

**Ernane Rosa Martins
(Organizador)**

A Abrangência da Ciência da Computação na Atualidade

Ernane Rosa Martins

(Organizador)

A Abrangência da Ciência da Computação na Atualidade

Atena Editora
2019

2019 by Atena Editora
Copyright © Atena Editora
Copyright do Texto © 2019 Os Autores
Copyright da Edição © 2019 Atena Editora
Editora Executiva: Prof^a Dr^a Antonella Carvalho de Oliveira
Diagramação: Karine de Lima
Edição de Arte: Lorena Prestes
Revisão: Os Autores

O conteúdo dos artigos e seus dados em sua forma, correção e confiabilidade são de responsabilidade exclusiva dos autores. Permitido o download da obra e o compartilhamento desde que sejam atribuídos créditos aos autores, mas sem a possibilidade de alterá-la de nenhuma forma ou utilizá-la para fins comerciais.

Conselho Editorial

Ciências Humanas e Sociais Aplicadas

Prof. Dr. Álvaro Augusto de Borba Barreto – Universidade Federal de Pelotas
Prof. Dr. Antonio Carlos Frasson – Universidade Tecnológica Federal do Paraná
Prof. Dr. Antonio Isidro-Filho – Universidade de Brasília
Prof. Dr. Constantino Ribeiro de Oliveira Junior – Universidade Estadual de Ponta Grossa
Prof^a Dr^a Cristina Gaio – Universidade de Lisboa
Prof. Dr. Deyvison de Lima Oliveira – Universidade Federal de Rondônia
Prof. Dr. Gilmei Fleck – Universidade Estadual do Oeste do Paraná
Prof^a Dr^a Ivone Goulart Lopes – Istituto Internazionele delle Figlie de Maria Ausiliatrice
Prof. Dr. Julio Candido de Meirelles Junior – Universidade Federal Fluminense
Prof^a Dr^a Lina Maria Gonçalves – Universidade Federal do Tocantins
Prof^a Dr^a Natiéli Piovesan – Instituto Federal do Rio Grande do Norte
Prof^a Dr^a Paola Andressa Scortegagna – Universidade Estadual de Ponta Grossa
Prof. Dr. Urandi João Rodrigues Junior – Universidade Federal do Oeste do Pará
Prof^a Dr^a Vanessa Bordin Viera – Universidade Federal de Campina Grande
Prof. Dr. Willian Douglas Guilherme – Universidade Federal do Tocantins

Ciências Agrárias e Multidisciplinar

Prof. Dr. Alan Mario Zuffo – Universidade Federal de Mato Grosso do Sul
Prof. Dr. Alexandre Igor Azevedo Pereira – Instituto Federal Goiano
Prof^a Dr^a Daiane Garabeli Trojan – Universidade Norte do Paraná
Prof. Dr. Darllan Collins da Cunha e Silva – Universidade Estadual Paulista
Prof. Dr. Fábio Steiner – Universidade Estadual de Mato Grosso do Sul
Prof^a Dr^a Girlene Santos de Souza – Universidade Federal do Recôncavo da Bahia
Prof. Dr. Jorge González Aguilera – Universidade Federal de Mato Grosso do Sul
Prof. Dr. Ronilson Freitas de Souza – Universidade do Estado do Pará
Prof. Dr. Valdemar Antonio Paffaro Junior – Universidade Federal de Alfenas

Ciências Biológicas e da Saúde

Prof. Dr. Benedito Rodrigues da Silva Neto – Universidade Federal de Goiás
Prof.^a Dr.^a Elane Schwinden Prudêncio – Universidade Federal de Santa Catarina
Prof. Dr. Gianfábio Pimentel Franco – Universidade Federal de Santa Maria
Prof. Dr. José Max Barbosa de Oliveira Junior – Universidade Federal do Oeste do Pará

Profª Drª Natiéli Piovesan – Instituto Federal do Rio Grande do Norte
Profª Drª Raissa Rachel Salustriano da Silva Matos – Universidade Federal do Maranhão
Profª Drª Vanessa Lima Gonçalves – Universidade Estadual de Ponta Grossa
Profª Drª Vanessa Bordin Viera – Universidade Federal de Campina Grande

Ciências Exatas e da Terra e Engenharias

Prof. Dr. Adélio Alcino Sampaio Castro Machado – Universidade do Porto
Prof. Dr. Eloi Rufato Junior – Universidade Tecnológica Federal do Paraná
Prof. Dr. Fabrício Menezes Ramos – Instituto Federal do Pará
Profª Drª Natiéli Piovesan – Instituto Federal do Rio Grande do Norte
Prof. Dr. Takeshy Tachizawa – Faculdade de Campo Limpo Paulista

Conselho Técnico Científico

Prof. Msc. Abrãao Carvalho Nogueira – Universidade Federal do Espírito Santo
Prof. Dr. Adaylson Wagner Sousa de Vasconcelos – Ordem dos Advogados do Brasil/Seccional Paraíba
Prof. Msc. André Flávio Gonçalves Silva – Universidade Federal do Maranhão
Prof.ª Drª Andreza Lopes – Instituto de Pesquisa e Desenvolvimento Acadêmico
Prof. Msc. Carlos Antônio dos Santos – Universidade Federal Rural do Rio de Janeiro
Prof. Msc. Daniel da Silva Miranda – Universidade Federal do Pará
Prof. Msc. Eliel Constantino da Silva – Universidade Estadual Paulista
Prof.ª Msc. Jaqueline Oliveira Rezende – Universidade Federal de Uberlândia
Prof. Msc. Leonardo Tullio – Universidade Estadual de Ponta Grossa
Prof.ª Msc. Renata Luciane Polsaque Young Blood – UniSecal
Prof. Dr. Welleson Feitosa Gazel – Universidade Paulista

Dados Internacionais de Catalogação na Publicação (CIP) (eDOC BRASIL, Belo Horizonte/MG)	
A161	A abrangência da ciência da computação na atualidade [recurso eletrônico] / Organizador Ernane Rosa Martins. – Ponta Grossa, PR: Atena Editora, 2019. Formato: PDF Requisitos de sistema: Adobe Acrobat Reader Modo de acesso: World Wide Web Inclui bibliografia ISBN 978-85-7247-488-7 DOI 10.22533/at.ed.887190908 1. Computação – Pesquisa – Brasil. I. Martins, Ernane Rosa. CDD 004
Elaborado por Maurício Amormino Júnior – CRB6/2422	

Atena Editora
Ponta Grossa – Paraná - Brasil
www.atenaeditora.com.br
contato@atenaeditora.com.br

APRESENTAÇÃO

A área da Ciência da Computação apresenta atualmente uma constante ascensão, seus profissionais estão sendo cada vez mais valorizados e requisitados pelas empresas, tornando-a mais importante, prestigiada e reconhecida. As empresas de todos os portes e setores necessitam de profissionais qualificados desta área, que apresentem potencial para promover inovação, desenvolvimento e eficiência.

A Ciência da Computação é uma área com amplas possibilidades de atuação, como por exemplo: a elaboração de programas e softwares, o gerenciamento de informações, a atuação acadêmica, a programação de aplicativos mobile ou ainda de forma autônoma. A abrangência da Ciência da Computação exige de seus profissionais conhecimentos diversos, tais como: novos idiomas, pensamento criativo, capacidade de comunicação e de negociação, além da necessidade de uma constante atualização de seus conhecimentos.

Dentro deste contexto, este livro aborda diversos assuntos importantes para os profissionais e estudantes desta área, tais como: API de localização da google, identificação de etiquetas RFID, ferramentas para recuperação de dados, ensino de computação, realidade virtual, interação humano computador, gestão do conhecimento, computação vestível, gerência de projetos, big data, mineração de dados, Internet das coisas, monitoramento do consumo de dados na Internet, pensamento computacional, análise de sentimentos, filtros ópticos, rede óptica elástica translúcida, algoritmo de roteamento, algoritmo de atribuição espectral, algoritmo de utilização de regeneradores e algoritmo genético.

Assim, certamente que os trabalhos apresentados nesta obra exemplificam um pouco a abrangência da área de Ciência da Computação na atualidade, permitindo aos leitores analisar e discutir os relevantes assuntos abordados. A cada autor, nossos agradecimentos por contribuir com esta obra, e aos leitores, desejo uma excelente leitura, repleta de boas reflexões.

Ernane Rosa Martins

SUMÁRIO

CAPÍTULO 1	1
UMA ABORDAGEM SOBRE SISTEMA DE LOCALIZAÇÃO MOBILE	
Paulo Roberto Barbosa	
DOI 10.22533/at.ed.8871909081	
CAPÍTULO 2	6
UMA ABORDAGEM BIDINÂMICA PARA A IDENTIFICAÇÃO DE ETIQUETAS RFID	
Shalton Viana dos Santos	
Paulo André da S. Gonçalves	
DOI 10.22533/at.ed.8871909082	
CAPÍTULO 3	23
TESTE DE FERRAMENTAS DE RECUPERAÇÃO DE IMAGENS PARA SISTEMAS DE ARQUIVOS EXT3 E EXT4	
Diego Vinícius Natividade	
DOI 10.22533/at.ed.8871909083	
CAPÍTULO 4	34
REDIMENSIONAMENTO DO ENSINO DA COMPUTAÇÃO NA EDUCAÇÃO BÁSICA: O PENSAMENTO COMPUTACIONAL, O UNIVERSO E A CULTURA DIGITAL	
Melquisedec Sampaio Leite	
Sônia Regina Fortes da Silva	
DOI 10.22533/at.ed.8871909084	
CAPÍTULO 5	47
REALIDADE VIRTUAL, UTILIZANDO DAS MELHORES PRÁTICAS DA INTERAÇÃO HUMANO COMPUTADOR	
Bruno Moreira Batista	
Guiliano Rangel Alves	
Hellen Corrêa da Silva	
Rhogério Correia de Souza Araújo	
DOI 10.22533/at.ed.8871909085	
CAPÍTULO 6	52
ORGANIZAÇÃO DO CONHECIMENTO PARA A MEMÓRIA EMPRESARIAL: UM RELATO TÉCNICO SOBRE A EXPERIÊNCIA DO SEBRAE/RJ	
Leandro Pacheco de Melo	
DOI 10.22533/at.ed.8871909086	
CAPÍTULO 7	65
GERÊNCIA DE PROJETOS EM COMPUTAÇÃO VESTÍVEL: DIRETRIZES PARA O DESENVOLVIMENTO DE PRODUTOS VESTÍVEIS INTELIGENTES	
Renan Gomes Barreto	
Lucas Oliveira Costa Aversari	
Renata Gomes Barreto	
Gabriela Ferreira Marinho Barreto	
DOI 10.22533/at.ed.8871909087	

