação. Para ter qualidade garantida, as etapas retiram parte dos defeitos que foram inseridos, porém, quando as etapas não são efetuadas corretamente deixa-se de serem retirados os defeitos.

Os testes de qualidade têm por objetivo prevenir e encontrar falhas no sistema e verificar se os mesmos foram corrigidos, visando garantir uma alta qualidade no produto evitando um problema no futuro. Pressman (2005) afirma que a qualidade do software é a conformidade com os requisitos funcionais e não funcionais que têm sido explicitamente declarados; padrões de desenvolvimento que tenham sido claramente documentados e características implicitamente esperadas de todo software a ser desenvolvido.

Este trabalho se propõe a apresentar a estruturação de um processo de teste de qualidade em um setor de desenvolvimento em uma empresa de tecnologia, bem como explicar todo o processo de testes de qualidade e sua importância no processo

Referencial Teórico

Com a inovação dos computadores após a segunda guerra mundial, teve-se um grande aumento no estímulo da indústria de computadores e o processo de construir softwares que automatizam processos que antes eram realizados manualmente se tornou parte do cotidiano das organizações.

Segundo Prikladnicki (2014, p. 18), “a indústria de software tem evoluído para se tornar uma das mais importantes indústrias da era moderna.” Atualmente, o software está presente em uma grande parte do nosso dia a dia, desde os afazeres mais básicos aos mais complexos e diversos. O ambiente no qual as empresas de software estão inseridas é altamente competitivo e, por isso, elas se esforçam para serem as primeiras a lançar novidades, produtos e serviços diferenciados.

A inserção de processos de qualidade de software é algo que vem deixando as empresas inquietas. Inicialmente, a atividade de teste era encarada simplesmente como a tarefa de navegar pelo código e corrigir problemas já conhecidos. Essas tarefas eram realizadas pelos próprios desenvolvedores, não existindo recursos dedicados exclusivamente a essa atividade, de maneira que os testes eram realizados tardiamente. Apesar dessa situação estar associada a uma má prática de desenvolvimento, ela continua presente em muitas organizações (BARTIÉ, 2002).

O objetivo dos testes é revelar defeitos, para que estes possam ser corrigidos antes da entrega de uma versão ou da entrega final do produto (Bartié, 2002). De acordo com Pressman (2006, p. 288), “Uma estratégia de teste integra métodos de projeto de casos de testes em uma série bem planejada de passos que resultam na construção bem sucedida de software.”

Conceitos Gerais

Antes de entender sobre o processo de teste de software, é preciso entender alguns conceitos. Para Bartié (2002, p. 22), “Teste é um processo sistemático e planejado que tem por finalidade única a identificação de erros”. Então, após a identificação dos erros é possível definir qual será a estratégia utilizada que garanta a qualidade do software. Além disso, os testes estão presentes em diversas fases do processo de desenvolvimento de um software, e com isso, existem alguns níveis de testes que é preciso entender:

Testes de unidade: os testes de unidade são os mais básicos e costumam consistir em verificar se um componente individual do software (unidade) foi implementado corretamente. Esse componente pode ser um método ou procedimento, uma classe completa ou, ainda, um pacote de funções ou classes de tamanho pequeno e moderado. Os testes de unidade costumam ser realizados pelo próprio programador, e não pela equipe de teste. (WAZLAWICK, 2011, P. 292)

Testes de integração: para Humble e Fearley (2014, p. 97), devem ser aplicados “se sua aplicação se comunica com vários sistemas externos por meio de diferentes protocolos, ou

se a aplicação em si consiste em uma série de módulos de baixo acoplamento com interações complexas entre eles”.

Testes de sistema: conforme Pressman (2011, p. 409), são “[...] uma série de diferentes testes cuja finalidade primária é exercitar totalmente o sistema”. Os testes de sistema são executados pensando no sistema como um todo e em como o usuário vai utilizar o mesmo.

Testes automatizados: Bartié (2002, p. 198) define que os testes automatizados como “a utilização de ferramentas de teste que possibilitem simular usuários ou atividades humanas de forma a não requerer procedimentos manuais no processo de execução de testes”. Os testes automatizados podem ser utilizados para realização de testes funcionais, de regressão e de carga ou performance.

O objetivo de um teste é verificar todos os requisitos funcionais e não funcionais de um sistema, como segurança, performance, robustez (Delamaro, 2007). Um bom planejamento de testes é importante para que os requisitos sejam contemplados conforme o esperado sem nenhum bug ou falha. Um planejamento do processo de teste consiste na definição dos propósitos e a abordagem do teste como especificar técnicas e tarefas de teste adequadas e formular um cronograma de teste.

Abordagem de Testes

Existem duas abordagens de testes: caixa preta e caixa branca. As duas são muito importantes e devem fazer parte de todo e qualquer planejamento de testes de qualidade de software. Ambas se diferenciam na forma de realizar os testes e abordaremos cada uma abaixo:

Caixa preta (funcional): os testes funcionais são gerados a partir de uma análise entre os dados de entrada e de saída, não se preocupando com o código. Testes de caixa preta servem para verificar a conformidade do que foi implementado aos respectivos requisitos funcionais, sendo utilizados, geralmente, no final do processo de desenvolvimento (Franzen, 2005).

Exemplos de testes funcionais: particionamento de equivalência, análise de valor limite ou condição-limite, teste exaustivo, teste de comparação, teste orientado por sintaxe, teste baseado em tabela de decisão e grafo de causa-efeito.

Caixa branca (estrutural): ao contrário dos testes de caixa preta, os testes de caixa branca, ou testes de estrutura, se baseiam no código-fonte do software, onde as ferramentas de diagnóstico procuram pontos fracos e erros estruturais, fornecendo uma lista que permite que sejam feitas ações corretivas posteriormente. Em geral, esse teste é realizado por desenvolvedores, e não por testadores do sistema e podem ser aplicados desde o início do processo de desenvolvimento (Rup, 2010).

Exemplos de testes estruturais: teste de caminho básico ou abrangência de instrução, abrangência de ramificação/condição, abrangência de caminho, teste de caminho em grafos com loop e teste de estrutura de controle (testes de condição, de fluxo de dados e de laços).

Tipos de Testes

Atualmente se tem a necessidade de criar softwares cada vez mais confiáveis e com isso, muitas técnicas surgem diariamente com o intuito de verificar e validar a qualidade dos produtos. Dentre os vários tipos de testes que podem ser aplicados, apresentarei alguns exemplos a seguir:

Teste funcional: este tipo de teste serve para verificar se todos os requisitos funcionais foram implementados de acordo com as regras do negócio. Assim, garantem que não haja diferença entre requisitos funcionais e o comportamento do software construído. Este teste submete a todas as possíveis entradas, podendo detectar assim, todos os possíveis defeitos (DELAMARO, 2007).

Teste de desempenho: o teste de performance, como também é conhecido, tem como principal objetivo verificar se o desempenho está consistente com os requisitos definidos,

empregando um volume de transações nas situações previstas de pico máximo de acesso e concorrência, e tempo de resposta obtidos nos testes, e assim, compará-los com os valores limites especificados (BARTIÉ, 2002).

Teste de código: este teste tem por objetivo deixar o código mais legível, reusável e bem estruturado, verificando se o código está de acordo com as padronizações de linguagem, estruturação, e afins.

Teste de usabilidade: são testes que enfatizam fatores humanos, estética, consistência na interface do usuário, ajuda on-line e contextual, assistentes e agentes, documentação do usuário e material de treinamento (RUP, 2010).

Teste de segurança e controle de acesso: estes testes são realizados para garantir segurança ao usuário ao efetuar login no sistema, limitar acessos aos dados disponíveis, assegurando que somente pessoas autorizadas poderão ter acesso a informações restritas (RUP, 2010).

Integridade dos dados: testes destinados a avaliar a robustez do objetivo do teste (resistência a falhas) e a compatibilidade técnica em relação à linguagem, sintaxe e utilização de recursos. Esse teste é implementado e executado em vários objetivos do teste, como unidades e unidades integradas (RUP, 2010).

Teste de interface do usuário: o teste de interface do usuário verifica a interação do usuário com o sistema, com o objetivo de assegurar que a interface fornecerá ao usuário o acesso e navegação adequados, além disso, o teste deve assegurar que os objetos contidos na interface funcionem como esperado.

Planejamento da Qualidade do Software

Segundo PMBOK (2004), os processos de gerenciamento da qualidade do projeto incluem todas as atividades da organização executora que determinam as responsabilidades, os objetivos e as políticas de qualidade, de modo que o projeto atenda às necessidades que motivaram sua realização. Eles implementam o sistema de gerenciamento da qualidade através da política, dos procedimentos e dos processos de planejamento da qualidade, garantia da qualidade e controle da qualidade, com atividades de melhoria contínua dos processos conduzidas do início ao fim, conforme adequado.

De acordo com Certified Tester Syllabus Foundation Level CTFL 3.1, a organização de um processo de desenvolvimento de software é um fator criterioso para o sucesso do modelo adotado e tem como base para os principais modelos de qualidade. Não existe um processo universal de teste de software, mas um processo de teste específico e apropriado em qualquer situação depende de fatores como: de que maneira que a atividade vai ser implementada? Quais atividades de testes estão envolvidas? Quando as atividades de testes serão executadas?

Entretanto, o grande objetivo dos programas de qualidade e produtividade é a melhoria contínua. Para Harrington (1993, s. p.), a melhoria contínua é a busca da perfeição. Para tal, a melhoria contínua vai além da definição de qualidade que, segundo ele, “é sempre fazer corretamente o trabalho”, assumindo que, perfeição, “é sempre fazer corretamente o trabalho certo”, com o objetivo de satisfazer os clientes internos e externos.

Segundo Martin (1998), é comum pensar somente nas grandes melhorias, porém pequenas mudanças podem ser grandes nos quesitos qualidade e produtividade. A melhoria não é um fim em si próprio, portanto precisa ser contínua. O autor coloca, também, que a melhoria contínua de processos é baseada no método japonês chamado KAIZEN, em que “todos melhoram tudo, o tempo todo” (MARTIN, 1998, s. p.) Ou seja, cada participante da organização identifica problemas, fazem análises e propõem soluções.

Estudo de caso

A escolha desse caso para estudo foi em função de trabalhar aspectos referentes ao aperfeiçoamento da Qualidade de Software de um sistema de uma grande empresa. Nesta empresa, uma equipe do setor de desenvolvimento utilizava o Scrumban, e o próprio desenvolvedor era quem realizava os testes unitários ao finalizar uma tarefa, garantindo a funcionalidade da aplicação. Porém, o processo de desenvolvimento da empresa mudou, fazendo com que os setores readequassem suas atividades e com isso surgiu a necessidade de reorganizar o processo de testes de qualidade da equipe.

Um estudo foi realizado com base em uma análise dos processos acerca da equipe e sua funcionalidade para o melhor funcionamento do negócio, juntamente com todos da equipe. Foram adotadas algumas novas ferramentas e a ferramenta Azure DevOps foi a selecionada na equipe para ser utilizada na nova rotina.

DevOps é um conjunto de boas práticas de engenharia de software que unifica desenvolvimento (dev) e operação (ops). Com o objetivo de aumentar a produtividade e a integração entre equipes, DevOps proporciona melhores processos para automação e monitoramento de todas as fases da construção de um software.

Quando falamos de Azure DevOps, falamos de uma solução exclusiva do sistema do Microsoft Azure, que combina pipelines e open-sources muito personalizados e extremamente automatizadas que se adequam às demandas de trabalho. O Azure DevOps Service engloba todos os serviços para elevar os fluxos de trabalho. Vamos ressaltar algumas aplicações e como o Azure DevOps pode ser usado:

Planejamento: definição e detalhamento das propriedades do trabalho a ser desenvolvido. Nesta etapa a equipe também monitora o desenvolvimento das tarefas de cada um.

Desenvolvimento: nesta etapa o Azure DevOps dispõe de mais ferramentas que são fornecidas para deixar esse processo mais produtivo, qualificado e automatizado fazendo uma integração mais contínua das fases de codificação, testes, correções, entre outros.

Fornecimento: o Azure DevOps auxilia na melhora da implantação e configuração da demanda para que o fornecimento dos aplicativos seja feito com mais confiança e tranquilidade.

Operação: para garantir que a entrega do aplicativo seja concluída com efetividade, é preciso confirmar sua execução e desempenho. Com o auxílio do Azure DevOps, a equipe pode supervisionar o funcionamento do aplicativo, solucionando rapidamente os possíveis bugs.

Com esta ferramenta, tornou-se possível criar testes de planos desde o início do refinamento de negócios, garantindo uma maior qualidade no critério de aceite da tarefa, e, na fase do desenvolvimento, os desenvolvedores passaram a realizar testes de unidade a partir dos testes de casos definidos inicialmente no refinamento. Na fase de homologação, passou a ser realizado um teste de caixa preta pelo Quality Assurance, seguindo o teste de caso e garantindo a total aderência dos critérios de aceitação da tarefa.

Impacto

O impacto na implantação do processo, como já era esperado, foi grande, uma vez que os desenvolvedores já estavam acostumados com o antigo processo. O principal impacto observado foi a adaptação dos desenvolvedores a ter a presença de um QA desde o início do planejamento das atividades. Outro ponto foi a adaptação dos testadores, pois passaram a achar muitos defeitos no software, principalmente nos testes de regressão, que eram pouco aplicados anteriormente.

Ainda será implementado um teste de regressão com Cypress, uma técnica do teste de software que consiste em manter a funcionalidade do sistema mesmo com as novas alterações, porém, após o segundo mês utilizando a nova ferramenta e o processo passou a ser bem visto pelos envolvidos, sendo possível realizar o planejamento cumprindo todas as fases do processo no seu devido escopo de prazo e qualidade.

Resultados

Os resultados encontrados com o processo implantado foram muito positivos. A princípio, com a implantação dos novos processos, muitos dos envolvidos tiveram dúvidas, mas com o desenvolvimento das atividades, passaram a ver o processo como um todo. Com o processo de qualidade implantado, foi possível também iniciar um mapeamento de bugs, essencial para o controle da gestão e melhoria de processos.

Os stakeholders, com a satisfação dos usuários do sistema, sentiram o impacto positivo com o processo implantado no sistema e pretendem expandir o mesmo processo para outros projetos, visto que a efetividade na identificação dos bugs antes de chegar ao cliente final aumentou 40%.

Considerações finais

Os testes de qualidade de software estão se mostrando cada vez mais importantes e necessários para manter o padrão e competitividade nas empresas, uma vez que a falta dele pode vir a ocasionar grandes prejuízos. Desta forma, os testes de qualidade são utilizados para verificar se o software está funcionando conforme o esperado e identificar possíveis erros. Baseando-se nisso, este artigo foi feito para demonstrar a implantação do processo de testes de uma empresa de desenvolvimento de software a fim de promover uma melhoria na qualidade dos produtos.

O estudo foi realizado com base em uma análise dos processos acerca da equipe e sua funcionalidade para o melhor funcionamento do negócio, juntamente com todos da equipe. O processo de desenvolvimento do software se atualiza e se adapta em um ritmo muito acelerado, existindo cada dia novas tecnologias, processos e ferramentas. As empresas necessitam acompanhar essas mudanças para que continuem competitivas no mercado, portanto é necessário se adaptar diariamente.

6. Referências

Azure DevOps: o que é e como começar a usar. Storm, 02/03/2021 <<https://storm.kumuluz.com.br/azure-devops/>>. Acesso em: 11 de junho de 2022.

BARTIÉ, A. **Garantia da qualidade de software**: adquirindo maturidade organizacional. Editora Campus: Rio de Janeiro, 2002

BAUNGARTNER, C. **Descubra a importância e os benefícios que a Qualidade de Software pode gerar para a sua empresa**. Testing Company, 2021. Disponível em: <<https://testingcompany.com.br/blog/descubra-a-importancia-e-os-beneficios-que-a-qualidade-de-software-pode-gerar-para-a-sua-empresa>>. Acesso em: 11 jun. 2022.

DELAMARO, M. E.; MALDONADO, J. C.; JINO, M. **Introdução ao Teste de Software**. Rio de Janeiro: Elsevier, 2007.

MARTIN, J. **A grande transição**. Ed. Futura: São Paulo, 1998.

PMBOK. Um guia do conjunto de conhecimentos em gerenciamento de projetos. Project Management Institute (PMI). 3 ed. 2004.

PRIKLADNICKI, R. **Métodos Ágeis para Desenvolvimento de Software**. Editora Bookman: Porto Alegre, 2014

PRESSMAN, R. S. **Engenharia de Software**. 6 ed. São Paulo: McGraw-Hill, 2006.

FRANZEN, M. B.; BELLINI, C. G. Porto. Arte ou Prática em Testes de Software?

REAd, vol. 11 n. 3, mai./jun., 2005.

HARRINGTON, J. **Aperfeiçoando processos empresariais**. Makron Books Editora: São Paulo, 1993.

RUP. Rational Unified Process. Disponível em: <<http://www.wthreex.com/rup/portugues/>>. Acesso em: 30 jan. 2023.

TAFNER, J.; SILVA, A. C. da; WEIDUSCHAT, I. **Normas para apresentação de trabalhos acadêmicos**. 3. ed. Indaial: ASSEVI, 2004.

WAZLAWICK, R. **Engenharia de Software: Conceitos e Práticas**. Ed Elsevier: São Paulo, 2011.

OOB! (OUT OF THE BOX): DESENVOLVIMENTO DE UM APLICATIVO EDUCACIONAL CONSTRUTIVISTA

OOB! (OUT OF THE BOX): DEVELOPMENT OF A CONSTRUCTIVE EDUCATIONAL APPLICATION

João Vitor Maia Neves Cordeiro¹

Rodrigo Ramos Nogueira²

Resumo: O construtivismo é uma teoria proposta por Jean Piaget embasada na descrição de que o aprendizado é construído pelo aluno em estágios e conforme suas experiências. Assim, aplicando conceitos construtivistas, o presente artigo descreve as etapas do desenvolvimento de uma aplicação educacional, que através de experiências construídas seguindo conceitos do movimento, utilizando-se de padrões e linguagens Web para promover uma interação entre o aluno e esse ambiente de conhecimento, por meio de animações e gráficos relacionados às disciplinas de matemática, física, química, biologia e programação de computadores

Palavras Chave: Educação. Aprendizado. Explicações exploráveis. Apache Cordova.

Abstract: Constructivism is a theory proposed by Jean Piaget based on the description that learning is constructed by students in stages and according to their experiences. Thus, applying constructivist concepts, this article describes the stages of development of an educational application that, through experiences built following concepts of the movement, using Web standards and languages to promote interaction between the student and this knowledge environment, through through animations and graphics related to the disciplines of mathematics, physics, chemistry, biology and computer programming

Keywords: Education. Learning. Explorable explanations. Apache Cordova.

Introdução

A utilização da informática por motivos educacionais no Brasil iniciou na década de 70 com o projeto Educom. O projeto visava utilizar computadores para o ensino de física na USP de São Carlos (Ministério da educação, 2009). Desde então, houve diversas reformas educacionais propondo o objetivo de “Educação de qualidade para todos”, mas usando o ensino superior como exemplo, houve um aumento de 6,6% de alunos entre os anos 2000 e 2010 (IBGE, 2010). Porém, enquanto a quantidade de alunos aumentou significativamente, a qualidade do ensino não acompanhou o crescimento, sendo o Brasil o 66º lugar de 70 países avaliados pelo PISA (OECD, 2015).

Pode-se afirmar, que desde a década de 70 a maior alteração no nosso modo de vida foi a entrada da tecnologia no cotidiano da população. Dessa maneira, é natural que a informática esteja intrinsecamente ligada à nossa vida. Entretanto, esse fato não é observado em relação à educação, já que segundo dados do TIC Educação 2016, apenas 51% dos professores da rede pública utilizam computadores e/ou a internet de forma didática em sala de aula (CETIC, 2016).

O pensamento da teoria construtivista, segundo Piaget, descreve que o aprendizado é construído pelo indivíduo ao observar e interagir com o mundo tangível, não sendo assim o conhecimento inato ou um mero objeto que possa ser passado diretamente do professor para o aluno (FERREIRA, 1998). Esse pensamento serviu de base e inspiração para diversos projetos educacionais, como o movimento Explorable Explanations (Explicações Exploráveis), que visa apre-

1 Acadêmico do Curso de Ciência da Computação (UFSC).joaocampo2102@gmail.com

2 Docente Uniasselvi. Mestre em Ciências da Computação (UFSCAR), Especialista em Desenvolvimento Web e Mobile (Faculdade Sant’ana), Bacharel em Informática (UEPG). Balneário Camboriú, SC. rodrigo.nogueira@uniasselvi.com.br

sentar ao alunos conhecimento de uma maneira diferente e interativa, por meio de explicações audiovisuais e jogos (BRET, 2011).

Neste contexto, este artigo apresenta o OOB, sigla derivada da expressão em inglês *out of the box*, uma aplicação mobile na qual o conhecimento é adquirido de uma maneira baseada no construtivismo. Dentro do ambiente o aprendizado é algo contínuo e construído pelo usuário em seu próprio tempo de observação. O OOB foi construído visando atender alunos que tenham alguma dificuldade com métodos tradicionais (instrucionismo) e Papert propõe o construcionismo (proporcionar que o aluno seja seu próprio guia na aquisição do conhecimento) contribuindo assim, para uma maior aprendizagem a partir do mínimo de ensino por meio da descoberta.

O OOB foi construído utilizando da informação fornecida pelos usuários na forma de experiências dentro do aplicativo. Seymour Papert 'intercâmbio' entre o aprendizado e o artefato (computador), o que levou à aprendizagem autodirigida como forma de aprendizagem (incide sobre o contexto e as diferenças individuais), partindo desse princípio, as experiências são explicações curtas, entre 5 e 10 minutos de duração e conversam com o aluno em uma linguagem natural e interativa, introduzindo o assunto, explicando as aplicações práticas daquele conteúdo e ao final deixando vários caminhos para o aluno seguir e exercer sua criatividade.

Desse modo, este artigo tem como objetivo apresentar esse ambiente virtual de aprendizado baseado na metodologia construtivista, empregando conceitos de Explicações Exploráveis, bem como toda a arquitetura e tecnologias envolvidas no processo de desenvolvimento. O aplicativo engloba conteúdos de cinco áreas do conhecimento: física, química, matemática, biologia e programação de computadores. Durante toda a construção do pensamento são apresentadas pontes entre essas disciplinas, lugares onde o conhecimento de uma área se interliga com outra e aplicações práticas diferentes desse conteúdo podem ser apresentadas.

Referencial teórico

A metodologia construtivista pode ser aplicada nas mais vastas formas de ensino, entre elas uma que vem crescendo muito nos últimos anos, a divulgação científica em mídia audiovisual por meio da plataforma Youtube (VISBAL, 2018).

O canal alemão Kurzgesagt (palavra alemã para a expressão “*in a nutshell*”, “em uma casca de noz”, traduzindo literalmente) possui grande alcance, tendo atingido mais de 400 milhões de visualizações com o seu trabalho de explicar coisas, segundo eles mesmos, “utilizando nosso entusiasmo pelo universo”. O grupo conta com vídeos animados de cerca de 15 minutos, em que são abrangidos diversos temas (KURZGESAGT, 2018).

Bardini (2017), desenvolveu um aplicativo que prima pela ludicidade, utilizando conceitos muito próximos deste projeto, que, em seu estilo em formato de jogo. No decorrer do jogo, o usuário se depara com funções de lógica de programação e deve preencher lacunas com as opções que lhe são apresentadas. Empregando a filosofia construtivista e uma metodologia muito próxima à deste trabalho. Jesus *et al.* (2010) desenvolveu um aplicativo que propicia a construção do conhecimento musical ao possibilitar a exploração de materiais musicais diversos, incluindo gêneros e instrumentos musicais de várias culturas. Hoffmann (2009) utiliza do construtivismo para o ensino de lógica matemática através de uma aplicação desenvolvida para a Web.

Empregando a mesma filosofia construtivista deste trabalho, Santana (2009) desenvolveu um ambiente para simulação médica. O ambiente apoia o processo de ensino e aprendizagem de cursos construtivistas de medicina. O ambiente foi desenvolvido utilizando a linguagem Java, na plataforma Enterprise Edition (JEE), o padrão Model-ViewController (MVC), e os frameworks Java Server Faces (JSF) e Rich Faces.

O já citado movimento Explorable Explanations, proposto por Victor Bret, desenvolveu um ambiente on-line para a divulgação do movimento e o estímulo da criação de mais explicações exploráveis (BRET, 2011), em que os usuários podem interagir diretamente com os trabalhos publicados na Web.

Metodologia

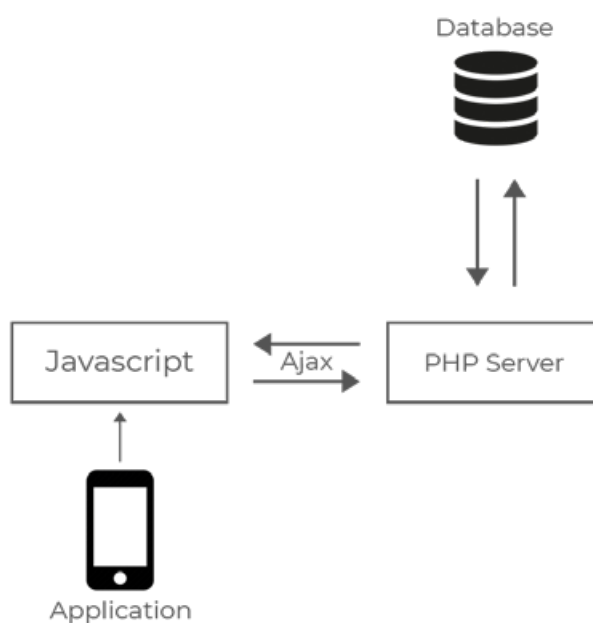
Este trabalho é categorizado como uma pesquisa aplicada e tecnológica (JÚNIOR *et al.*, 2010), tendo início com a elaboração do repertório conceitual, em que foi realizada pesquisa exploratória através artigos e livros, bem como uma pesquisa de viabilidade com alunos do ensino técnico integrado.

O processo de modelagem conceitual de software foi iniciado com o levantamento de requisitos e delimitação de escopo. Posteriormente, foi realizada a modelagem da base de dados do sistema e, por fim, o desenvolvimento dos protótipos da interface do aplicativo utilizando o framework Apache Cordova. O ambiente desenvolvido é uma aplicação híbrida, tipo que agrega as tecnologias de desenvolvimento e programação de aplicações nativas e Web com as linguagens de marcação de hipertexto HTML5 e folhas de estilo em cascata, transformando-as em um aplicativo mobile.

Além disso, para a utilização dos gráficos e animações propostas, fez-se o uso de uma framework de Javascript para manipulação do elemento Canvas, essa ferramenta é mantida pela Processing Foundation sobre o nome de p5.js. Referente ao conteúdo didático que alimenta o aplicativo, está embasado nos parâmetros curriculares nacionais para o ensino médio (BRASIL, 2000), além de se integrar com as mais diversas artes, na elaboração das animações visuais e trilhas sonoras.

Outra particularidade das aplicações híbridas são restrições quanto à arquitetura da aplicação, já que como todo o código será compilado e armazenado client-side, as requisições ao servidor serão feitas exclusivamente por AJAX (Asynchronous Javascript and XML) como visto na Figura 1.