CAPÍTULO 8	76
EXPLORING <i>BIG DATA</i> CONTENT AND INFORMATION METRICS: INTERSECTIONS AND ANALYSIS TO SUPPORT DECISION-MAKING	
Rafael Barcellos Gomes Vânia Lisboa da Silveira Guedes	
DOI 10.22533/at.ed.8871909088	
CAPÍTULO 9	92
DEMOCHAIN - FRAMEWORK DESTINADO A CRIAÇÃO DE REDES BLOCKCHAIN HÍBRIDAS PARA DISPOSITIVOS IOT	
Lorenzo W. Freitas Carlos Oberdan Rolim	
DOI 10.22533/at.ed.8871909089	
CAPÍTULO 10	107
CONSUMO DO TRÁFEGO DE DADOS EM APLICAÇÕES DE VÍDEO SOB DEMANDA- YOUTUBE E NETFLIX	
Patricia Emilly Nóbrega da Silva Éwerton Rômulo Silva Castro	
DOI 10.22533/at.ed.88719090810	
CAPÍTULO 11	112
COMPUTAÇÃO NA ESCOLA: ABORDAGEM DESPLUGADA NA EDUCAÇÃO BÁSICA	
Christian Puhmann Brackmann Marcos Román-González Rafael Marimon Boucinha Dante Augusto Couto Barone Ana Casali Flávia Pereira da Silva	
DOI 10.22533/at.ed.88719090811	
CAPÍTULO 12	128
COLETA DE DADOS E ANÁLISE DE SENTIMENTOS NAS REDE SOCIAIS ON LINE	
Maurilio Alves Martins da Costa Bruna Emidia de Assis Almeida Fraga	
DOI 10.22533/at.ed.88719090812	
CAPÍTULO 13	137
ANÁLISE DO IMPACTO DO CASCATEAMENTO DE FILTROS ÓPTICOS EM UM CENÁRIO DE REDES ÓPTICAS ELÁSTICAS	
Gabriela Sobreira Dias de Carvalho William Silva dos Santos Lucas Oliveira de Figueiredo Helder Alves Pereira	
DOI 10.22533/at.ed.88719090813	

CAPÍTULO 14	143
ANÁLISE DE REDE ÓPTICA ELÁSTICA TRANSLÚCIDA CONSIDERANDO DIFERENTES ALGORITMOS DE ROTEAMENTO	
Arthur Hendricks Mendes de Oliveira	
William Silva dos Santos	
Helder Alves Pereira	
Raul Camelo de Andrade Almeida Júnior	
DOI 10.22533/at.ed.88719090814	
CAPÍTULO 15	149
ANÁLISE DE REDE ÓPTICA ELÁSTICA TRANSLÚCIDA CONSIDERANDO ALGORITMOS DE ATRIBUIÇÃO ESPECTRAL	
Arthur Hendricks Mendes de Oliveira	
William Silva dos Santos	
Helder Alves Pereira	
Raul Camelo de Andrade Almeida Júnior	
DOI 10.22533/at.ed.88719090815	
CAPÍTULO 16	155
A NEW MULTI OBJECTIVE APPROACH FOR OPTIMIZING P-MEDIAN MODELING IN SCHOOL ALLOCATION USING GENETIC ALGORITHM	
Clahildek Matos Xavier	
Marly Guimarães Fernandes Costa	
Cícero Ferreira Fernandes Costa Filho	
DOI 10.22533/at.ed.88719090816	
SOBRE O ORGANIZADOR	168
ÍNDICE REMISSIVO	169

UMA ABORDAGEM SOBRE SISTEMA DE LOCALIZAÇÃO MOBILE

Paulo Roberto Barbosa

Instituto Federal do Triangulo Mineiro
Uberlândia - MG

RESUMO: Com o advento do mundo mobile, ficou evidente o aprimoramento relacionado à localização referencial do dispositivo e conseqüentemente do indivíduo que está manipulando este. Os Serviços baseados em localização são fortemente utilizados na atualidade. Existem vários métodos e estudos na área de localização em que o desafio está no acerto da precisão entre o que é mostrado no dispositivo (mapa pré-definido) e a localização da pessoa/objeto. O trabalho em questão destaca a demonstração de como a API de localização da google trabalha com GPS, Cell-Id (rede de telefonia celular) e triangulação WIFI. Como resultado pode-se refletir sobre o estudo de localização em ambientes indoor (LPS – Local Positioning System), já que a dificuldade de chegada de sinal é mais comprometida.

PALAVRAS-CHAVE: posicionamento, dispositivo móvel, WI-FI, GPS, LPS.

AN APPROACH ON MOBILE LOCATION SYSTEM

ABSTRACT: With the advent of the mobile world, it was evident the improvement related

to the referential location of the device and consequently of the individual who is manipulating it. Location-based services are heavily used today. There are several methods and studies in the area of location where the challenge lies in the accuracy between what is shown in the device (predefined map) and the location of the person / object. The work in question highlights the demonstration of how Google's location API works with GPS, Cell-Id, and WIFI triangulation. As a result one can reflect on the study of localization in indoor environments (LPS), since the difficulty of arrival of signal is more compromised.

KEYWORDS: positioning, mobile device, WI-FI, GPS, LPS.

1 | INTRODUÇÃO

Sistemas de Posicionamento Indoor podem ser entendidos como ambientes fechados, tais como construções, em que se deseja analisar ou prover uma solução tecnológica utilizando um sistema de localização operando sobre um hardware em que se pode destacar a utilização de smartphones e tablets.

Zahid Farid, Rosdiadee Nordin e Mahamod Ismail (2013) definem um sistema de posicionamento indoor como qualquer sistema que prove uma posição dentro de uma estrutura

fechada como supermercado, aeroporto, hospital, metro ou campos universitários.

Quando se trabalha com sistema de localização indoor, é necessário se levar em conta várias situações inerentes a essa característica. Adena Schutzberg (2016) relata os seguintes pontos: GPS não funcionam bem em ambientes indoors; algumas soluções de posicionamento indoor trabalham com características similares ao GPS, outras soluções utilizam luz ou campo magnético para determinação da localização; RFID e sistemas inertes (os quais necessitam de uma localização inicial para funcionar, como exemplo podemos citar: acelerômetros, giroscópios, relógios de orientação e distância) trabalham de forma muito diferente; posicionamento indoor detecta a localização de uma pessoa ou objeto, mas nem sempre sua orientação ou direção; a melhor solução para sistema de posicionamento indoor e outdoor pode ser híbrida; várias são as demandas para a utilização de posicionamento indoor. Entre elas destacam-se hospitais, supermercados, prover ajuda em navegação, melhor fomentação de mercado para clientes, prover informações em tempo real via áudio para turistas, etc; os maiores “players” de tecnologia (Apple, Google e Microsoft) estão trabalhando em espaço indoor e por fim, posicionamento indoor requer mapas “indoor”.

Vários são os métodos e trabalhos relatados nessa área para melhor se chegar em uma precisão na localização. Cálculos matemáticos complexos são cada vez mais treinados e estudados a fim de garantir economia de bateria no dispositivo, além do ganho de performance e exatidão.

Cricket system é um exemplo de hardware criado exclusivamente para trabalho com localização indoor. Nissanka Bodhi (2005) define o Cricket system como sendo uma plataforma de hardware constituída de um transmissor de rádio frequência, um microcontrolador e outro hardware associado, gerando e recebendo sinal.

A Microsoft possui outro projeto relacionado a indoor positioning system denominado RADAR. Padmanabhan Venkat e Victor Bahl (2005) designam o RADAR como um sistema de posicionamento indoor que muito bem aproveita a força do sinal e pode ser utilizado para determinar a localização do usuário e máquina no interior de edifícios permitindo o uso de aplicações ditas interiores (location-aware).

Outra pesquisa relacionada a Wireless LAN indoor está pautada no Horus. Moustafa Youssef e Ashok Agrawala garantem que o Horus WLAN location satisfaz dois objetivos: alta precisão e baixos requisitos computacionais.

Por último, porém não menos importante, pode-se citar o projeto COMPASS que utiliza tecnologia WI-FI para triangulação, juntamente com a bússola, atualmente presente na maioria dos aparelhos. King Thomas, Stephan Kopf, Thomas Haenselmann, Christian Lubberger e Wolfgang Effelsberg mostraram na conclusão do trabalho que a orientação do usuário pode melhorar consideravelmente a precisão do sistema de posicionamento.

2 | MATERIAIS E MÉTODOS

Os recursos utilizados nessa análise são pautados em antenas WI-FI, Smartphone para teste e Eclipse como plataforma de análise além de foco principal na estratégia de localização com a utilização da classe Location Manager e interface LocationListener, disponibilizada pela google em sua plataforma Android. A vantagem na utilização dessa abordagem de utilização de uma API (application program Interface) está associada ao fato de que não é necessário o desenvolvedor saber detalhes de cálculos matemáticos complexos, simplesmente utilizando o recurso no processo de desenvolvimento. A desvantagem pode estar relacionada a decepção da precisão não ser a esperada, já que o método de captura de posição é encapsulado na API. O método é experimental com teste da API nas três possíveis abordagens (GPS, CellID e triangulação WI-FI) para que no futuro o trabalho seja evoluído para um protótipo de navegação indoor.