Figura 1 – Arquitetura das requisições Ajax



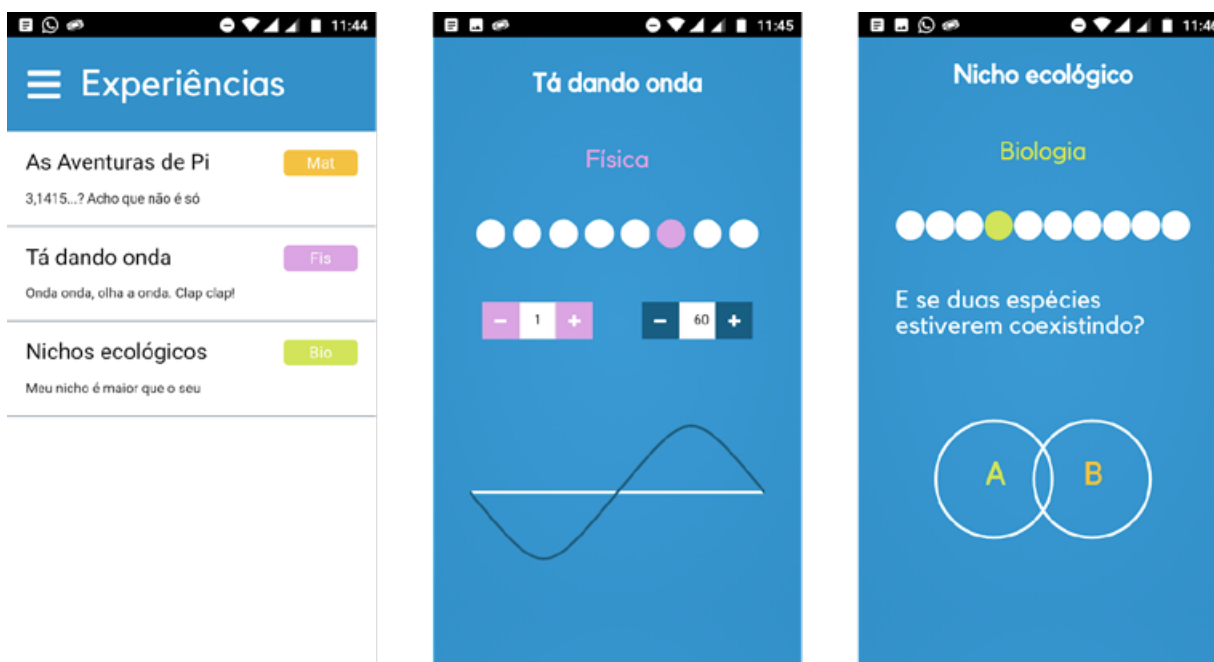
Fonte: os autores.

Resultados e discussão

As experiências são divididas entre as disciplinas e podem ser escolhidas livremente pelo usuário, ainda que seja recomendado uma ordem para a melhor absorção dos conteúdos. Cada experiência tem como objetivo se tornar uma abordagem integrada e imersiva para os usuários.

Aquelas experiências que integram mais de uma disciplina, alternam entre cores para situar o aluno. O esquema de cores foi uma das maiores preocupações durante o desenvolvimento da interface, já que ele ajudaria o usuário a entender quais das disciplinas estão em uso naquele momento em determinada experiência, isso se reflete em uma interface que, apesar de possuir cores vivas e distintas, ainda tem um visual limpo, como visto na Figura 2.

Figura 2 – Interface e exemplos de experiências



Fonte: os autores.

O ambiente integra as cinco disciplinas, mostrando ao usuário apenas aquelas que o mesmo tenha optado por utilizar no seu aprendizado. As experiências são divididas entre as disciplinas e podem ser escolhidas livremente pelo usuário, ainda que seja recomendado uma ordem para a melhor absorção dos conteúdos.

A Figura 3 mostra duas experiências desenvolvidas na tela da esquerda, que demonstra uma das experiências, intitulada “As aventuras de PI”, o padrão visto é formado ao se usar o dígito atual e o próximo dígito de π para gerar coordenadas cartesianas. No lado direito da figura mostra uma experiência intitulada “Tá dando onda”, que recebe o tamanho e propagação da onda retornando uma onda visualmente, bem como utiliza de recursos de áudio para trazer sua representação sonora conforme os valores são alterados.

Figura 3 – Experiência Tá dando onda e as Aventuras de π



Fonte: os autores.

Considerações finais

O presente trabalho demonstrou as tecnologias e metodologias necessárias para desenvolver um ambiente de aprendizado com bases na filosofia construtivista proposta por Jean Piaget. Neste sentido, utilizando-se da capacidade do próprio aluno de construir o seu conhecimento a partir do mundo a sua volta, bem como parte de seu desenvolvimento ainda em processo, acentuando principalmente a questão visual que uma aplicação como esta deve possuir.

Com a conclusão desse ambiente, cada vez mais metodologias diferentes de ensino sejam difundidas pela comunidade brasileira, descentralizando assim o ensino na proposta instrucionista que é o tradicional para nós. Também se espera fazer uma contribuição significativa para o movimento de explicações exploráveis, visto que o ambiente se encaixa nos moldes do mesmo. Como trabalhos futuros se tem a aplicação deste ambiente no ensino para estudantes do ensino médio e comparação com as tradicionais metodologias de ensino.

Referências

BARDINI, A. *et al.* Aplicativo Lúdico-Pedagógico para Ensino de Programação e Robótica Educacional. *In: Workshops do Congresso Brasileiro de Informática na Educação. [...] Anais.* 2017. p. 222.

BRET, V. **Explorable Explanations**. 2011. Disponível em: <<http://worrydream.com/ExplorableExplanations/>>. Acesso em: 29 out. 2022.

CETIC. **TIC Educação 2016**: Professores. 2016. Disponível em: <<http://cetic.br/tics/educacao/2016/professores/E3/>>. Acesso em: 11 nov. 2022.

HOFFMANN, D. S.; MARTINS, E. F.; BASSO, M. V. de A. Experiências física e lógico-matemática em Espaço e Forma: uma arquitetura pedagógica de uso integrado de recursos manipulativos digitais e não-digitais. *In: Brazilian Symposium on Computers in Education. [...] Anais.* 2009.

JESUS, E. A; de; URIARTE, M; Z.; RAABE, A. L. A. **Zorelha**: Um objeto de aprendizagem para auxiliar o Desenvolvimento da percepção musical em crianças de 4 a 6 anos. *Brazilian Journal of Computers in Education*, v. 18, n. 01, p. 91, 2010.

JONASSEN, D. **O uso das novas tecnologias na educação a distância e a aprendizagem construtivista**. 1996. Brasília. Disponível em:< <http://emaberto.inep.gov.br/index.php/emaberto/article/viewFile/2082/2051>>. Acesso em: 29 out. 2022.

JUNIOR, Vanderlei FREITAS et al. A pesquisa científica e tecnológica. *Espacios*, v. 35, n. 9, 2014.

Kurzgeasgt. Kurzgeasgt: in a nutshell. 2018. Disponível em:< <http://kurzgesagt.org/>>. Acesso em 29 de março de 2018.

SANTANA, Eduardo Felipe Zambom et al. Um Ambiente Baseado na WEB 2.0 para Atividades de Simulação na Educação Médica Construtivista. In: *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*. 2009.

USO DA TECNOLOGIA E INTELIGÊNCIA ARTIFICIAL NO AUXÍLIO DE APRENDIZADO DAS CRIANÇAS.

Use of Technology and Artificial Intelligence to aid Children's Learning.

Kennedy Anderson Panzenhagen¹

Pedro Sidnei Zanchett

RESUMO: Com o aprimoramento tecnológico, a educação vem ganhando cada vez mais força, gerando rico potencial pedagógico a fim de facilitar no aprendizado das pessoas em especial para crianças. O uso de recursos como robótica, inteligência artificial e a própria expansão dos aplicativos com novos métodos de ensino vem dinamizando ainda mais a forma de ensinar e aprender. Esta pesquisa tem como objetivo apresentar um exemplo prático da especificação de um chatbot com inteligência artificial utilizando IBM Watson que seja capaz de tirar dúvidas no ensino das crianças. Inicialmente se faz um estudo bibliográfico da área e posteriormente traz exemplos de chatbot para auxiliar professores. Conclui-se o artigo com a proposta de um chatbot para ensinar álgebra aos estudantes da matemática detalhando suas funcionalidades.

Palavras-chave: Educação. Inteligência artificial. Chatbot.

ABSTRACT: With technological improvement, education has been gaining more and more strength, generating a rich pedagogical potential in order to facilitate people's learning, especially for children. The use of resources such as robotics, artificial intelligence and the very expansion of applications with new teaching methods has been further streamlining the way of teaching and learning. This research aims to present a practical example of the specification of a chatbot with artificial intelligence using IBM Watson that is capable of answering questions in teaching children. Initially, a bibliographical study of the area is carried out and later it brings examples of chatbots to help teachers. The article concludes with the proposal of a chatbot to teach algebra to mathematics students, detailing its functionalities.

Keywords: Education. Artificial intelligence. Chatbot.

Introdução

Segundo a IBM (2020), a Inteligência Artificial (IA) é a capacidade das máquinas de realizarem tarefas que seriam normalmente denominadas para nós humanos. O caminho para chegar na IA iniciou a partir dos conceitos do cientista da computação Alan M. Turing em seu artigo denominado “Computing Machinery and Intelligence”, onde ele questionava, “As máquinas seriam capazes de pensar?”.

Para descobrir se uma máquina seria capaz de se comportar como um ser humano, Turing elaborou um método, o “Jogo da Imitação”, onde seu objetivo é que uma máquina e uma pessoa tentem confundir um interrogador, ao ponto que ele não saiba distinguir quem é a máquina. Em seu artigo, Turing previa que em cinquenta anos seria possível programar um computador para ganhar seu jogo. Apesar de não ter acertado em seu palpite, levou-se pouco mais de 50 anos para que uma máquina fosse capaz de derrotar um humano. De acordo com a Coventry University (2020), em 2014, Eugene Goostman, um chatbot russo, convenceu os juízes, fazendo com que eles não o identificassem como uma máquina, passando-se por uma criança ucraniana de 13 anos.

¹ Centro Universitário Leonardo Da Vinci – UNIASSELVI – Rodovia BR 470 – Km 71 – nº 1.040 – Bairro Benedito – Caixa Postal 191 – 89130-000 – Indaial/SC Fone (47) 3281-9000 – Fax (47) 3281-9090 – Site: www.uniassevi.com.br

Com o avanço dessa tecnologia, já se pode encontrar chatterbots voltados especificamente para a área educacional, como por exemplo da Professora Elektra, um chatterbot desenvolvido pela UFRGS. A Prof^ª. Elektra visou inicialmente responder perguntas sobre Física para alunos do ensino secundário que estivessem se preparando para o vestibular (THAISE, 2016).

Diferente dos chatterbots convencionais, os chatbots mais comuns, que apenas são capazes de solucionar questões pré-definidas, aqueles que possuem IA, são capazes de aprender com cada nova pergunta, além de várias outras ações que podem ser realizadas, como compra de passagens aéreas ou compra de produtos on-line.

Atualmente, lidamos diariamente com chatterbots e/ou com a IA, seja em um site de vendas ou um canal de ouvidoria. Por exemplo, o chatterbot com IA e que faz uso do IBM Watson é a BIA utilizada por aplicativos do banco Bradesco. Segundo Segura (2018) A BIA nasceu da parceria entre IBM e Bradesco em 2015, que estavam dispostos a iniciar um dos maiores e mais inovadores projetos de tecnologia cognitiva no mundo. Primeiramente, ela foi testada internamente pelos colaboradores de algumas filiais, mas com apenas alguns meses de treinamento ela já conseguia entender 100% das perguntas escritas e 83% das perguntas faladas, e após 10 meses o sistema já respondia 96% das perguntas corretamente. Atualmente, a BIA já atende clientes finais e aprende integrações.

De acordo com o IBGE (2010), “quase 46 milhões de brasileiros, cerca de 24% da população, declararam possuir pelo menos uma das deficiências investigadas (mental, motora, visual e auditiva)”. Diante disso, diversos projetos para inclusão foram desenvolvidos, um deles pelo Museu do Amanhã, em parceria com a IBM, denominado de IRIS+.

A nova experiência chamada IRIS+ também será acionada por esse cartão (da IRIS). Mas a diferença é que, ao final do percurso, ao lado da área NÓS, a IRIS ganhará voz – deficientes auditivos, ou quem preferir, poderão conversar por chat – e falará diretamente com o visitante (MUSEU DO AMANHÃ, 2017).

Este projeto não foi desenvolvido para somente responder os visitantes, mas também para formular perguntas e encorajar que o visitante pense sobre seu papel na sociedade e que ele aja para que tenhamos um amanhã mais consciente.

Com base no contexto atual, bem como existem diversas tecnologias voltadas para o auxílio a pessoas com deficiências, porém muito poucas voltadas para o aprendizado infantil, este trabalho irá executar uma pesquisa para que possam ser beneficiadas essas pessoas através de um aplicativo.

Objetivo geral

Especificação de um chatbot com inteligência artificial utilizando IBM Watson que seja capaz de tirar dúvidas no auxílio de aprendizado infantil.

Fundamentação teórica

Segundo o dicionário on-line Dicio (2018), inteligência significa: Faculdade de entender, de compreender e solucionar, já artificial segundo o mesmo, tem seu significado como: Algo que é produzido não pela natureza, mas sim por uma técnica, algo que age de maneira fingida, escondendo realmente suas reais intenções, com esses dois conceitos fixados podemos dizer que a inteligência artificial é algo criado para trabalhar sozinha, sem a presença de um humano, tendo o conhecimento sobre o conceito desses termos, o presente capítulo será apresentado da seguinte maneira:

Inteligência Artificial (IA)

Segundo Ortigara Bittencourt e Osório (2006, *apud* ORTIGARA, 2018), a IA pode ser dividida em três épocas:

- I. Clássica, sendo de 1956 até 1970, onde: Baseado em técnicas de busca heurística. Os sistemas gerais desenvolvidos nesta época obtiveram resultados interessantes.
- II. Romântica, sendo de 1970 até 1980, que: Houve uma crescente exigência de formalização matemática, foram incluídas decidibilidade, completude e complexibilidade.
- III. Moderna, sendo de 1980 até 1990, onde: Marcou a época do surgimento dos Sistemas Especialistas (SE's). Devido ao sucesso, empresários partiram em busca de obter algo comercializável.

Com a transformação digital pode-se notar que a IA está presente em nossas atividades rotineiras, como recomendações em sites de compras na internet ou em nosso aparelho de retirar pó que reconhece os objetos, e que muitos são os benefícios para quem desfruta desta tecnologia, tais como: redução de custo e mão de obra, maior eficiência e agilidade, pois não seria mais preciso que um ser humano interferisse sobre uma enorme massa de dados.

A inteligência artificial vem realizando coisas extraordinárias desde os tempos de sua aprimoração, e alguns estados já estão se adaptando a essas mudanças. Segundo Knapp (2011), no estado de Nevada, nos Estados Unidos, em 2012 aprovou uma lei que permitia carros sem motoristas (autônomos). De acordo com Bernardes (2019), cerca de 86% das empresas prestadoras de serviços de saúde, ciências biológicas e fornecedoras de tecnologia de saúde estão utilizando tecnologia da inteligência artificial, neste cenário de tecnologias podemos citar desde próteses robóticas, prontuários eletrônicos, diagnósticos gerados automaticamente por robôs, cirurgias robóticas até a criação de remédios. Assim como estes citados, Schwab (2016) também cita em seu livro outros benefícios da IA como:

1. Decisões racionais, orientadas por dados com menos vies.
2. Eliminação da “exuberância irracional”.
3. Reorganização das burocracias ultrapassadas,
4. Dependência energética.
5. Avanços na ciência médica, a erradicação de doenças.

Tendo em vista todos esses aspectos positivos, pode se deparar com a seguinte situação: “Com tanta tecnologia assim, poderia existir algum impacto negativo neste meio?”. Schwab (2016) também nos mostra em seu livro alguns pontos negativos da IA:

1. Prestação de contas, quem é o responsável, direitos fiduciários.
2. Hacking/Cibercrimes.
3. Responsabilidade e responsabilização, governança.
4. Aumento de desigualdade.
5. Ameaças existenciais para a humanidade.
6. Perdas de trabalhos.

A 4ª Revolução Industrial

A palavra revolução nos remete a algo que mudou radicalmente, que houve uma mudança abrupta, durante os tempos fomos apresentados a três grandes revoluções industriais que, segundo o Conselho Federal Administrativo (2019), a primeira de 1760 e 1840 caracterizada pelas máquinas a vapor e ferrovias, a segunda entre 1850 a 1945, e é caracterizada pela chegada da eletricidade e da linha de montagem, a terceira que vai de 1950 a 2010, foi neste

ponto da história que se caracterizou o grande avanço da tecnologia da informática, surgindo os computadores e os aparelhos móveis, a criação da internet, da robótica e da eletrônica.

Para entender o quão grande é a quarta revolução basta imaginar bilhões de pessoas conectadas em seus celulares, próteses feitas a partir de impressões 3D, a extrema abundância máquinas inteligentes realizando serviços que antes era realizados somente por pessoas, carros autônomos que levam e trazem pessoas a qualquer lugar, nanotecnologia e a computação quântica (SCHWAB, 2016).

Educação 4.0

Segundo Borsa (2007), a criança desde que nasce, ela já se torna um membro de um grupo social, pois as suas necessidades fisiológicas são diretamente dependentes de outras pessoas e estão designadas a serem saciadas em sociedade. Como Führ (2018) nos apresenta, a educação teve até o presente momento, quatro classificações.

- **Educação 1.0:** Onde que nesta fase o educador era a figura mais importante da organização e no trabalho da formação do acadêmico. Ele era o detentor do saber, e por isso, seus estudantes, com uma atitude de admiração recebiam seus ensinamentos. No princípio as escolas eram chamadas de Escolas Paroquiais, pois eram limitadas apenas a formação de eclesiásticos, as aulas aconteciam em igrejas e seus ensinamentos eram baseados em textos sagrados.

Durante vários séculos este tipo de educação prevaleceu, pois atendia às expectativas da sociedade da época, pois a educação era estritamente baseada em ensinamentos cristãos já que não queriam que as pessoas refletissem, pensassem ou tirassem suas próprias conclusões. Em resumo, a educação 1.0 era basicamente consistida no aprendizado da leitura, escrita, ensinamentos da bíblia, canto e um pouco de aritmética. Com o tempo foram incluídos os ensinamentos do Latim, gramática, retórica e dialética.

- **Educação 2.0:** Nesta fase as pessoas eram preparadas para trabalhar nas fábricas, ela teve forte influência da Revolução Industrial e apresenta as mesmas características observadas na produção industrial.

Tarefas repetitivas, mecânicas e trabalho individual. A sala de aula era vista como homogeneia e uma metodologia de ensino e aprendizagem que se caracterizava pela: padronização, concentração, centralização e sincronização. A educação passou a ter como objetivo o treinamento, alicerçado na aprendizagem informativa, à qual a memorização ficava evidenciada (FÜHR 2018, s.p.).

E mais uma vez, o conhecimento transmitido tinha como função educar o educando a sociedade e ao mercado de trabalho.

- **Educação 3.0:** Nesta fase da educação, o tutor careceria saber usar das novas tecnologias como potencial pedagógico. Consistia em uma nova concepção do que ensinar, de como ensinar e com o que ensinar. Neste novo tipo de educação unia as novas tecnologias com a aprendizagem, estimulando cada vez mais que o estudante desenvolvesse a autonomia, criatividade e a flexibilidade.
- **Educação 4.0:** Com a chegada da Quarta Revolução Industrial e da era digital, o conceito de educação nos apresenta um novo paradigma, em que a informação se encontra acessível a todos de forma geral, sem limites de tempo e espaço. O educador se torna um orquestrador das diversas informações, juntamente com o educando, onde ambos procuram organizar e sintetizar a informação, transformando-a em conhecimento.

Educação em tempos de pandemia

Em tempos de pandemia, onde as leis para o isolamento social já são uma realidade para os Brasileiros, e para todos no mundo, a educação a distância vem ganhando espaço e forças, sendo uma nova alternativa encontrada pelas instituições de ensino.

[...] em 2019, foram mais de 1,4 milhão de alunos no EAD, 52% do total de matriculados. Com a suspensão das aulas ocasionadas pelo isolamento social, o Ministério da Educação (MEC) autorizou o funcionamento dos cursos presenciais na educação a distância até que a situação seja normalizada no país (E+B EDUCAÇÃO; GABRIELE SILVA, 2020, s.p.)

Ainda segundo E+B Educação e Gabrile Silva, a educação infantil que costumava acontecer totalmente em regime presencial, foi a mais afetada.

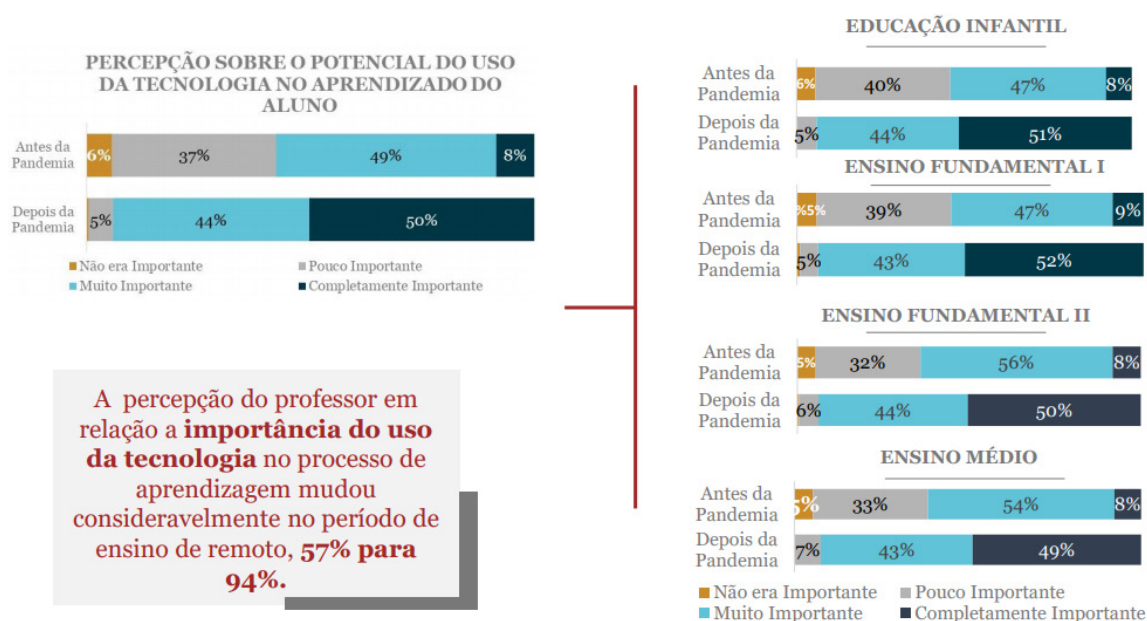
A educação infantil, etapa que costuma acontecer de forma totalmente presencial, tem sido a mais afetada. Os professores precisam realizar verdadeiros malabarismos para conseguir produzir as videoaulas e prender a atenção dos alunos (E+B EDUCAÇÃO; GABRIELE SILVA, 2020, s.p.)

Um estudo realizado este ano pelo Instituto Península, apontou que entre abril e maio, grande parte dos professores se sentiam despreparados ou que nunca havia dado aula de forma virtual.

Após a prática ter sido imposta pela pandemia e gestores de escolas públicas e privadas criarem formas para capacitar os docentes mesmo à distância, hoje, 49% afirmam que a falta de formação é um desafio para ensinar remotamente. Como consequência, 94% dos professores indicaram que agora enxergam a tecnologia como muito ou completamente importante no processo de aprendizagem dos alunos. Antes, apenas 57% tinham essa percepção (INSTITUTO PENINSULA, 2020)

Ainda segundo a pesquisa, a percepção dos professores, no que diz respeito ao índice “Completamente Importante” passou de apenas 8%, para 50%, maior depois da pandemia do COVID-19.

Figura 1 – Percepção sobre o uso da tecnologia na aprendizagem.



Fonte: Instituto Península (2020)

4 CHATBOTS

Segundo Cahn (2003 *apud* CAHN, 2010), um chatbot é um "sistema de diálogo human-computer online com linguagem natural. Sansonnet (SANSONNET, 2006 *apud* MAHN 2010) nos fornecem uma estrutura básica que mostra as funções esperadas dos chatbots da atualidade:

- **Agente Dialógico:** Deve compreender a entrada inserida pelo usuário, ou seja, fornecer função de compreensão. Bots recebem uma entrada de texto, que são analisadas com ferramentas de processamento de linguagem natural e usadas para gerar respostas apropriadas. (SANSONNET *et al.*, 2006 *apud* CAHN, 2010).
- **Agente Racional:** Deve ter acesso a uma base de conhecimento externa.
- **Agente Incorporado:** Deve fornecer a função de “Presença”. Essa função se mostra crucial no caso de usuários comuns. Até os primeiros robôs receberam nomes (ELIZA, ALICE, CHARLIE etc.) para satisfazer essa condição. Hoje, os desenvolvedores estão focados no uso de truques de linguagem para criar personas para chatbots, a fim de criar confiança com os usuários e dar a impressão de um agente incorporado. (SANSONNET *et al.*, 2006, *apud* CAHN, 2010).

No ano de 1996, Joseph Weizenbaun, criou o primeiro chatbot que mais se aproximou a imitar um ser humano. ELIZA. (CAHN, 2010). A partir da ELIZA, houve um significativo progresso no aperfeiçoamento dos chatbots, tornando-os mais inteligentes.

Em 1972, Kenneth Colby, em Stanford, desenvolveu o PARRY, um chatbot que personificava uma pessoa esquizofrênica e paranoica. (COLBY, 1981 *apud* CAHN, 2010). E em 1995, Richard Wallace desenvolveu o chatbot ALICE, este chatbot gerava as respostas por entradas de correspondências de padrão em comparação com os pares <padrão> (entrada) <modelo> (saída), que estavam armazenados em uma base de conhecimento. A escrita destes mesmo documentos é baseada em Artificial Intelligence Markup Language (AIML), uma extensão do XML, que ainda é usada até os tempos de hoje.

Cases de ChatBots para a Educação

A exemplo do chatbot professora Elektra que já foi mencionada anteriormente, existem diversos cases de chatbots voltados para a educação. É o caso do **Robô Ed**, onde integrantes do Programa Nacional de Racionalização do Uso dos Derivados do Petróleo e do Gás Natural (CONPET), juntamente com a equipe de desenvolvimento da Petrobras, lançaram um novo website, com o intuito de facilitar a comunicação com o público e otimizar o trabalho de conscientização da população (OLIVEIRA, 2004)

Figura 3 – Robô Ed



Fonte: CONPET (2020)

O gerente de Tecnologia e Marketing do CONPET, Frederico Marinho ressalta:

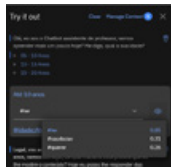
A idéia era atrair, principalmente, o público infanto-juvenil para o portal, com a novidade do robô Ed, o único com inteligência artificial funcionando via Internet no País. Ele pode responder em tempo real às perguntas da garotada e de demais internautas interessados em saber o que fazer para evitar o desperdício de combustíveis como o petróleo, gás de cozinha e outras fontes energéticas (MARINHO, 2004 *apud* OLIVEIRA, 2004).