3 | RESULTADOS E DISCUSSÃO

Como ilustração voltada para a parte de sistemas de localização, será detalhado nessa sessão algumas partes interessantes disponibilizadas pela biblioteca do Android. Vale ressaltar que se trata apenas de um exemplo, pois não foram encontradas verificações de que a biblioteca trabalha de forma assertiva quando se trata de localização indoor, sem depender de GPS ou rede de celular. Conforme for a aplicação será necessário fazer uma melhor análise do algoritmo de localização com melhor detalhamento do método. A classe LocationManager possui uma característica que identifica o tipo de localização habilitada no aparelho. Através da atribuição (código em linguagem Java, utilizado pelo Android) da Figura 1, se consegue obter o tipo de localização. **GPS_PROVIDER** para GPS e **NETWORK_PROVIDER** para redes de celulares ou WI-FI.

```
// Acquire a reference to the system Location Manager  
LocationManager locationManager = (LocationManager) this.getSystemService(Context.LOCATION_SERVICE);
```

Figura 1 – Código mostrando instanciação da classe LocationManager

Com o serviço de localização ativo, conforme figura 2, conclui-se que o **GPS_PROVIDER** está sendo utilizado. Ao rodar aplicação exemplo, pode-se verificar a latitude, longitude e raio obtido pelo GPS, conforme Figura 3.

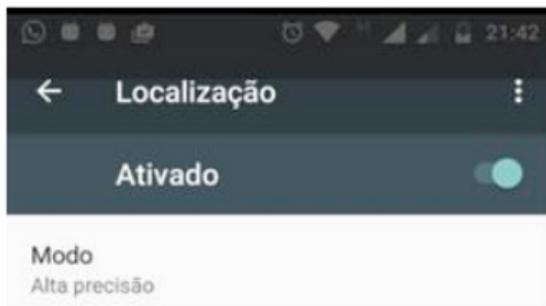


Figura 2 – GPS ativado

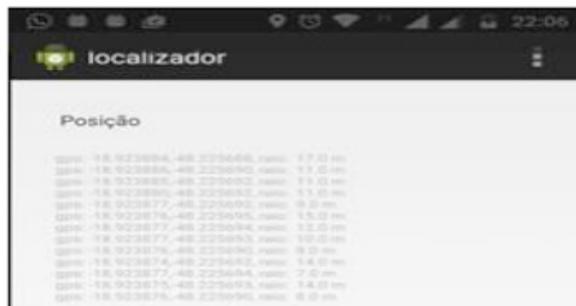


Figura 3 – Localização via GPS

Na mudança do serviço de localização (Figura 4), percebe-se que ainda há possibilidade de se utilizar o sinal via rede de telefonia (chamada de Cell-ID), Bluetooth ou WI-FI. Pela Figura 5 comprova-se o que o WI-FI está ativo.



Figura 4 – Ativação de localização

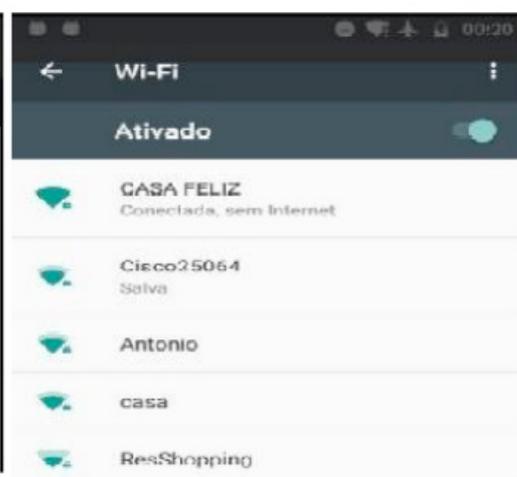


Figura 5 – Rede WI-FI ativa

via Cell-ID, Bluetooth ou WI-FI

Modo avião ativo (Figura 6), mostrando a possibilidade de se estar utilizando WI-FI e não rede de telefonia. Pela Figura 7 é possível concluir a possível triangulação (network e não GPS) via rede WI-FI.

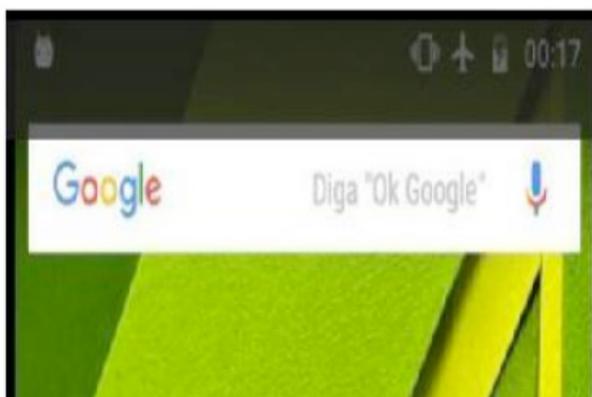


Figura 6 – Modo avião ativo

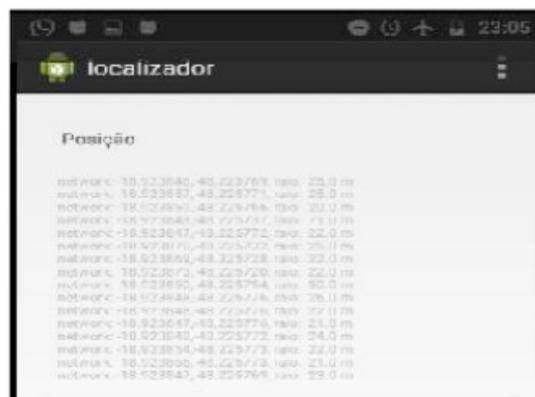


Figura 7 – Utilização de rede WI-FI

4 | CONSIDERAÇÕES FINAIS

Assim, é conclusivo que com ferramentas desse tipo é possível a realização de vários trabalhos que vão de encontro com necessidades de desenvolvimento de sistemas para controle tecnológico em ambientes indoor.

Podem ser agregadas funções de realidade virtual e aumentada para melhor complementar o sistema de mapeamento, localização e forma do objeto.

O trabalho não trata questões de precisão, somente mostra como pode ser o primeiro passo a ser dado na escolha de uma melhor ferramenta para se trabalhar com o tipo de tecnologia citado.

É necessário aprofundar o estado atual da pesquisa para a utilização da confecção de mapas a fim de se realizar o sensoriamento do local a ser implantado/utilizado o sistema, além de se estudar uma melhor tecnologia e estudo de precisão.

REFERÊNCIAS

BUSINESS NEWS DAILY. **Location-Based Services Definitions & Examples**. 2013. Disponível em: <<http://www.businessnewsdaily.com/5386-location-based-services.html>>. Acesso em 25 set. 2016.

DIRECTIONS MAGAZINE. **Ten Things You Need to Know About Indoor Positioning**. 2016. Disponível em: <<http://www.directionsmag.com/entry/10-things-you-need-to-know-about-indoor-positioning/324602>>. Acesso em 25 set. 2016.

JOURNAL OF COMPUTER NETWORKS AND COMMUNICATIONS. **Recent Advances in Wireless Indoor Localization Techniques and System**. 2013. Disponível em: <<https://www.hindawi.com/journals/jcnc/2013/185138>>. Acesso em 25 set. 2016.

KING Thomas; KOPF, Stephan; HAENSELMANN, Thomas; LUBBERGER, Christian; EFFELSBERG, Wolfgang. **COMPASS: A Probabilistic Indoor Positioning System Based on 802.11 and Digital Compasses**. Department of Computer Science, University of Mannheim. Mannheim, Germany.

PRIYANTHA, Nissanka Bodhi. **The Cricket Indoor Location System**. 2005. Disponível em: <<http://nms.csail.mit.edu/papers/bodhi-thesis.pdf>>. Acesso em 25 set. 2016.

VENKAT, Padmanabhan; VICTOR, Bahl. **RADAR**. 2005. Disponível em: <<https://www.microsoft.com/en-us/research/project/radar/>>. Acesso em 29 set. 2016.

YOUSSEF, Moustafa; AGRAWALA, Ashok. **The Horus WLAN Location Determination System. Cricket Indoor Location System**. Disponível em: <http://www.cs.umd.edu/~moustafa/papers/horus_userenix.pdf>. Acesso em 29 set. 2016.

UMA ABORDAGEM BIDINÂMICA PARA A IDENTIFICAÇÃO DE ETIQUETAS RFID

Shalton Viana dos Santos

Centro de Informática (CIn) – Universidade Federal de Pernambuco (UFPE) – Recife – PE – Brasil

Paulo André da S. Gonçalves

Centro de Informática (CIn) – Universidade Federal de Pernambuco (UFPE) – Recife – PE – Brasil

RESUMO: O DFSA (Dynamic Framed-Slotted Aloha) é um protocolo de acesso ao meio comumente usado em sistemas RFID. Tradicionalmente, o tamanho dos quadros é ajustado dinamicamente em função apenas do valor retornado pelo estimador de população de etiquetas empregado em abordagens sem reset de quadro. Este artigo propõe uma abordagem bidinâmica para o processo de identificação de etiquetas RFID. A proposta é duplamente dinâmica por contar com a dinamicidade no reajuste de tamanho de quadros do DFSA e por usar fatores multiplicativos dinâmicos no reajuste de estimativas retornadas pelo estimador utilizado no sistema RFID. As avaliações de desempenho da abordagem proposta foram realizadas através de simulação utilizando-se três estimadores distintos e requisitos temporais do canal de comunicação no padrão ISO 18000-6 Type C. Os resultados mostram que a abordagem bidinâmica permite otimizar o sistema para todos os estimadores estudados, reduzindo o tempo total de identificação de

etiquetas.

PALAVRAS-CHAVE: RFID, tempo de identificação, custo computacional.