Hoje em dia, ao acessarmos o site do CONPET, o próprio nos mostra que o Ed tirou “férias” desde 2016, ficando indisponível a partir desta data.

Professora Elektra: Segundo Lucchesi, (Lucchesi et al., 2018), a Elektra é um chatbot de bate-papo, com base na arquitetura A.L.I.C.E, projetada inicialmente para atender e interagir com os alunos do ensino secundário, com fins de apoiar a aprendizagem da Física.

Posteriormente, suas funcionalidades foram expandidas no intuito de prover atendimento contínuo aos alunos remotos, por meio de outras funcionalidades, como hiperlinks e múltiplas respostas para cada categoria (LUCCHESI *et al.*, 2018).

Figura 4 – Professora Elektra



Fonte: UFRGS (2020)

Katri: Um professor da UNA e do Uni-BH Leonardo Drummond Vilaça criou um chatbot chamado Katri, para ser usado como professor-assistente no mundo virtual, resultado de sua pesquisa de doutorado. (Lourenço, 2019). Segundo Leonardo:

Essa ideia surgiu porque trabalho como professor da UNA e do Uni-BH e também como líder de inovação acadêmica do Instituto Ânima. Ao pensar em como trazer inovações para a educação superior, estava discutindo com meu diretor e a nossa equipe sobre projetos que usassem inteligência artificial. No curso de relações internacionais, o professor Rafael Ávila já tinha experimentado criar um chatbot e eu entrevistei os alunos dele. Ouvi o que pensavam dessa experiência e também entendi como o professor criou o chatbot (VILAÇA, 2019 *apud* LOURENÇO, 2019).

IBM Watson

Segundo Ortigara (IBM, 2018 *apud* ORTIGARA, 2018), o IBM Watson pode ser definido como: “Conjunto de soluções técnicas de Inteligência Artificial, como Machine Learning (Aprendizado de Máquina, AM), Natural Language (Linguagem Natural, NT).”

A história do IBM Watson começa por volta dos anos de 1956, onde Arthur Samuel com a ajuda de um IBM 704, desenvolveu um programa baseado no jogo de Damas, onde o programa pudesse aprender com jogos anteriores, assim como também enfrentava oponentes humanos. (IBM BP Network, 2017). Após isso, em onze de maio de 1997, segundo Campbell (2017), um computador denominado Deep Blue, viera a derrotar o atual campeão mundial de xadrez, Garry Kasparov. A partida veio ao final de vários dias, durando seis jogos, com duas vitórias para a IBM, uma para o campeão Garry Kasparov e três empates. Em fevereiro de 2011, o computador da IBM, Watson, competiu contra os dois maiores campeões de todos os tempos em um programa de TV chamado Jeopardy.

Para que o Watson estivesse no mesmo nível dos competidores do Jeopardy, e que tivesse precisão em cada resposta por ele entregue, o Watson precisou analisar cada pergunta para determinar o que estava sendo solicitado e determinar o seu nível de confiança, para que isso seja possível durante o seu processo de analisar uma pergunta para determinar a melhor resposta, o Watson se utiliza de um processamento avançado de NT e captação e recuperação de informações.

Estudos de Caso da Tecnologia no Auxílio de Aprendizagem

Processamento de Linguagem Natural

Segundo Cahn (2017) a primeira parte da operação da aplicação é a interpretação das perguntas.

O Watson executa técnicas de processamento de linguagem natural para classificar o problema, incluindo análise superficial, análise profunda, identificação de funções semânticas, relações de conferência e reconhecimento de entidades nomeadas (FERRUCCI, 2010 *apud* CAHN, 2017).

Para que o Watson possa processar e analisar as sentenças de entrada, é realizado um processo de “tokenização” e segmentação destas entradas, que são seguidas de diversos processos de análises linguísticas, há exemplo de análises morfolexicais e sintáticas. Estes tokens são convertidos em frases para serem processados em sequência. Deste modo o Watson é capaz de dividir a frase em partes, rotular cada uma e entender o seu significado funcional e gramatical para o uso na geração das respostas.

Geração das Respostas

Logo após este processamento, o segundo passo é a geração das hipóteses, baseado no entendimento da pergunta, o Watson procura uma resposta em suas bases de conhecimento e gera todo o conteúdo possível.

O objetivo do Watson é obter 85% de recall dos 250 principais candidatos (por exemplo, para 85% das perguntas, a resposta correta estará contida em uma das 250 principais respostas de candidatos recuperadas) (FERRUCCI, 2010 *apud* CAHN 2017, p. 33).

Estão incluídas entre as técnicas de pesquisa realizadas pelo Watson: Pesquisa na base de conhecimento, mecanismos de pesquisa textual, pesquisas em documentos e também pesquisas de passagem que incluem várias consultas de pesquisa por pergunta.

Na pesquisa de documentos, o Watson extrai os títulos dos documentos que contêm o texto da consulta de pesquisa como hipóteses. Na pesquisa de passagem, técnicas de

reconhecimento de entidade nomeada são usadas em passagens com o texto da consulta de pesquisa e essas entidades nomeadas são geradas como hipóteses. A pesquisa de passagem e a pesquisa de documentos usam plataformas existentes, como Lucene e Indri para encontrar documentos relevantes que podem conter a resposta para a consulta de pesquisa (FERRUCCI, 2010 *apud* CAHN, 2017).

Ainda segundo Cahn (2017), o terceiro passo é a pontuação das hipóteses e evidências, no qual o sistema analisa cada entrada e aplica um algoritmo de classificação para determinar a resposta mais provável.

As funções de pontuação medem: “o grau de correspondência entre a estrutura predicado-argumento de uma passagem e a pergunta, a confiabilidade da fonte da passagem, a localização geoespacial, as relações temporais, a classificação taxonômica, as relações lexicais e semânticas nas quais o candidato é conhecido por participar, a correlação do candidato com termos da pergunta, sua popularidade (ou obscuridade), seus pseudônimos e assim por diante (FERRUCCI, 2010 *apud* CAHN 2017).

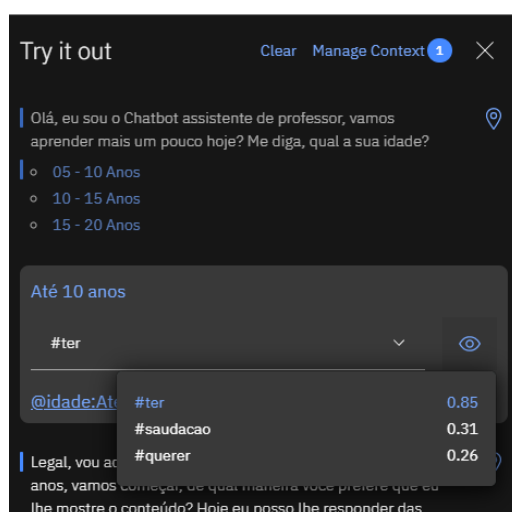
Incluído também no IBM Watson, há recursos reacionais, ou seja, o Watson foi desenvolvido com a capacidade de raciocínio temporal, ao modo que possa detectar inconsistências em datas, e também o raciocínio espacial, ao modo em que possa detectar inconsistências em geografias.

Todos os pontuadores mencionados acima são agregados em dimensões nas quais cada resposta é pontuada, como "Taxonômico, Geoespacial (local), Temporal, Confiabilidade da fonte, Gênero, Consistência do nome, Relacional, Suporte de passagem, Consistência da teoria e assim por diante (FERRUCCI, 2010 *apud* CAHN 2017).

E o quarto passo, Cahn (2017) nos mostra que é a classificação e a estimativa de confiança, onde o Watson se utiliza de uma abordagem de AM para realizar esta etapa. Ele usa as dimensões acima citadas como recursos, para treinar um classificador, criando “pesos” para cada dimensão.

O Watson treina classificadores separados para lidar com subconjuntos de recursos e, em seguida, usa um conjunto desses classificadores para gerar uma classificação geral (FERRUCCI, 2010 *apud* CAHN, 2017).

Figura 5 – Classificação das entidades do IBM Watson



Fonte: o autor.

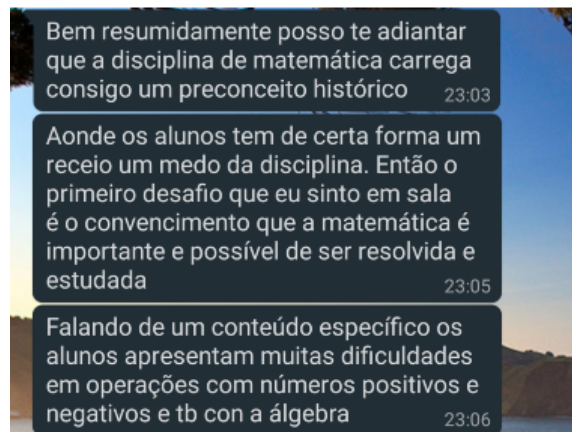
Exemplos práticos de casos no ensino de Matemática.

Neste tópico serão demonstrados dois exemplos práticos para a disciplina de matemática no ensino das crianças, objetivando apresentar a sua eficiência no ensino.

Especificações do case 1 - Ensino de Álgebra

Na sequência será mostrado como criar através do IBM Watson Assistant um exemplo passo-a-passo de uma atividade que possa ser aplicada dentro do contexto educacional. O exemplo será descrito para case de matemática no ensino sobre álgebra. Justifica se este case, conforme relato apresentado na figura abaixo.

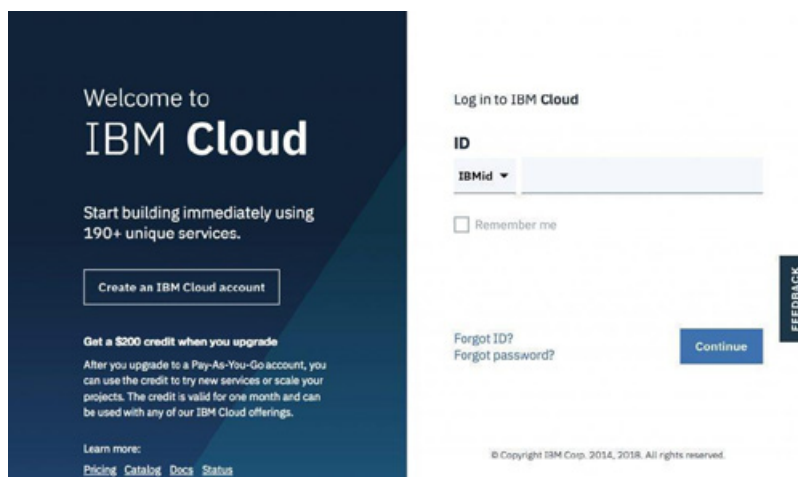
Figura 6 – Declaração de uma Professora de Matemática



Fonte: o autor.

Primeiramente deve-se criar uma conta na plataforma da IBM Cloud, para poder utilizar a tecnologia IBM Watson Assistant.

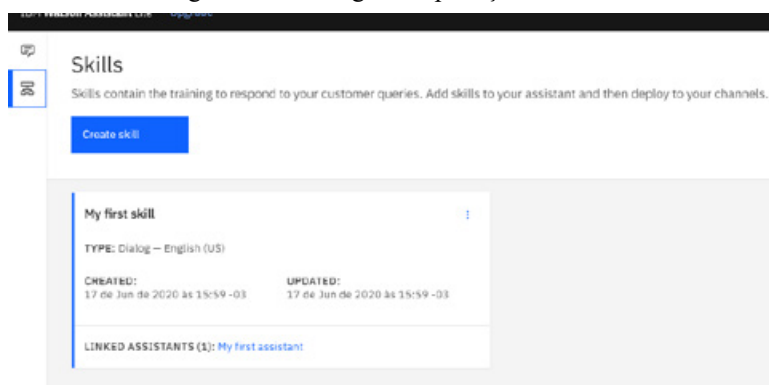
Figura 7 – Tela de Login IBM Cloud



Fonte: o autor.

Depois de criada a conta ou efetuado o login, vamos acessar a aba Catálogo e localizar o recurso IBM Watson Assistant.

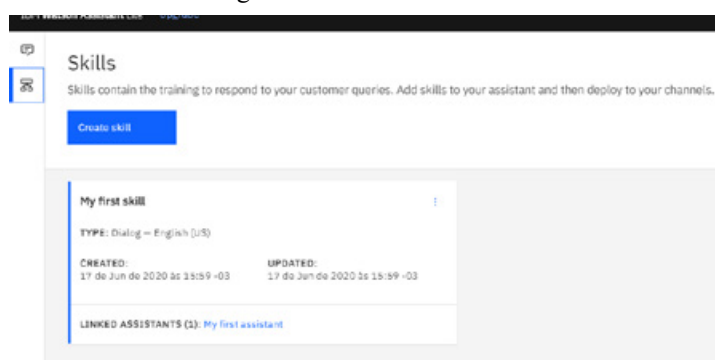
Figura 8 – Catálogo de Aplicações da IBM



Fonte: a autor.

Depois que selecionamos este recurso, a plataforma nos leva para a página de criação do recurso, onde ela vai nos solicitar a Região, o Plano, Nome e Grupos de Serviços, e opcionalmente as tags para identificação. Para este projeto, como será no plano Lite (Grátis), esse recurso é limitado a 100 nodos de diálogo. Depois de criado o nosso chatbot, iremos acessar a área de Skills para podermos começar a dar vida para nosso chatbot.

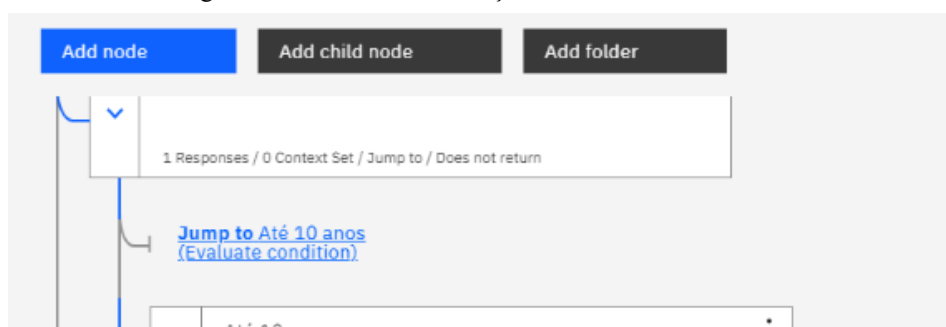
Figura 9 – Skills do ChatBot



Fonte: o autor.

Para que o IBM Watson possa interpretar as nossas frases ou perguntas, devemos adicionar entidades e intenções, possibilitando que a ferramenta se utiliza dos próprios recursos para compreender nossa escrita.

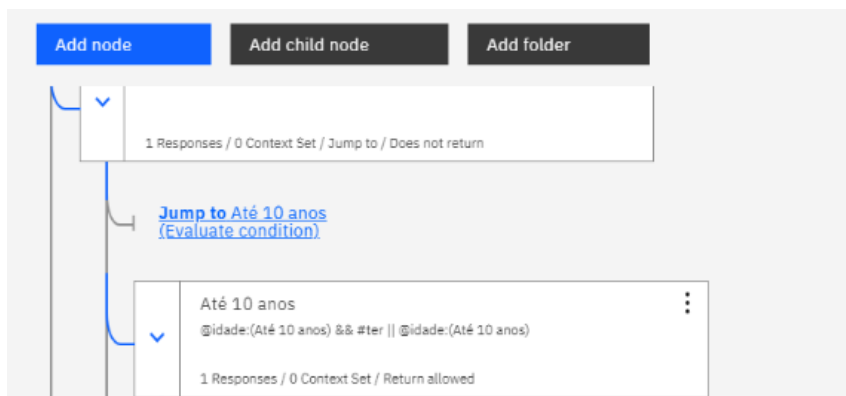
Figura 10 – Entidades e Intenções dentro do IBM Watson



Fonte: o autor.

Ao acessarmos o nodo “Dialog” vamos ter acesso a árvore de criação das respostas dos chatbots, por padrão existem duas condições comuns dentro do Watson, “Welcome” onde todo início de conversa será interpretado por esta condição e anything_else caso a entrada do usuário não condizer com nenhuma entidade ou intenção.

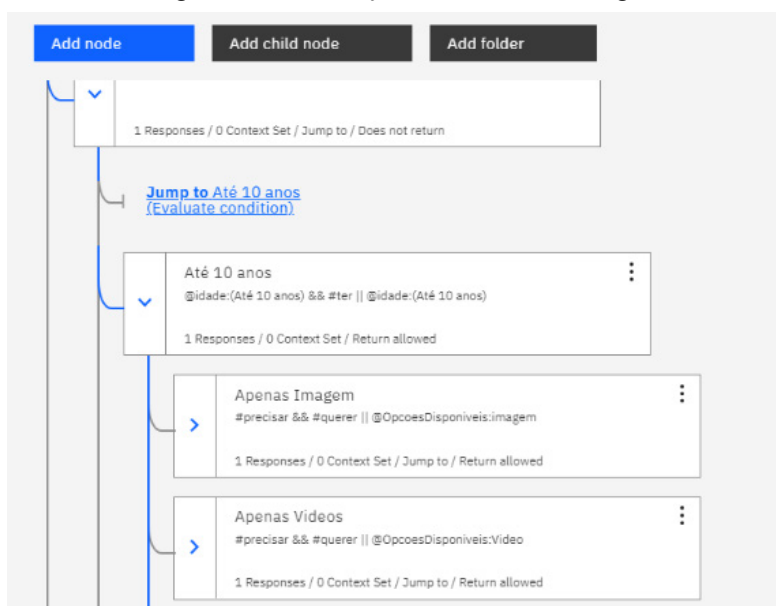
Figura 11 – Árvore de Diálogos do IBM Watson



Fonte: o autor.

E será dentro desta árvore que iremos construir o chatbot que irá auxiliar o estudo de quem solicitar. Segue mais detalhes nas imagens abaixo.

Figura 12 – Construção da Árvore de Diálogo



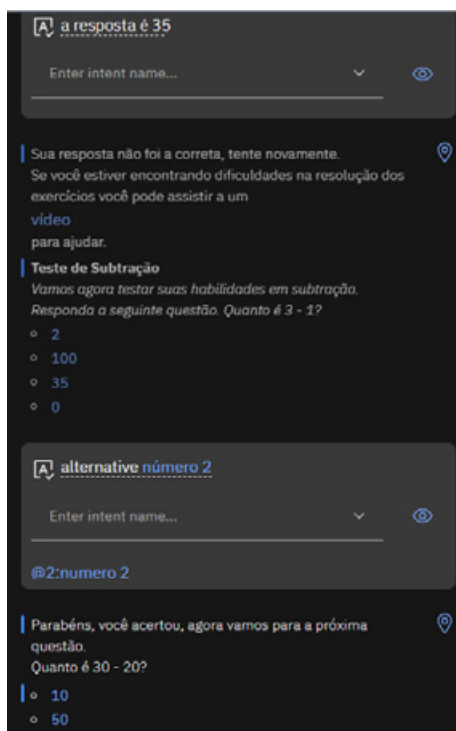
Fonte: o autor.

Neste artigo foi proposto somente resposta em texto sem edições, porém, a plataforma IBM Watson pode obter respostas de várias maneiras, por exemplo via Webhooks. As respostas também podem ser personalizadas, uma vez que a caixa de diálogo onde a resposta é adicionada também é um interpretador de HTML.

Agora será demonstrado como funcionará uma aplicação de um teste utilizando a plataforma.

Utilizando-se somente dos recursos gratuitos da plataforma, e aplicando um conhecimento básico da linguagem de marcação HTML, foi criado um teste de raciocínio, onde o professor define o texto de entrada, em seguida aplica a imagem e aguarda a entrada da resposta do aluno. A resposta foi definida manualmente, e o professor teve que inserir alguns modelos de possíveis respostas para que o chatbot compreendesse que a próxima entrada seria a resposta para a imagem, uma vez que ele não é aninhado e nem pré-definido para isso.

Figura 13 – Explicação do Watson para a resposta errada



Fonte: o autor.

Especificações do case 2 - Adição e Subtração

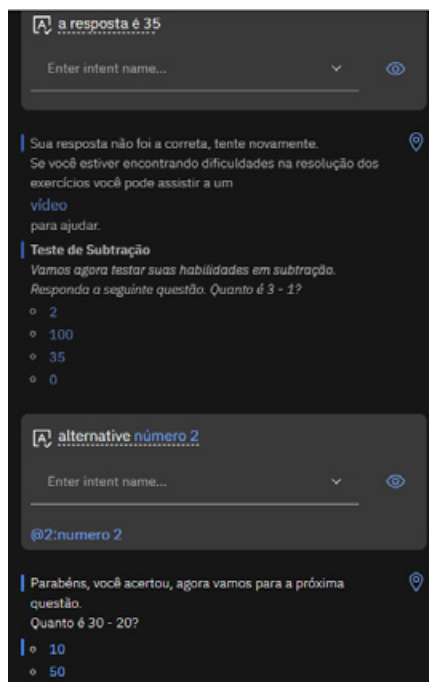
Agora, irei demonstrar outra maneira que a plataforma nos possibilita a aplicação de testes e provas, por meio das Opções. Elas permitem que o usuário, possa retornar a entrada para o assistente com base na opção escolhida dentre várias outras. (IBM CLOUD DOCS, 2020). Além de uma lista de opções, o assistente permite até outros dois tipos de escolha:

Seu app pode exibir as opções especificadas usando qualquer controle apropriado da interface com o usuário (por exemplo, um conjunto de botões ou uma lista suspensa). A propriedade `preferredControlType` opcional indica o tipo preferencial de controle que seu app deve usar (button ou dropdown), se suportado. Para uma experiência de usuário melhor, uma boa prática é apresentar três opções ou menos como botões e mais de três opções como uma lista suspensa (IBM CLOUD DOCS, 2020)

Na demonstração a seguir, foram adicionadas diversas opções para a pergunta realizada, e o aluno poderá escolher alguma opção disponível. Enquanto a resposta estiver errada, foi definido para que a pergunta seja refeita, ao acertar a pergunta, ele passa para a próxima

questão. Ao final, também mostramos ao aluno quantas ele perguntou e respondeu corretamente, e quantas não, disponibilizando também uma alternativa para as explicações.

Figura 14 – Demonstração da alternativa Opção dentro do IBM Watson



Fonte: o autor.

Como material de apoio referente a este case 2 que demonstra com mais um exemplo prático do chatbot no ensino da matemática. Existe uma vasta possibilidade de aprendizado através de vídeos no Youtube, como por exemplo o Título (Subtração para crianças - Aprendendo a subtrair com os dinossauros - Matemática para crianças).

Considerações Finais

Como mencionado acima, durante os tempos, a educação teve diversas fases, onde em algumas delas o professor era a entidade maior dentro da sala de aula, e o único detentor do saber, com o passar dos tempos este conceito foi mudando, nos tempos atuais o professor ainda continua sendo a entidade maior da sala, porém não mais o único detentor de todo o saber, nestes tempos atuais, existe uma interação aluno e professor muito maior, onde professor transfere conhecimento para o aluno e aluno também transfere conhecimento para o professor.

Com isso, estando presente em nossa vida de diversas maneiras, a inteligência artificial e os chatbots vem nos auxiliando e tornando atividades rotineiras e repetitivas obsoletas. Com o intuito de amenizar os problemas causados pelas dificuldades enfrentadas pelos estudantes, o estudo acima realizado, baseou-se em artigos e literaturas para propor o uso de tecnologias para auxiliar e complementar o tempo que uma criança passa estudando.

A alta taxa de aceitação da tecnologia para o ensino das crianças pelos professores em meio a pandemia da COVID-19 em 2020, a inteligência artificial fornecida pelo IBM Watson seria uma combinação perfeita de praticidade e comodidade, uma vez que o requisito mínimo para uso da plataforma seria ter acesso a internet, independente de que o aluno esteja em um aparelho celular ou em um computador de mesa. A junção do conhecimento do professor, com a facilidade de programar e implantar as perguntas e respostas na plataforma

Fundamentado pelos estudos e pela pesquisa, a proposta de utilizar inteligência artificial e chatbots no ensinamento é válida, e pode sim complementar o estudo das crianças, mas existem, muitos fatores que tornam esta realidade distante, ao menos no Brasil. Onde nem todas as pessoas possuem computadores ou aparelhos celulares, ou ao menos possuem acesso a internet. Quando evoluirmos neste aspecto, esta realidade não estará mais tão distante.

Referências

- BERNARDES, I. **Impactos da Inteligência Artificial na saúde**. 2019. Disponível em: <<https://www.med.puc-rio.br/notcias/2019/5/14/impactos-da-inteligencia-artificial-na-sade>>. Acesso em: 26 abr. 2020.
- BORSA, J. C. **O papel da escola no processo de socialização infantil**. 2007. Disponível em: <<https://www.psicologia.pt/artigos/textos/A0351.pdf>>. Acesso em: 19 maio 2020
- CAHN, J. **CHATBOT: Architecture, Design, & Development**. Disponível em: <http://www.academia.edu/download/57035006/CHATBOT_thesis_final.pdf>. Acesso em: 25 mai. 2020.
- CAMPBELL, M. **20 Years After Deep Blue, a New Era in Human-Machine Collaboration**. 2017. Disponível em: <<https://www.ibm.com/blogs/think/2017/05/deep-blue/>>. Acesso em: 03 maio 2020.
- COVENTRY UNIVERSITY. **Turing test transcripts reveal how chatbot 'Eugene' duped the judges**. 2015. Disponível em: <<https://www.coventry.ac.uk/primary-news/turing-test-transcripts-reveal-how-chatbot-eugene-duped-the-judges/>>. Acesso em: 10 mar. 2020.
- CRUZ, E. P. **Brasil tem 24,3 milhões de crianças e adolescentes que usam internet**. 2019. Disponível em: <<https://agenciabrasil.ebc.com.br/geral/noticia/2019-09/brasil-tem-243-milhoes-de-criancas-e-adolescentes-utilizando-internet>>. Acesso em: 3 maio 2020.
- DICIO. **Dicionário Online de Português**. Disponível em: <<https://www.dicio.com.br/inteligencia/>>. Acesso em: 19 maio 2020.
- E+B EDUCAÇÃO. SILVA, G. **Covid-19: importância da Educação a Distância durante a pandemia**. 2020. Disponível em: <<https://www.educamaisbrasil.com.br/educacao/dicas/covid19-importancia-da-educacao-a-distancia-durante-a-pandemia>>. Acesso em: 1 out. 2020.
- FREY, C. B.; OSBORNE, M. **The future of employment: how susceptible are jobs to computerisation?** University of Oxford. 2013. Disponível em: <<https://www.oxfordmartin.ox.ac.uk/publications/the-future-of-employment/>>. Acesso em: 21 abr. 2020.
- FÜHR, R. C. **EDUCAÇÃO 4.0 E SEUS IMPACTOS NO SÉCULO XXI**. CONEDU. Disponível em: <https://www.editorarealize.com.br/revistas/conedu/trabalhos/TRABALHO_EV117_MD4_SA19_ID5295_31082018230201.pdf>. Acesso em: 19 mai. 2020.
- IBGE. **Número de católicos cai e aumenta o de evangélicos, espíritas e sem religião**. IBGE. 2010. Disponível em: <<https://censo2010.ibge.gov.br/noticias-censo?id=3&idnoticia=2170&view=noticia>>. Acesso em: 22 mar. 2020.