A BIDINÂMIC APPROACH FOR RFID LABEL IDENTIFICATION

ABSTRACT: DFSA (Dynamic Framed Slotted ALOHA) is an access medium control protocol commonly used in RFID systems. Typically, the size of the frames under approaches without frame size resetting is dynamically adjusted in accordance with the specific value returned by the tag estimation method used. We propose a different and dual dynamic approach for the tag identification process. Our approach is doubly dynamic because it takes advantage of the dynamism of DFSA while using dynamic multiplicative factors for readjusting the specific value returned by the tag estimation method used. We evaluate our approach through simulation by using three different tag estimation methods. Time requirements of the communication channel are modelled in accordance with ISO 18000-6 Type C standard. The results show that our dual dynamic approach allows to optimize the system by minimizing the tag identification time under all the tag estimation methods evaluated.

KEYWORDS: Key-word: RFID, identification time, computational cost.

1 | INTRODUÇÃO

A Internet das Coisas ou *Internet of Things* (IoT) é uma nova e promissora abordagem de rede composta por objetos físicos inteligentes (e.g. sensores, atuadores, eletrodomésticos, dispositivos eletrônicos, veículos, prédios, produtos em supermercados, produtos industrializados em geral) que podem, dependendo da aplicação, realizar comunicações autônomas, interagir entre si e trocar dados com a Internet [Perera et al. 2015] [Al-Fuqaha et al. 2015]. Existem aplicações de IoT no cotidiano onde é interessante que exista alguma maneira de contar e identificar de forma rápida e automática uma determinada quantidade de objetos. Como exemplo, quanto mais rápido um caixa de supermercado contar e identificar os produtos de uma determinada compra, menores seriam as filas e mais satisfeitos ficariam os clientes com a eficiência dos caixas. Outro exemplo está após o processo de fabricação de grandes quantidades de produtos (e.g. canetas, parafusos, tampas), onde uma contagem automática deve ser rapidamente feita no momento de embalagem, distribuição e recepção pelo revendedor.

Os sistemas RFID (*Radio Frequency Identification*) aparecem como os mais promissores no contexto de IoT não só para a identificação automática de objetos como também para outras aplicações. As principais razões para isso decorrem dos seguintes fatores: 1) a comunicação é através de sinais de radiofrequência (RF), não requerendo linha de visada direta; 2) o alcance de comunicação varia de centímetros a vários metros; 3) etiquetas RFID permitem suporte a aplicações de sensoriamento, localização, contabilização e indexação de objetos; 4) as etiquetas RFID são econômicas em termos de recursos de *hardware*, o que reduz custos financeiros; e 5) as etiquetas RFID podem funcionar sem bateria e sem necessidade de recarga constante caso usem bateria.

Os sistemas RFID mais simples são compostos por um leitor, possuindo ou não um servidor associado, e uma ou mais etiquetas. Os leitores podem ser ou não portáteis e, geralmente, não são tão restritos em termos de processamento e memória quanto as etiquetas. A depender do uso de bateria para a tarefa de comunicação, as etiquetas podem ser classificadas em passivas, semi-ativas ou ativas. As passivas não possuem bateria. Elas obtêm energia do sinal RF recebido do leitor e transmitem usando uma técnica conhecida por *backscatter*. As etiquetas ativas possuem bateria para comunicação e processamento. As etiquetas semi-ativas utilizam *backscatter* para comunicação e a bateria para processamento. Cada etiqueta possui um identificador (ID) único, podendo ser colada ou embutida em um objeto. E a depender da aplicação, o leitor requisita o ID das etiquetas em seu alcance de comunicação para identificação dos objetos. Nesse processo, como existe a possibilidade de haver colisões de transmissões, é necessário utilizar um protocolo anti-colisão de etiquetas a fim de se resolver os conflitos de transmissão e permitir uma rápida identificação de todos os objetos.

Dentre os protocolos anticolisão para sistemas RFID, o DFSA (*Dynamic Framed Slotted ALOHA*) é um dos mais populares [EPC Global 2015, Klair et al. 2010]. A execução do DFSA é orientada pelo leitor que organiza o tempo em um ou mais quadros. Cada quadro é ainda subdividido em *slots* de tempo. As etiquetas são requisitadas a transmitir em um *slot*, aleatoriamente escolhido a cada quadro, até serem identificadas pelo leitor. Contudo, o DFSA requer o uso de alguma técnica para reajuste dinâmico do tamanho dos quadros. Existem vários estudos que fazem isso com base na estimativa da população de etiquetas competindo por *slots* [Andrade and Gonçalves 2013, Klair et al. 2010] ou com base em algum algoritmo [EPC Global 2015].

Os estimadores ou algoritmos de ajuste de tamanho de quadro para o DFSA em sistemas RFID podem ser divididos em duas classes básicas: com e sem *reset* de quadro. A primeira classe recalcula o valor do próximo quadro a ser aberto somente ao término da execução de um quadro completo [Andrade and Gonçalves 2013, Eom and Lee 2010, Li and Wang 2011, Wu and Zeng 2010, Tong et al. 2009, Chen 2009, Vogt 2002, Schoute 1983]. A segunda classe decide ao término de cada *slot* ou de um subgrupo de *slots* consecutivos se cancela ou não a execução dos *slots* restantes e recalcula o tamanho do próximo quadro [Šolić et al. 2016, Chen 2014, EPC Global 2015].

Tradicionalmente, o tamanho dos quadros no DFSA na classe sem *reset* é ajustado dinamicamente em função apenas do valor retornado pelo estimador de população de etiquetas utilizado. Assim sendo, há apenas um fator dinâmico sendo utilizado no processo de identificação de etiquetas RFID. De forma diferente, este artigo propõe uma abordagem bidinâmica para esse processo. A proposta é duplamente dinâmica por contar com a dinamicidade no reajuste de tamanho de quadros do DFSA e por usar fatores multiplicativos dinâmicos no reajuste de estimativas retornadas pelo estimador utilizado no sistema RFID. Para demonstrar a potencialidade da proposta, foram realizadas avaliações de desempenho através de simulações com os estimadores clássicos *Lower Bound* e Schoute bem como com o estimador Eom-Lee, sendo este último um dos mais acurados já propostos na literatura. A modelagem do canal de comunicação no estudo de simulação é feita com base na estrutura de temporização do canal de comunicação do padrão ISO 18000-6 Type C, o qual é também conhecido como EPC Class-1 Generation 2 [EPC Global 2015].

O foco deste trabalho está em abordagens que analisam todo o quadro em execução para cálculo do tamanho do quadro subsequente. Os estudos apresentados neste artigo utilizam o *Lower Bound* e o Schoute devido ao fato de serem os estimadores mais simples já propostos e com custo computacional muito baixo. Em contrapartida, eles possuem baixa acurácia, o que reflete, geralmente, em uma maior quantidade de *slots* utilizados no processo de identificação de etiquetas em relação a estimadores mais acurados e pior tempo de identificação de etiquetas [Andrade and Gonçalves 2011]. Por outro lado, o estimador Eom-Lee possui um custo computacional elevado, sendo porém um dos estimadores mais acurados já propostos na literatura [Andrade

and Gonçalves 2013, Eom and Lee 2010]. Desta forma, é possível mostrar o impacto da proposta deste artigo com tipos de estimadores distintos em relação ao custo computacional, à qualidade de estimação e ao tempo de identificação.

Os resultados alcançados mostram que a abordagem bidinâmica proposta permite otimizar o sistema com qualquer um dos estimadores sem *reset* estudados, reduzindo o tempo total de identificação de etiquetas. As principais contribuições deste artigo são a proposição de um otimizador de estimadores sem *reset* de quadros e a demonstração de que é possível se alcançar um desempenho similar ao do Eom-Lee utilizando o simples *Lower Bound* com o otimizador resultante da proposta bidinâmica e com custo computacional significativamente menor. O trabalho proposto é uma extensão e generalização de abordagem inicialmente proposta em [Andrade and Gonçalves 2011] que utiliza fatores multiplicativos fixos, independentes da saída do estimador, em vez de fatores multiplicativos dinâmicos, dependentes da saída do estimador do sistema RFID para reajuste de quadros. A otimização de abordagens derivadas do estimador Vogt [Andrade and Gonçalves 2013, Li and Wang 2011] ou que resetam o tamanho dos quadros em momento de execução a partir da análise de cada *slot* ou um subgrupo de *slots* consecutivos será abordada em trabalhos futuros.

O restante deste artigo está organizado da seguinte forma: a Seção 2 apresenta os trabalhos relacionados. A Seção 3 detalha e discute especificamente os estimadores *Lower Bound*, Schoute e Eom-Lee. A Seção 4 apresenta e discute a abordagem bidinâmica proposta para a identificação de etiquetas RFID. A Seção 5 apresenta as avaliações de desempenho realizadas com canal de comunicação sob requisitos temporais do padrão ISO 18600-6 Type C. Por fim, a Seção 6 conclui este trabalho.

2 I TRABALHOS RELACIONADOS

Esta seção apresenta os principais estimadores e algoritmos para o DFSA com e sem *reset* de quadros. Para facilitar a leitura, a Tabela 1 apresenta os principais parâmetros utilizados e o significado de cada um.

Parâmetro	Significado
f	tamanho do quadro executado ou sendo executado
\hat{f}	tamanho do quadro subsequente calculado pelo estimador
\hat{n}	estimativa de etiquetas não identificadas (<i>backlog</i>) calculada pelo estimador
s_s	quantidade de <i>slots</i> bem sucedidos
s_v	quantidade de <i>slots</i> vazios
s_c	quantidade de <i>slots</i> em colisão

Tabela 1. Lista de Parâmetros.

2.1 Sem Reset

A Figura 1 ilustra o funcionamento da classe sem *reset*. Basicamente, um quadro de tamanho f é aberto e cada etiqueta escolhe aleatoriamente um *slot* para transmitir informações. Ao término do quadro, caso haja ao menos um *slot* em colisão, executa-

se um estimador para se determinar o tamanho \hat{f} do quadro subsequente a ser utilizado. Os parâmetros utilizados como entrada para um estimador variam de acordo com a proposta. Contudo, essas entradas costumam ser os parâmetros s_v , s_s , s_c e f ou um subconjunto deles, onde s_v , s_s e s_c representam, respectivamente, o total de *slots* vazios, bem sucedidos e em colisão no quadro executado com tamanho f . O estimador calcula o tamanho \hat{f} do quadro subsequente com base na estimativa do quantitativo de etiquetas que não foram identificadas no quadro de tamanho f executado. Essa estimativa é representada neste trabalho por \hat{n} .

Existem diversos estimadores sem *reset* de quadro propostos na literatura como o *Lower Bound* [Vogt 2002], o Schoute [Schoute 1983], o Vogt [Vogt 2002], o Eom-Lee [Eom and Lee 2010], o Chen [Chen 2009], o IV-II [Andrade and Gonçalves 2013], o CMEBE [Li and Wang 2011] e estimadores Bayesianos como os apresentados em [Wu and Zeng 2010] e [Tong et al. 2009]. Os estimadores *Lower Bound*, Schoute e Eom-Lee são utilizados nas avaliações de desempenho neste artigo e, por isso, são apresentados e discutidos separadamente na Seção 3. Os demais estimadores são apresentados a seguir.

O estimador Vogt [Vogt 2002] modela o processo de identificação de etiquetas com base em uma distribuição binomial. Basicamente, o estimador procura um quantitativo de etiquetas que minimiza a norma Euclidiana da diferença entre dois vetores: um com os valores de s_v , s_s e s_c e outro com os valores esperados para esses parâmetros. Esse quantitativo de etiquetas é a quantidade estimada de etiquetas que competiram por *slots* no quadro de tamanho f analisado. O tamanho \hat{f} do quadro subsequente é igual a esse quantitativo menos s_s . Em [Eom and Lee 2010] é sugerido que o Vogt possui maior custo computacional do que o Eom-Lee. Contudo, a literatura ainda carece de estudos mais detalhados com comparativos de custos entre o Vogt e o Eom-Lee em função do quantitativo de etiquetas a serem identificadas.

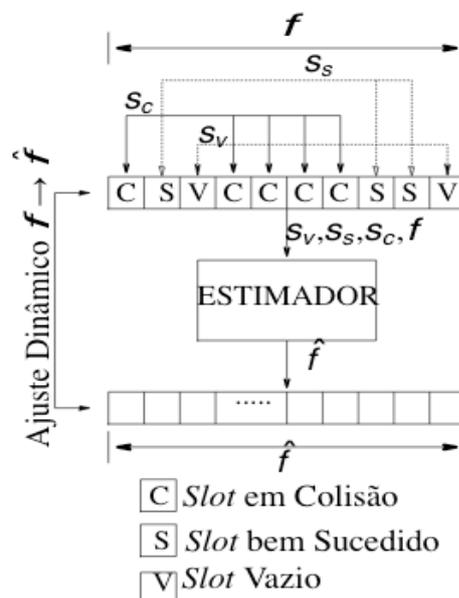


Figura 1. Abordagem sem reset de quadro.

O IV-II [Andrade and Gonçalves 2013] identifica e trata a deficiência do estimador Vogt na produção de estimativa de etiquetas quando o quadro analisado possui todos os *slots* em colisão. Os resultados sugerem que o IV-II e o Eom-Lee são equivalentes em termos de quantidade de *slots* totais gerados no processo de identificação com um quadro inicial de 64 *slots*. Quando o quadro inicial é de 128 *slots*, os dois estimadores apresentam desempenho equivalente até 700 etiquetas. Acima disso, o IV-II apresenta melhor desempenho, chegando a usar até 69 *slots* a menos para identificar 1.000 etiquetas. Adicionalmente, o IV-II possui desempenho idêntico ou melhor do que o do Vogt, chegando a usar até menos 232 *slots* para identificar 1.000 etiquetas. O custo do IV-II é ligeiramente maior do que o do Vogt.

O estimador Chen [Chen 2009] também modela inicialmente o processo de identificação de etiquetas com base em uma distribuição binomial. Contudo, ele procura determinar a probabilidade de se obter exatamente sv vazios, ss sucessos e sc colisões dado um quadro de tamanho f . A solução para esse problema deriva uma equação de probabilidades $P(nlf, sv, sc, ss)$ que requer o cálculo de vários fatoriais e exponenciações. O valor de n que maximiza a probabilidade $P(nlf, sv, sc, ss)$ é a estimativa de etiquetas que competiram por *slots* no quadro de tamanho f . Em geral, o estimador Chen usa mais *slots*, é menos acurado e mais custoso do que os estimadores Eom-Lee e IV-II [Eom and Lee 2010, Andrade and Gonçalves 2013]

Os estimadores em [Wu and Zeng 2010] e [Tong et al. 2009] se diferenciam dos demais por utilizarem um processo Bayesiano para a determinação da estimativa da quantidade de etiquetas que competiram por *slots* num quadro e para o cálculo do tamanho ótimo do quadro subsequente. A contrapartida é o custo computacional associado ao processo e, por isso, técnicas de redução de espaço de busca precisam ser utilizadas para tornar viável o uso de estimadores desse tipo. Em [Wu and Zeng 2010] é mostrado como se reduzir o espaço de buscas e como se reduzir a complexidade computacional.

Em [Andrade and Gonçalves 2011] é apresentada a proposta de uma função de cálculo de tamanho de quadros para o DFSA. A função proposta reajusta o tamanho de quadro obtido pelo estimador empregado. Os estudos foram realizados com o *Lower Bound*, Schoute e Eom-Lee. A função ajusta o valor do tamanho de quadro da saída do estimador, multiplicando-o por um fator fixo cujo valor depende do estimador, mas independe do valor de saída do estimador. Os resultados mostram que é possível reduzir significativamente o tempo total de identificação de etiquetas dependendo da população de etiquetas a ser identificada e da relação temporal entre os diferentes tipos de *slots* de tempo.

Outros estimadores sem *reset* focam em melhorar a qualidade das estimativas em cenários com canais de comunicação sob efeito de captura como o CMEBE [Li and Wang 2011]. Esse estimador pode ser visto como uma variante do Vogt que considera um canal de comunicação com efeito de captura e calcula o tamanho \hat{f} do

quadro subsequente com base em uma estimativa \hat{n} de etiquetas e uma estimativa da probabilidade média de captura em *slots* em colisão.

2.2 Com Reset

Outros estudos na literatura focam no desenvolvimento de estimadores com *reset*, onde o quadro de tamanho f , em execução, é analisado ao término de cada *slot* ou de um conjunto consecutivo de *slots*. Com base na análise realizada, o estimador pode decidir cancelar a abertura dos *slots* restantes e abrir imediatamente um novo quadro com um tamanho \hat{f} recalculado. Comparada com a abordagem sem *reset*, a com *reset* incorre em processamento extra a cada *slot* ou conjunto de *slots* devido ao algoritmo de decisão de anulação de *slots*. Além disso, o cancelamento de *slots* aumenta o custo das mensagens do leitor para as etiquetas visto a necessidade de informá-las sobre o cancelamento de um quadro para a geração de novo número pseudoaleatório.

O estimador Chen II [Chen 2014] pode ser visto como uma versão do Schoute que analisa grupos consecutivos de *slots*. Ao analisar um grupo de *slots*, o estimador verifica se os parâmetros de desempenho obtidos a partir do grupo analisado correspondem ao esperado para se maximizar a eficiência do sistema. Caso não correspondam, os *slots* restantes são cancelados e um novo quadro é aberto. A estimativa \hat{n} é igual a $(ss + 2,39sc)f/i$, onde $4 \leq i \leq f$ é o índice do último *slot* do subgrupo sendo analisado. O tamanho do quadro subsequente \hat{f} é dado por \hat{n} menos o quantitativo de *slots* bem sucedidos até o *slot* i . Caso não haja *reset* do quadro, o tamanho do quadro subsequente e a estimativa de etiquetas são obtidas conforme o estimador Schoute.

O padrão EPC Classe 1 Gen2 [EPC Global 2015] especifica dois procedimentos de identificação de etiquetas: um com quadros de tamanho fixo igual a 2^Q , ou seja, um FSA (*Framed Slotted Aloha*) simples e outro com quadros ajustados dinamicamente com tamanho 2^Q , porém, com Q recalculado ao término de cada *slot* com base no algoritmo-Q. Este segundo procedimento resolve ineficiências típicas de um sistema FSA e nada mais é do que um DFSA com *reset* de quadro ao término de cada *slot* vazio ou em colisão. Quando um *slot* termina, o leitor analisa o tipo de *slot*. De acordo com o resultado da análise, o leitor atualiza o valor do parâmetro Q_{fp} que nada mais é do que a versão em ponto flutuante do parâmetro inteiro Q e é calculado como segue: $Q_{fp} = \min(15, Q_{fp} + C)$ para *slot* em colisão, $Q_{fp} = \max(0, Q_{fp} - C)$ para *slot* vazio e $Q_{fp} = Q_{fp} + 0$ para *slot* bem sucedido. O novo valor do parâmetro Q passa a ser $\text{round}(Q_{fp})$. Valores para C são dinâmicos e estão no intervalo de 0,1 a 0,5. A dificuldade está em se determinar valores adequados para C visto que valores ótimos são dependentes do número de etiquetas competindo por *slots*. O padrão não define como calcular valores adequados para C . Em [Khanna and Uysal 2015] é proposta uma modificação no algoritmo-Q. A proposta utiliza informações da camada física para estimar a quantidade de etiquetas que transmitiram em um mesmo *slot*e, assim, calcular um valor otimizado para o parâmetro Q .

O ILCM SbS [Šolić et al. 2016] é uma extensão de trabalhos anteriores dos autores e foca em produzir um valor Q otimizado para sistemas baseados no algoritmo-Q. Embora tenha como base uma equação de probabilidades $P(nlf, sv, sc, ss)$ que também requer o cálculo de vários fatoriais e exponenciações, os autores conseguem encontrar funções menos custosas. Os resultados para um quantitativo de até 250 etiquetas mostram que o ILCM SbS permite um tempo de identificação inferior ao de uma versão do Vogt com *reset* de quadros.