IBM BP Network. **Interactive Timeline: See the Past and Explore the Future with IBM.** Disponível em: <<https://www.ibmnetwork.com/blog/history-of-ibm>>. Acesso em: 3 maio 2020.

IBM Cloud Docs. **Implementando Respostas.** Disponível em: <<https://cloud.ibm.com/docs/assistant?topic=assistant-api-dialog-responses>>. Acesso em: 11 out 2020.

IBM. **Get started with artificial intelligence, Developer IBM.** Disponível em <<https://developer.ibm.com/technologies/artificial-intelligence/gettingstarted/>>. Acesso em: 11 mar. 2020.

IBM. **Museu do Amanhã comemora dois anos com lançamento de experiência em Inteligência Artificial.** Disponível em:< <https://www.ibm.com/blogs/ibm-comunica/museu-do-amanha/>>. Acesso em: 22 mar. 2020.

INSTITUTO PENINSULA. **Sentimento e percepção dos professores brasileiros nos diferentes estágios do coronavírus no Brasil.** Instituto Península, Disponível em: <https://institutopeninsula.org.br/wp-content/uploads/2020/08/Sentimentos_-fase-3.pdf>. Acesso em: 05 out. 2020.

KNAPP, A. **Nevada Passes Law Authorizing Driverless Cars.** Forbes. 2011. Disponível em: <<https://www.forbes.com/sites/alexknapp/2011/06/22/nevada-passes-law-authorizing-driverless-cars/#51dcf6e81332>>. Acesso em: 26 abr. 2020.

LOURENÇO, A. **Professor cria assistente no mundo virtual.** Belo Horizonte, 29 jul. 2019. Disponível em: <https://www.em.com.br/app/noticia/especiais/educacao/2019/07/29/inter-nas_educacao,1073214/chatbot-educacional.shtml>. Acesso em: 25 mai. 2020.

LUCCHESI, I. L. *et al.* **Avaliação de um chatbot no contexto educacional: um relato de experiência com metis.** Novas Tecnologias na Educação, Porto Alegre, v. 16, n. 1, dez., 2018. Disponível em: <<https://www.seer.ufrgs.br/renote/article/view/85903/49294>>. Acesso em: 25 mai. 2020.

MUSEU DO AMANHÃ. **Iris+ | Uma nova experiência no Museu do Amanhã.** Disponível em: <<https://museudoamanha.org.br/pt-br/iris-mais>>. Acesso em: 22 abr. 2020.

OLIVEIRA, E. O. **Robô Virtual ensina a Economizar.** Disponível em: <<https://in.bot/imprensa/jornaldocomercio/2004-10-25.html>>. Acesso em 25 mai. 2020.

SCHWAB, Klaus Martin, **A Quarta Revolução Industrial**, 1º Edição, Edipro, 2016.

SEGURA, Mauro, **Bradesco é estrela da campanha mundial da IBM**, Mauro Segura. Disponível em: <<http://www.maurosegura.com.br/bradesco-e-estrela-da-campanha-mundial-da-ibm/>>. Acesso em: 17 mar. 2020.

THAISE, D. **Conheça a Profª Elektra, um chatbot desenvolvido pela UFRGS.** Disponível em: <<https://medium.com/fred-s-a/conhe%C3%A7a-a-prof%C2%AA-elektra-um-chatbot-desenvolvido-pela-ufrgs-aaf8a4fc412>>. Acesso em: 20 abr. 2020.

VIEIRA, B. B. **Timidez e exclusão-inclusão escolar: um estudo sobre identidade.** Disponível em: <<https://tede2.pucsp.br/handle/handle/15973>>. Acesso em: 3 maio 2020.

DESENVOLVIMENTO DE UM SISTEMA DE DETECÇÃO AUTOMATIZADA DE FAKE NEWS UTILIZANDO MACHINE LEARNING

Development of an automated fake news detection system using machine learning

Roger Oliveira Monteiro¹

Rodrigo Ramos Nogueira²

Resumo: Com o avanço da Internet, a facilidade e a velocidade no compartilhamento de notícias, o problema da disseminação de fake news aflige a sociedade como um todo, afetando cada vez mais o nosso cotidiano. Tendo em vista os problemas causados pela desinformação, este trabalho tem como objetivo o estudo e análise dos métodos de machine learning para desenvolver um mecanismo de coleta de dados de forma inteligente a partir de datasets de notícias e a implementação de algoritmos como filtros. Por fim, foi desenvolvido um sistema que permite a classificação de notícias em verdadeiras e fake news.

Palavras Chave: Notícias Falsas. Checagem. Mineração de Dados. Aprendizado de Máquina.

Abstract: With the advancement of the Internet, the ease and speed of sharing news, the problem of the spread of fake news afflicts society as a whole, increasingly affecting our daily lives. In view of the problems caused by disinformation, this work aims to study and analyze machine learning methods to develop an intelligent data collection mechanism from news datasets and the implementation of algorithms such as filters. Finally, a system was developed that allows the classification of news into true and fake news.

Keywords: Fake News. Checking. Data Mining. Machine Learning.

Introdução

Desde o início da Web, o volume de dados que estão nos repositórios na rede mundial tem crescido de forma exponencial, atualmente são cerca de 200 milhões de sites ativos na Internet, dos quais, apenas a rede social Twitter gera, em média, 500 milhões de postagens por dia. Tal explosão de dados, levou a um estudo do IDC (Institute Data Corporation) que estima que até 2020 serão gerados 44 zettabytes de dados em todo mundo (IDC, 2012).

Nos diferentes nichos de redes sociais que surgiram, observou-se maneiras diferentes de redigir críticas, propiciadas pelas características das aplicações. Sites específicos, como especializados em críticas de filmes, permitem que usuários escrevam textos relativamente longos. Os microblogs, por outro lado, impõem limites na quantidade de caracteres das mensagens e não são ambientes exclusivamente destinados para publicação de críticas. No processo de descoberta e pesquisa que prosseguiu nas redes sociais, surgiu a necessidade de expressar opiniões de forma mais direta (VON LOCHTER, 2015).

Segundo Nogueira (2018), os sites de notícias são o terceiro maior veículo de informação mais acessado da Internet, perdendo apenas para aplicativos de mensagens e redes sociais. Esta informação reflete a importância do uso de sites de notícias e seu impacto no cotidiano das pessoas.

Juntamente com a importância de textos de notícias e seu compartilhamento das mesmas em redes sociais, vem a ascensão e disseminação das fake news. Desde meados de 2017, a

1 Acadêmico Egresso do curso de Análise e Desenvolvimento de Sistemas. roger.omonteiro@gmail.com

2 Docente Uniasselvi. Mestre em Ciências da Computação (UFSCAR), Especialista em Desenvolvimento Web e Mobile (Faculdade Sant'ana), Bacharel em Informática (UEPG). Balneário Camboriú, SC. rodrigo.nogueira@uniasselvi.com.br

quantidade de eventos e debates acerca deste fenômeno que vem sendo chamado de fake news cresceu de forma. Fake news pode ser definida como artigos de notícias que são intencional e verificadamente falsos e podem enganar os leitores. Em nessa definição de fake news inclui artigos de notícias fabricados intencionalmente, como um artigo amplamente compartilhado do agora extinto site denvergardian.com com a manchete “FBI agent suspected in Hillary email leaks found dead in apparent murder-suicide” (Agente do FBI suspeito de vazamento de e-mail de Hillary encontrado morto em aparente assassinato-suicídio) (DELMAZO, 2017).

Dado seu destaque, tem sido realizadas diversas discussões multidisciplinares sobre o tema. Almejando contribuir com tais pesquisas, este trabalho tem como objetivo acoplar à etapa de ETL (Extract, Transform, Load) de um Data Warehouse de Notícias o enriquecimento semântico através de classificação do tipo de notícias: real ou falsa.

Referencial teórico

No que se refere à notícias falsas e a aplicação de *Machine Learning*, GRUPPI *et al.* (2018) construíram um dataset com notícias, em português e inglês, tendo por objetivo construir um classificador para predizer se a fonte da notícia é ou não confiável. Utilizando um algoritmo de SVM com um kernel linear, foi possível estabelecer as características mais importantes, bem como sua classificação. Como resultado, o algoritmo de classificação obteve acurácia de 85% para os datasets brasileiros e 72% para datasets Americanos. Em uma contribuição para a área de classificação de notícias, Monteiro *et al.* (2018) utilizam o dataset Fake.br com o objetivo de avaliar os principais métodos de pré-processamento de textos para avaliar o desempenho do método SVM. Os melhores resultados foram obtidos com a combinação de bag-of-words com sentimentos, bem como o uso de todos os atributos, ambos com acurácia de 90%.

Marumo (2018) coletou notícias de sites com notícias verídicas e sites com notícias falsa e/ou de cunho satírico, com o objetivo de encontrar o melhor método para detecção de fake news. Como parte do pré processamento dos dados, utilizou-se o framework Gensim para remoção de caracteres não alfabéticos, a substituição de espaçamentos e quebra de linhas para espaços únicos, remoção de palavras com menos de 3 caracteres e a conversão de letras maiúsculas para minúsculas. Também foi utilizado o framework keras para tokenização dos dados. Com a aplicação dos algoritmos de classificação LSTM e SVM, conseguiu-se uma acurácia acima de 90%.

No que se refere ao enriquecimento semântico em ambientes de Data Warehouse através do emprego de técnicas de *Machine Learning*, é o caso Mansman (2014), que obteve um modelo multidimensional da rede social Twitter e desenvolveu um ambiente de Data Warehouse que permitiu a criação de um cubo de dados, bem como a análise de sentimentos. <Removed for double-blind review>, em uma abordagem similar, desenvolveu um ambiente de Data Warehouse que coleta notícias em inglês em tempo real, no qual após avaliação regressão logística, Naïve Bayes, SVM e Perceptron tiveram resultados próximos, dos quais o este último foi utilizado para realizar o enriquecimento semântico na etapa de ETL.

Overfitting constitui-se um grande problema em se tratando de base de dados textuais. Sendo assim, FENG, *et. al.* (2017), utilizaram o algoritmo AdaBoost, conhecido por obter grande sucesso para redução de overfitting em detecção de faces, reconhecimento de caracteres (OCR) e classificação de veículos. Em seus experimentos, foram utilizados datasets de 20 grupos de notícias, dataset Reuters, que consiste em 22 arquivos com um total de 21,758 documentos, e um dataset da BioMed, o qual é dividido em 10 tópicos, cada um contendo entre 1966 e 5022 artigos. Os resultados foram uma média de 86% de acurácia no algoritmo AdaBoost (Bonzaiboost).

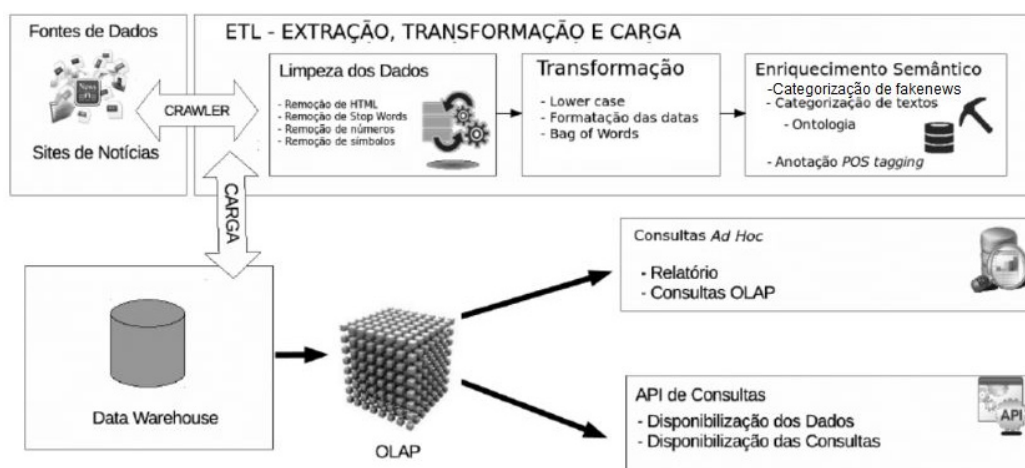
Metodologia

A primeira etapa deste projeto foi dedicada ao levantamento bibliográfico (PIZZANI et al, 2012), onde através de artigos e livros se obteve a fundamentação teórica e estado da arte, foi de suma importância para se obter os melhores métodos de aprendizado de máquina empregados durante os experimentos.

Em uma segunda etapa, foi desenvolvido um script de coleta e análise de notícias, permitindo com que esta pesquisa também se enquadre como pesquisa tecnológica de acordo com (JUNIOR et al. ,2014), pois o produto final é conjunto de arquitetura, software, complementado de um conjunto de dados.

No que se refere à base de dados, após pesquisa bibliográfica sobre dados com fake news, pode se verificar que existem poucos recursos disponíveis no idioma Português do Brasil, no qual o dataset mais utilizado é o Fake.br (MONTEIRO *et al.*, 2018). Tendo como em vista complementar este conjunto de dados e obter melhores resultados este trabalho também se propõe a coletar dados de notícias. A metodologia de desenvolvimento prático deste trabalho é baseada na arquitetura proposta por Nogueira (2018), na qual o classificador gerado será acoplado a etapa de ETL de um Data Warehouse, gerando o enriquecimento semântico em uma nova dimensão.

Figura 1 – Arquitetura utilizada



Fonte: adaptada de Nogueira (2018).