3 I OS ESTIMADORES ESTUDADOS

Esta seção detalha e discute os estimadores *Lower Bound*, Schoute e Eom-Lee dado que serão os utilizados nas avaliações de desempenho neste artigo. Como explicitado na Seção 1, a escolha dos estimadores *Lower Bound* e Schoute é em virtude da simplicidade e do baixo custo computacional deles. Em contrapartida, são estimadores de baixa acurácia, o que reflete, geralmente, em um maior tempo de identificação de etiquetas em relação a estimadores mais acurados. Por outro lado, o estimador Eom-Lee possui um custo computacional significativamente maior, porém é um dos mais acurados já propostos na literatura. Assim, é possível avaliar o impacto da proposta neste artigo com estimadores distintos em relação ao custo computacional, à qualidade de estimação e ao tempo total de identificação de etiquetas. A Tabela 2 apresenta o custo de operações em ponto flutuante (FLOP) [Vales-Afonso et al. 2015] para análise dos estimadores.

Operação	Custo
adição, subtração, multiplicação	1
comparação	2
divisão, raiz quadrada	10
exponenciação, logaritmo	50
fatorial	100

Tabela 2. Custos de operações em ponto flutuante (FLOP).

3.1 Lower Bound

Formalmente, o estimador *Lower Bound* define [Vogt 2002]:

$$\hat{n} = s_s + 2 \cdot s_c, \quad (1)$$

$$\hat{f} = 2 \cdot s_c. \quad (2)$$

A lógica desse estimador é simples: há ao menos duas etiquetas envolvidas em uma colisão. Logo, a menor quantidade possível de etiquetas que competiram por *slots* em um quadro sendo analisado pode ser facilmente calculada, sendo igual a

quantidade total de *slots* com transmissões bem sucedidas somada com o dobro da quantidade de *slots* em colisão. O *backlog* é a quantidade de etiquetas que ainda não foram identificadas e possui valor igual ao dobro da quantidade de *slots* em colisão. Para se maximizar a eficiência do sistema quando todos os *slots* possuem o mesmo tamanho, o quadro subsequente ao quadro sendo analisado deve ter tamanho igual ao *backlog* [Schoute 1983].

O *Lower Bound* é um estimador grosseiro, pois sempre estima a quantidade mínima possível de etiquetas restantes dentro de uma gama de possibilidades. Assim, ele pode ser visto como uma fronteira para se avaliar a qualidade de outros estimadores. Contudo, o custo FLOP dele é baixo. Com o auxílio da Tabela 2 é possível notar que o custo FLOP é igual 2 para o cálculo conjunto de cada tamanho de quadro e \hat{n} . Note que \hat{n} é obtido através de simples soma tendo \hat{f} já calculado. A Figura 2(a) apresenta o custo FLOP total de uso do estimador em função da quantidade de etiquetas a serem identificadas. Os resultados na Figura 2 são médias obtidas considerando-se a Tabela 2 e simulações de desempenho do estimador com os mesmos parâmetros apresentados na Seção 5, exceto pelo passo para o quantitativo de etiquetas. O intervalo de confiança é de 99% e é representado por barras de erro quase sempre imperceptíveis.

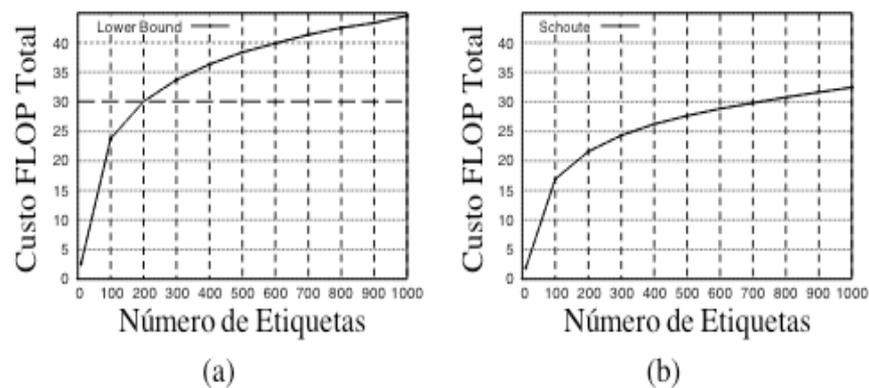


Figura 2. FLOP para os estimadores Lower Bound e Schoute.

3.2 Schoute

Formalmente, o estimador Schoute [Schoute 1983] define:

$$\hat{n} = s_s + 2,39 \cdot s_c, \quad (3)$$

$$\hat{f} = 2,39 \cdot s_c. \quad (4)$$

A mudança em relação ao *Lower Bound* está no uso do fator multiplicativo 2,39 ao invés de 2. Esse novo fator é o número esperado de etiquetas que transmitirão em

cada *slot* em colisão no próximo quadro ao que está sendo analisado. A formulação de Schoute é obtida considerando-se um processo de chegadas do tipo Poisson. Para outras distribuições, as Equações (3) e (4) trazem resultados subótimos.

Note com base nas Figuras 2(a) e 2(b) que o Schoute possui, em geral, um custo FLOP total menor do que o do *Lower Bound* apesar de ambos terem o mesmo custo FLOP no cálculo conjunto do tamanho do quadro e da estimativa de etiquetas. Isso ocorre porque o Schoute utiliza, em geral, menos quadros no processo de identificação de etiquetas. Assim, ele é executado menos vezes, impactando positivamente no custo FLOP total de uso do estimador. Para sistemas que requerem apenas o valor de \hat{f} , o custo FLOP total mostrado nas Figuras 2(a) e 2(b) cai pela metade.

3.3 O Estimador Eom-Lee

O estimador Eom-Lee [Eom and Lee 2010] utiliza um algoritmo iterativo tanto para estimar a quantidade de etiquetas que competiram por *slots* em um quadro analisado de tamanho L quanto para calcular o tamanho do quadro subsequente. O estimador Eom-Lee define que

$$\hat{f} = \gamma \cdot s_c, \quad (5)$$

onde

$$\gamma = \frac{1 - e^{-\frac{1}{\beta}}}{\beta(1 - (1 + \frac{1}{\beta})e^{-\frac{1}{\beta}})}. \quad (6)$$

Contudo, não é trivial encontrar uma solução fechada para se determinar os valores de γ e β a partir da Eq. (6). O problema é contornado calculando-se γ e β de forma iterativa. Considere γ_k e β_k , respectivamente, uma aproximação para o valor de γ e de β na k -ésima iteração do algoritmo. Essas aproximações são obtidas de acordo com as seguintes equações:

$$\beta_k = \frac{L}{\gamma_k \cdot s_c + s_s}, \quad (7)$$

$$\gamma_k = \frac{1 - e^{-\frac{1}{\beta_k}}}{\beta_k(1 - (1 + \frac{1}{\beta_k})e^{-\frac{1}{\beta_k}})}. \quad (8)$$

No primeiro passo do algoritmo iterativo, considera-se $\beta_1 = \infty$ e $\gamma_1 = 2$ e em cada passo k seguinte se determina uma nova aproximação para β e γ com o auxílio das Eqs. (7) e (8), respectivamente. Quando $|\gamma_{k-1} - \gamma_k|$ for menor que um

limiar pré- definido $\varrho_{threshold}$, o processo iterativo é interrompido. Os valores $\gamma_{k^{*-1}}$ e γ_{k^*} representam, respectivamente, a aproximação anterior e atual para o valor de γ . A partir de então, o tamanho \hat{f} do próximo quadro e a quantidade estimada \hat{n} de etiquetas são obtidos, respectivamente, pelas Eqs. (9) e (10), onde β_{k^*} é a aproximação mais recente para o valor de β .

$$\hat{f} = \gamma_{k^*} \cdot s_c . \quad (9)$$

$$\hat{n} = \frac{\hat{f}}{\beta_{k^*}} . \quad (10)$$

O custo do Eom-Lee por cálculo de tamanho de quadro depende do número de iterações calculando-se β_k e γ_k até se atingir o critério de parada. Após atingir o critério de parada, há ainda o cálculo de \hat{f} que adiciona 1 ao custo total das iterações e o cálculo de \hat{n} que adiciona mais 10 ao custo. A Figura 3(a) apresenta o número total de iterações realizadas pelo estimador ao longo do processo de estimação de etiquetas. Note que o estimador possui um aumento abrupto de iterações a partir de 300 etiquetas. A Figura 3(b) apresenta o custo FLOP total de uso do estimador em função da quantidade de etiquetas a serem identificadas. Note que o custo FLOP total é significativamente maior do que o dos estimadores *Lower Bound* e Schoute. Para sistemas que requerem apenas o cálculo de \hat{f} , não há mudança significativa no custo visto que o número de iterações e cálculos nessas iterações possuem um peso muito mais importante. Os resultados na Figura 3 são médias obtidas considerando-se a Tabela 2 e simulações de desempenho do estimador com os mesmos parâmetros apresentados na Seção 5. Os intervalos de confiança são de 99%.

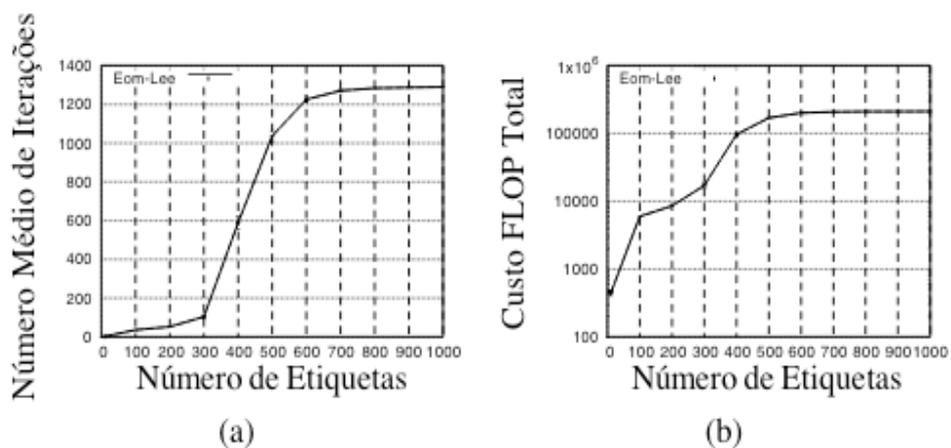


Figura 3. Número médio de iterações e custo FLOP total para o Eom-Lee.

4 | A ABOARDAGEM BIDINAMICA PROPOSTA

Conforme mencionado na Seção 1, o tamanho dos quadros no DFSA é ajustado dinamicamente em função apenas de valor retornado pelo estimador de população de etiquetas utilizado. Desta forma, nota-se que há apenas um fator dinâmico sendo utilizado no processo de identificação de etiquetas. A proposta deste trabalho consiste em uma abordagem que adicionalmente analisa o valor retornado pelo estimador e o reajusta também dinamicamente tendo como base uma função de otimização aqui proposta. Assim, o bidinamismo da proposta é consequência do reajuste dinâmico de tamanho de quadros e do reajuste dinâmico da saída do estimador empregado. A Figura 4 apresenta uma esquematização da abordagem bidinâmica proposta.

Note que o diferencial em relação à Figura 1 está no módulo *Otimizador* que altera dinamicamente o valor de \hat{f} que é retornado pelo estimador utilizado. O módulo *Otimizador* proposto neste artigo é uma generalização inspirada a partir dos achados apresentados em [Andrade and Gonçalves 2011]. Nesse trabalho, os autores demonstram ser possível reduzir o tempo total de identificação de etiquetas ao se multiplicar todo \hat{f} retornado pelo estimador por um mesmo fator fixo calculado de forma adequada. Em vez de utilizar um mesmo fator fixo, o módulo *Otimizador* proposto usa um fator otimizado que varia de acordo com o resultado de saída do estimador.

Seja p a população de etiquetas a serem identificadas. Seja $\bar{\Omega}(i, p, \delta)$ uma função que representa o tempo médio total de identificação das p etiquetas utilizando-se um quadro inicial com i slots e um fator multiplicativo de correção de tamanho de quadro \hat{f} igual

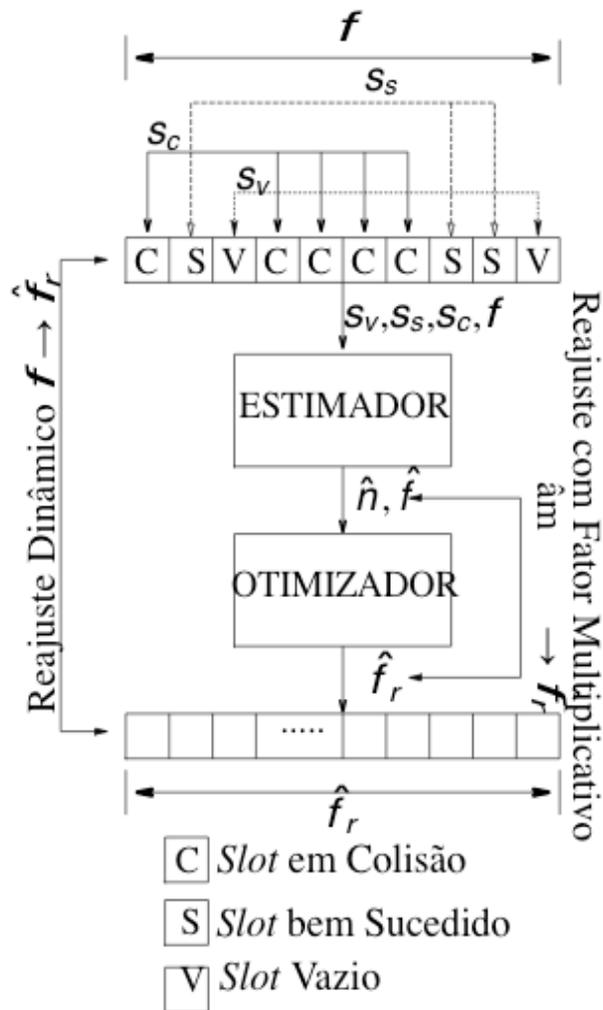


Figura 4. Abordagem bidinâmica para ajuste de tamanho de quadros.

a δ . A função $\bar{\Omega}$ é facilmente obtida através de simulação de Monte-Carlo para qualquer estimador desejado e seguindo o tamanho dos *slots* conforme norma utilizada pelo sistema RFID. A partir da função $\bar{\Omega}$ o tamanho do quadro corrigido dinamicamente para qualquer estimativa \hat{n} , respectivo tamanho de quadro \hat{f} , calculados pelo estimador empregado é dado por:

$$\hat{f}_r = \hat{f} \times \underset{\delta > 0}{\operatorname{argmin}} \bar{\Omega}(i, p, \delta) \quad \text{para } p = \hat{n}, \quad (11)$$

significando que \hat{f}_r é igual a estimativa \hat{f} do estimador empregado multiplicada pelo valor de δ que minimiza o valor da função $\bar{\Omega}$ para $p = \hat{n}$ ou seja, para p igual ao valor da estimativa de etiquetas calculado pelo estimador.

A função $\bar{\Omega}$ não deve ser calculada em tempo de execução por causa de seu custo de processamento. Ela deve ser pré-computada e os dados produzidos podem ser pré-instalados em forma de uma árvore binária em memória com p sendo chave de busca para resultado pré-computado de $\underset{\delta > 0}{\operatorname{argmin}} \bar{\Omega}(i, p, \delta)$ e dado i . Isso permite que apenas simples consultas sejam feitas em tempo de execução, reduzindo-se de

forma importante

o tempo de processamento do otimizador. Portanto, o custo do otimizador proposto em relação ao uso exclusivo do estimador escolhido está no uso adicional de memória no leitor ou servidor associado, na realização adicional de uma multiplicação e na realização de uma busca em uma árvore binária na memória. Não há alterações nas etiquetas visto que o otimizador e o estimador são sempre executados no leitor RFID ou servidor associado. O custo de memória não é significativo. Cada nó da árvore precisa armazenar um único valor de δ para o valor de i definido no sistema RFID considerado. Valores típicos de i são 32, 64 e 128 *slots*. Além disso, o uso de árvore binária balanceada reduz significativamente o custo das consultas visto a complexidade de pior caso ser igual a $O(\log pmax)$, onde $pmax$ é o número total de nós da árvore. Estudos na literatura consideram tipicamente a identificação de até 1.000 etiquetas, o que limita o valor de $pmax$.

51 AVALIAÇÃO DE DESEMPENHO

Esta seção apresenta um estudo de simulação para avaliação do desempenho da abordagem bidinâmica proposta face a abordagem tradicional, considerando-se os estimadores *Lower Bound*, *Schoute* e *Eom-Lee*. Para isso, o simulador do protocolo DFSA para RFID desenvolvido por [Andrade and Gonçalves 2013] foi estendido com o módulo *Otimizador*. As simulações apresentadas consideram um sistema básico RFID com um leitor e um determinado quantitativo de etiquetas a serem identificadas. Esse quantitativo é variado de 10 a 1.000 em passos de 10. As métricas de avaliação são o tempo total de identificação e a redução média percentual no tempo total de identificação de etiquetas. Ambas as métricas consideram apenas o tempo gasto com *slots* vazios, em colisão e bem sucedidos. Esses tempos são os seguintes conforme norma ISO 18000-6 Type C: 1) tempo de *slot* bem sucedido = $TS = T4 + TQuery + 2T1 + 2T2 + TRN16 + TACK + TPC + EPC + CRC16 + TQREP = 2312 \mu s$; 2) tempo de *slot* em colisão = $TC = T1 + T2 + TRN16 = 337,5 \mu s$ e; 3) tempo de *slot* vazio = $TE = T1 + T3 = 67,5 \mu s$.

Nos gráficos apresentados, cada ponto em cada curva apresentada é uma média dos resultados de 2.000 simulações. O canal de comunicação é livre de erros visto o intuito de se estudar o impacto isolado dos estimadores e deles com o otimizador. Contudo, a proposta é aplicável com estimadores que levam em conta efeitos de captura, o que será melhor investigado em trabalhos futuros. Em particular às simulações com o estimador *Eom-Lee*, o parâmetro $q_{threshold}$ é igual a 0,001. Todas as simulações consideram um quadro inicial de 64 *slots*. Todos os resultados foram calculados com intervalo de confiança de 99%.

A Figura 5(a) mostra a redução média percentual no tempo total de identificação de etiquetas ao se utilizar a abordagem bidinâmica proposta. Cada curva revela o ganho obtido com a inclusão do otimizador em relação ao uso de cada estimador de forma

isolada. Note que o otimizador consegue melhorar o tempo de identificação com os três estimadores e possui maior impacto com o *Lower Bound* visto que o ganho alcança perto de 9% para 1.000 etiquetas. Embora o Eom-Lee seja um estimador acurado, o gráfico mostra que alguma otimização ainda é possível. A Figura 5(b) apresenta o tempo total de identificação de etiquetas utilizando-se apenas as versões originais do *Lower Bound* e do Eom-Lee. O desempenho do Eom-Lee começa a se diferenciar de forma discreta do desempenho do *Lower Bound* a partir de 100 etiquetas. É importante observar que para diversos quantitativos de etiquetas, o maior custo FLOP total do Eom-Lee não implica em uma melhoria significativa no tempo total de identificação de etiquetas. Já para 300 e 1.000 etiquetas, por exemplo, o Eom-Lee permite um tempo total de identificação em torno de 6,3% e 8,5% menor, respectivamente. São ganhos consideráveis, mas a contra- partida é um custo de processamento significativo. As Figuras 5(c) e 5(d) apresentam um resultado interessante. Elas comparam o tempo total de identificação de etiquetas utilizando-se somente o Eom-Lee original com a versão do *Lower Bound* otimizada pela abordagem bidinâmica proposta. Note que o otimizador permite levar o *Lower Bound*, um estimador inacurado e de baixo custo computacional, a um desempenho similar ao do Eom-Lee, um estimador acurado mas com custo computacional significativo. E isso tro- cando processamento por memória no leitor ou servidor associado e adicionando apenas uma multiplicação e uma busca em árvore binária balanceada que possui custo baixo.