Para realizar os experimentos foi desenvolvido um web crawler, utilizando a linguagem python, juntamente com a biblioteca beautifulsoup para a coleta inicial dos dados. Foi construído um dataset composto por 1744 títulos e corpo de notícias falsas coletadas dos sites <boatos.org> e <g1.globo.com/fato-ou-fake>, e 3185 títulos e corpo de notícias verdadeiras coletadas do site brasil.elpais.com. Inicialmente será efetuado testes utilizando apenas os títulos das notícias, posteriormente o corpo juntamente com título e fazer um comparativo entre ambos. Para isso, serão utilizados os algoritmos de aprendizado de máquina (Machine Learning), Regressão Logística, AdaBoost, Naive Bayes e SVM (KOSALA, 2000).

A partir da criação de um sistema de coleta, com um algoritmo acoplado à etapa de ETL, este irá automaticamente classificar os dados coletados, aumentando assim a acurácia do classificador, e gerando uma base maior de dados para futuros trabalhos de combate a fake news. Também foi construído uma interface Web, onde o usuário será capaz de submeter um link e verificar se este é ou não uma notícia verdadeira, servindo este como protótipo antes de ser submetido a etapa de ETL (sendo esta, o propósito geral deste trabalho).

Resultados e discussão

Após a aplicação dos algoritmos Regressão Logística (Logistic Regression), AdaBoost, Naive Bayes e SVM (kernel linear), os mesmos obtiveram a acurácia de 88,85%, 81,37%, 86,22% e 87,45% respectivamente, no modelo de testes. Como técnica de avaliação dos modelos empregados, foi utilizado a validação cruzada com o método k-fold = 10.

Novamente o dataset foi dividido entre treino e teste, juntando agora os títulos ao corpo das notícias. Receberam os mesmos tratamento acima citados, obtendo a acurácia de 90,88%, 84,23%, 91,19% e 91,16% nos algoritmos Regressão Logística (Logistic Regression), AdaBoost, Naive Bayes e SVM respectivamente. A aplicação do método de validação cruzada, revelou um overfitting em alguns casos.

Por fim, o dataset foi dividido para utilização apenas dos corpos das notícias. Foram empregados os mesmos métodos utilizados anteriormente em relação ao tratamento e limpeza dos dados. A aplicação dos algoritmos resultou em 90,88%, 94,23%, 91,19% e 91,16% de acurácia nos algoritmos Regressão Logística (Logistic Regression), AdaBoost, Naive Bayes e SVM respectivamente.

Tabela 1 – Comparativo entre os *datasets* em relação à acurácia e validação cruzada.

| | Regressão Logística | AdaBoost | Naive Bayes | SVM (kernel Linear) |
|----------------|---------------------|----------|-------------|---------------------|
| Título | 88,85% | 81,37% | 86,22% | 87,45% |
| K-fold | 0,88 | 0,75 | 0,86 | 0,55 |
| Corpo | 97,40% | 95,12% | 97,80% | 98,62% |
| K-fold | 0,97 | 0,95 | 0,97 | 0,64 |
| Título + Corpo | 90,88% | 84,23% | 91,19% | 91,16% |
| K-fold | 0,90 | 0,84 | 0,91 | 0,54 |

Fonte: os autores.

A partir da análise de resultados, o método de Naive Bayes foi selecionado o melhor método, pelo fato de obter uma alta acurácia, complementado de ser um método de aprendizado incremental (on-line).

Posterior ao acoplamento foi desenvolvido a interface de classificação de fake news, mostrada pela Figura 2. E está disponível no servidor <https://detectorfakenews.herokuapp.com/>. A ferramenta espera como parâmetro o link de um site de notícia, e retorna se ele é ou não uma notícia falsa (fake news).

Figura 2 – Interface Web da Aplicação desenvolvida.



Fonte: os autores.

Considerações finais

O overfitting se constitui em um problema recorrente em bases textuais. Alguns algoritmos chegaram a resultados bastante relevantes, mas ao aplicarmos a validação cruzada com $k=10$, notou-se um grande *overfitting* em alguns casos. Sendo assim, observou-se que o algoritmo Naive Bayes obteve, além da alta acurácia, tolerância ao overfitting.

Para futuros trabalhos, tem-se como objetivo avaliar outras características técnicas de pré-processamento, aumentar a base de treino, aplicar os novos resultados a interface web, e posteriormente, o acoplamento a ETL do Data Warehouse.

Referências

DELMAZO, C.; VALENTE, J. CL. Fake news nas redes sociais online: propagação e reações à desinformação em busca de cliques. *Media & Jornalismo*, v. 18, n. 32, p. 155-169, 2018.

IDC. GANTZ, J., & REINSEL, D. The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. IDC iView: IDC Analyze the future, 2007(2012), p. 1-16.

KOSALA, R.; BLOCKEEL, H. Web mining research: A survey. *ACM Sigkdd Explorations Newsletter*, v. 2, n. 1, p. 1-15, 2000.

JUNIOR, V.; F. et al. A pesquisa científica e tecnológica. *Espacios*, v. 35, n. 9, 2014.

MONTEIRO, R. A. et al. Contributions to the Study of Fake News in Portuguese: New Corpus and Automatic Detection Results. In: *International Conference on Computational Processing of the Portuguese Language*. [...] Anais. Springer, Cham, 2018. p. 324-334.

NOGUEIRA, R. R. O Poder do Data Warehouse em Aplicações ed Machine Learning: Newsminer: Um Data Warehouse Baseado em Textos de Notícias. São Paulo: Nea, 2018.

VON LOCHTER, J. et al. Máquinas de classificação para detectar polaridade de mensagens de texto em redes sociais. 2015.

BANCO DE DADOS PARA ORGANIZAÇÕES

Database for organizations

Juliana Farias Pires¹

Cristiano Domingues Rodrigues¹

Telmo Muniz Pacheco Junior¹

Resumo: O presente artigo abordará o tema banco de dados para as organizações, a fim de apresentar o funcionamento dele dentro de uma organização. O banco de dados armazena, com segurança, dados, incluindo documentos, endereços, nomes de clientes, informações, o que é de suma importância para a empresa que busca êxito a partir de tomadas de decisões, e, quando da implementação desse banco de dados, a organização garante eficácia na gestão.

Palavras-chave: Banco de dados. Empresa. Organização. Dados.

Abstract: This article will address the topic Database for Organizations, presenting its functioning within the organization. The database securely stores data, such as documents, addresses, customer names, information that is of paramount importance for the company that seeks success in its decision-making, and from the implementation of the database, the organization guarantees efficiency in its management.

Keywords: Database. Company. Organization. Data.

Introdução

O presente artigo tem a finalidade de apresentar a importância do banco de dados dentro de uma organização, a fim de buscar, de forma objetiva, identificar os tipos de informações armazenados no interior dele, com os benefícios e os problemas trazidos para a realidade de uma organização.

O banco de dados é considerado um conjunto de elementos integrados entre si, os quais se relacionam de forma lógica. Consolida registros armazenados de forma separada, em arquivos, fornecidos para aplicações em um banco, independentemente dos programas de aplicação utilizada e de dispositivos já armazenados.

Funcionamento de um banco de dados

Um banco de dados agrupa dados de um mesmo assunto, a fim de armazenar, assim, diversos dados, como documentos, endereços, serviços, clientes, entre outros. Dessa forma, serve para que esses dados sejam armazenados e consumidos posteriormente.

Fundamentação teórica

O banco de dados é de suma importância para uma empresa que busca organização de dados, segurança de informação e uma rápida busca em um domínio específico. É muito comum que a empresa precise desses dados organizados para uma tomada de decisões.

A importância do banco de dados, para uma empresa, é que ele organiza e armazena as informações em determinado domínio.

¹ Centro Universitário Leonardo da Vinci – UNIASSELVI – Rodovia BR 470 - Km 71 - no 1.040 – Bairro Benedito – Caixa Postal 191 – 89130-000 – Indaial/SC Fone (47) 3281-9000 – Fax (47) 3281-9090 – Site: www.uniasselvi.com.br

Conforme Santos (*apud* MACHADO; ABREU, 1996, p. 14), “a modelagem de dados (ou de informações) está baseada no princípio de que, comprovadamente, os dados são estáveis no decorrer da vida de uma empresa, ou organização, não tendo volatilidade dependente de fatores pessoais, governamentais e temporais. Já os procedimentos possuem essa característica de volatilidade, pois sofrem constantes alterações [...] por fatores pessoais, quando existe troca de pessoas e métodos; por decisões governamentais e legislativas; por fatores de calamidade; e outros, externos às atividades normais da empresa”.

Para Laudon e Laudon (1999), os dados são independentes dos programas-aplicativos, pois o SGBD tem a capacidade de distinguir as visões físicas e lógicas, de modo que os programas possam utilizar os dados de um banco de dados compartilhado.

Com relação à redundância e à inconsistência, são reduzidas. Como os dados são independentes dos programas-aplicativos, não há a necessidade de construção de arquivos isolados, nos quais os mesmos elementos de dados são repetidos a cada vez que uma nova aplicação é chamada. Os dados são mantidos em, apenas, um lugar (LAUDON; LAUDON, 1999).

Conforme Laudon e Laudon (1999), o acesso a informações e o modo de uso são mais fáceis, pois o banco de dados estabelece relacionamentos com diferentes registros e aplicações que podem ser acessados e combinados. Qualquer fragmento do banco de dados pode ser acessado, desde a direção descendente, com o início do segmento e o prosseguimento para camadas subordinadas. Ainda, a organização dos arquivos e dos bancos de dados provoca um profundo efeito na forma através da qual as informações podem ser fornecidas, assim, deve, o projeto de um banco de dados, ser considerado com muito cuidado. Uma solução de sistema de informação precisa incluir os projetos lógico e físico desse banco.

Depois que o projeto lógico está pronto, é transposto para um banco de dados físico, para a forma através da qual os dados são, realmente, arranjados e armazenados no meio de armazenamento do computador.

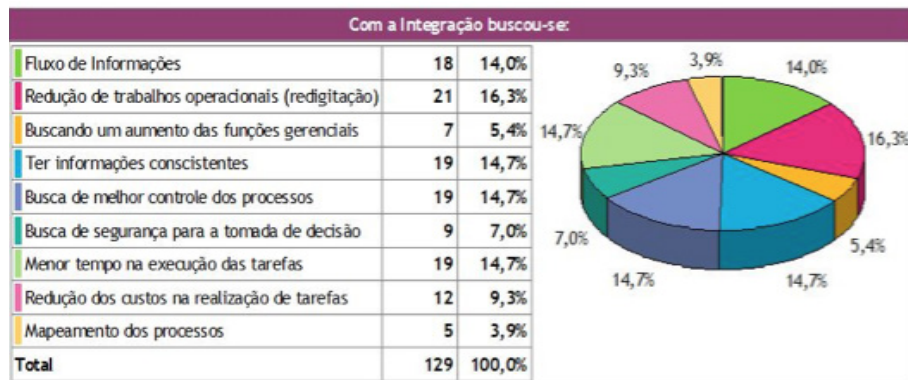
A meta do projeto físico é arranjar os dados de uma forma que torne a atualização e a recuperação rápidas e eficientes. Os padrões de acesso aos dados; e a frequência de utilização destes, pelos especialistas da empresa, são considerações importantes para tal projeto físico (LAUDON; LAUDON, 1999).

O conceito de banco de dados se distingue entre visões físicas e lógicas de dados. Pela visão lógica, tais dados são apresentados através da percepção por usuários finais, ou por especialistas da empresa. A visão física mostra como são, de fato, organizados e estruturados no meio de armazenamento físico (LAUDON; LAUDON, 1999).

É possível que pessoas que não são profissionais da área de informática consigam montar um projeto de banco de dados para sistemas de aplicação, através de técnicas estruturadas, como modelagem conceitual de dados.

É necessário proteger os dados particulares de pessoas e de uma organização, além de realizar backup, para que tais dados sejam restaurados após alguma falha de hardware, ou de energia.

Gráfico 1. Integração.



Fonte: Os autores (2022).

No gráfico, trazem-se as características que as empresas buscaram ao integrar módulos com a contabilidade. Os resultados não foram muito dispersos. A maioria englobou a redução de trabalhos operacionais, como a redigitação, com 16,3%; seguida por um melhor controle de processos e por informações consistentes, com 14,7% cada.

Resultados e discussão

A utilização de técnicas, para a construção de SGBD, apresenta-se como uma alternativa eficiente, em função do reaproveitamento do código das funcionalidades comuns, o que possibilita implementações específicas para diversos ambientes operacionais, com diferentes seguimentos de regra.

Um novo tipo de estrutura de banco de dados vem ganhando popularidade e mostrando resultados operacionais de recuperação de objetos, como documentos, de forma que possa ser capaz de gerenciar dados por vídeo, som, imagens e texto em um único registro, aplicados em multimídia, o que possibilita a atualização. Essa estrutura é denominada de banco de dados orientados a objetos, como C++, Java ou Delphi.

A proposta deste trabalho foi a de identificar os tipos de informações possíveis de serem encontrados armazenados em um determinado banco de dados, de modo que possa demonstrar benefícios e problemas trazidos para a realidade de uma determinada organização.

A manutenção dos bancos de dados, de registros e de objetos precisa ser realizada da forma mais rápida, e, constantemente, adicionada ao acesso desses dados, para que possam ser produzidas informações em resposta a pedidos de usuários finais.

Atualmente, os novos desafios exigem, dos antigos profissionais, uma postura, para que possam assumir novos padrões de comportamento perante a evolução da tecnologia e interagir com a transformação do mundo.

Referências

LAUDON, K. C.; LAUDON, J. P. **Sistemas de informação: organizando as informações: arquivos e bancos de dados.** 4. ed. Rio de Janeiro: J.C. Editora, 1999.

MACHADO, F. N. R.; ABREU, M. P. de. **Projeto de banco de dados: uma visão prática.** São Paulo: Érica, 1996.