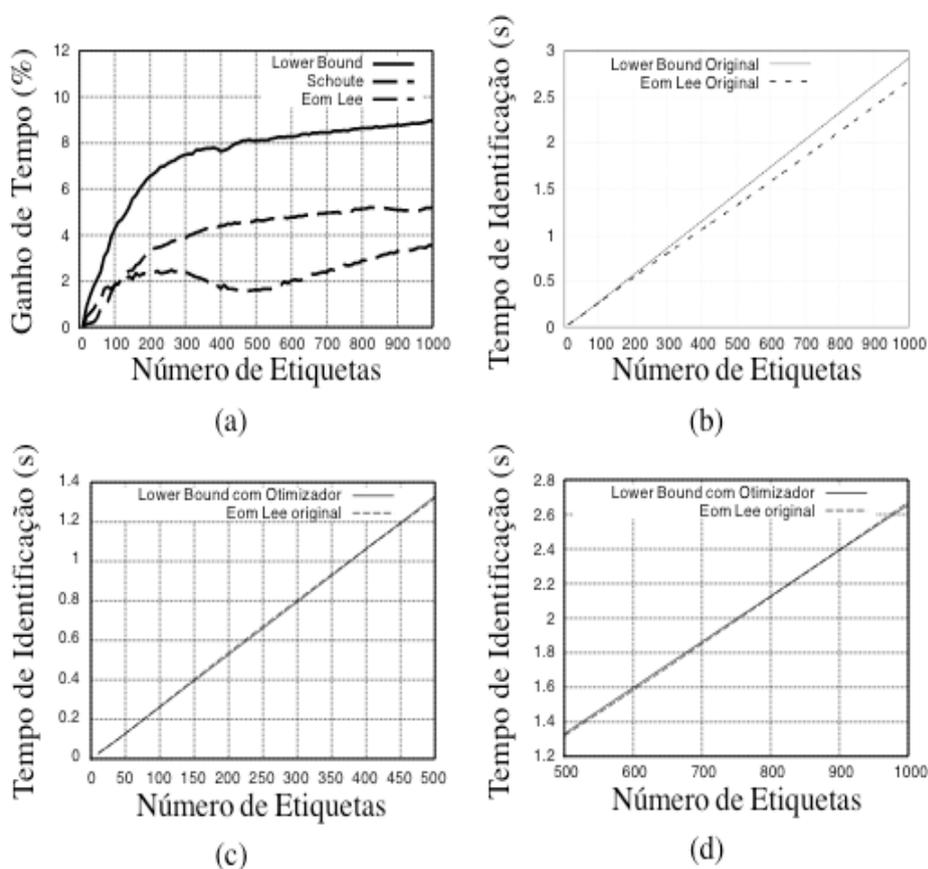


Figura 5. Resultados das Avaliações de Desempenho.

6 | CONCLUSÕES

Este artigo propôs uma abordagem bidinâmica para a rápida identificação de etiquetas em sistemas RFID baseados no DFSA sem *reset* de quadros. A proposta otimiza dinamicamente a saída do estimador empregado pelo sistema RFID. As principais contribuições deste artigo foram a proposição de um otimizador de estimadores sem *reset* de quadros e a demonstração de que é possível se alcançar um desempenho similar ao do Eom-Lee utilizando apenas o simples *Lower Bound* associado ao otimizador resultante da proposta bidinâmica, mas com custo computacional significativamente menor do que o do Eom-Lee. Para isso, trocou-se processamento por memória e adicionou-se uma multiplicação e uma busca em árvore binária por tamanho de quadro calculado. Os trabalhos futuros incluem 1) o mapeamento do impacto da abordagem proposta com outros estimadores sem *reset*; 2) a extensão da proposta para uso com estimadores com *reset* e os impactos de desempenho e; 3) a extensão da proposta para sistemas baseados no algoritmo-Q e impactos de desempenho.

REFERÊNCIAS

- Al-Fuqaha, A., Guizani, M., Aledhari, M., and Ayysah, M. (2015). Internet of things: A Survey on Enabling Technologies, Protocols and Applications. *IEEE Communications Surveys Tutorials*, 17(4):2347 – 2376.
- Andrade, J. D. and Gonçalves, P. A. S. (2011). Uma Função de Cálculo de Tamanho de Frames para o Protocolo DFSA em Sistemas RFID. In *Proc. of XVI Workshop de Gerência e Operação de Redes (WGRS)*, pages 61–74, Campo Grande, MS.
- Andrade, J. D. and Gonçalves, P. A. S. (2013). Um Estimador Acurado para o Protocolo DFSA em Sistemas RFID. In *Proc. of XXXI Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC)*, pages 325–338, Brasília, DF.
- Chen, W.-T. (2009). An Accurate Tag Estimate Method for Improving the Performance of an RFID Anticollision Algorithm Based on Dynamic Frame Length ALOHA. *IEEE Transactions on Automation Science and Engineering*, 6(1):9 –15.
- Chen, W.-T. (2014). A Feasible and Easy-to-Implement Anticollision Algorithm for the EPCglobal UHF Class-1 Generation-2 RFID protocol. *IEEE Transactions on Automation Science and Engineering*, 11:485 – 491.
- Eom, J.-B. and Lee, T.-J. (2010). Accurate Tag Estimation for Dynamic Framed-slotted ALOHA in RFID Systems. *IEEE Communications Letters*, 14:60–62.
- EPC Global, I. (2015). *EPC radio-frequency identify protocols Generation-2 UHF RFID Specification for RFID air interface protocol for communications at 860MHz-960 MHz - ratified*, 2.0.1 edition.
- Khanna, N. and Uysal, I. (2015). Q-frame-collision-counter: a novel and dynamic approach to RFID Gen 2's Q algorithm. In *Proc. of the IEEE International Conf. on RFID Technology and Application (RFID-TA)*, pages 120–125.
- Klair, D., Chin, K.-W., and Raad, R. (2010). A Survey and Tutorial of RFID Anti-Collision Protocols.

IEEE Communications Surveys Tutorials, 12(3):400–421.

Li, B. and Wang, J. (2011). Efficient Anti-collision Algorithm Utilizing the Capture Effect for ISO 18000-6C Protocol. *IEEE Communications Letters*, 15:352–354.

Perera, C., Liu, C. H., and Jayawardena, S. (2015). The Emerging Internet of Things Marketplace From an Industrial Perspective: A Survey. *IEEE Transactions on Emerging Topics in Computing*, 3(4):585–598.

Schoute, F. C. (1983). Dynamic Frame Length ALOHA. *IEEE Transactions on Communications*, 31:565–568.

Tong, Q., Zou, X., and Tong, H. (2009). Dynamic Framed Slotted ALOHA Algorithm Based on Bayesian Estimation in RFID System. In *Proceedings of the WRI World Congress on Computer Science and Information Engineering*, pages 384–388, Washington, DC, USA. IEEE Computer Society.

Vales-Afonso, J., Bueno-Delgado, V., Egea-Lopez, E., Gonzales-Castano, F., and Alcaraz, J. (2015). Multiframe maximum-likelihood tag Estimation for RFID Anticollision Protocols. *IEEE Transactions on Industrial Informatics*, 7(3):487–496.

Vogt, H. (2002). Efficient Object Identification with Passive RFID Tags. In *Proc. of the First International Conf. on Pervasive Computing*, pages 98–113, London, UK.

Šolić, P., Radić, J., and Rožić, N. (2016). Early Frame Break Policy for ALOHA-Based RFID Systems. *IEEE Trans. on Automation Science and Engineering*, 13(2):876–881.

Wu, H. and Zeng, Y. (2010). Bayesian tag estimate and optimal frame length for anti-collision aloha rfid systems. *IEEE Transactions on Automation Science and Engineering*, 7(4):963–969.

TESTE DE FERRAMENTAS DE RECUPERAÇÃO DE IMAGENS PARA SISTEMAS DE ARQUIVOS EXT3 E EXT4

Diego Vinícius Natividade

e-mail: natividade@bol.com.br

Universidade Federal de Lavras
Lavras - MG

RESUMO: Com o aumento do uso de sistemas GNU/Linux, recuperação de dados nestes sistemas tornou-se muito comum. Usar a ferramenta ideal para cada tipo de sistema de arquivos é de suma importância para o êxito da tarefa, seja no quesito desempenho ou eficiência. Este trabalho apresentou de forma prática e objetiva, três das principais e mais conhecidas ferramentas para recuperação de dados e as comparou em sistemas de arquivos ext3 ext4, de forma a possibilitar que o leitor possa escolher a solução mais adequada para cada ocasião. O resultado mostrou que, em geral, o Photorec obteve o melhor desempenho.

PALAVRAS-CHAVE: recuperação de dados; recuperação de imagens; ext3; ext4

TEST OF IMAGE RECOVERY TOOL FOR EXT3 AND EXT4 FILE SYSTEMS

ABSTRACT: With the increasing use of GNU/Linux systems, the retrieval of data in these systems has become very common. Using the ideal tool for each type of file system is paramount to the success of the task, be it

performance or efficiency. This work presented in a practical and objective way three of the most well known tools for data recovery and compared them in ext3 and ext4 file systems, in order to allow the reader to choose the most suitable solution for each occasion. The result showed that, in general, Photorec achieved the best performance.

KEYWORDS: data recovery; image recovery; ext3; ext4

1 | INTRODUÇÃO

Com o aumento do uso de sistemas Linux tanto em servidores quanto em *desktops*, ter em mãos ferramentas adequadas que otimizem o trabalho de recuperação de dados para estes sistemas é de fundamental importância, seja para técnicos, analistas, peritos ou mesmo usuários e interessados no assunto.

O objetivo principal deste trabalho foi testar três ferramentas de recuperação de dados em sistemas de arquivos ext3 e ext4, comparando-as quanto à eficiência na recuperação de arquivos de imagens.

Como resultado, tem-se uma visão geral das ferramentas apresentadas podendo optar por uma ou outra, dependendo do sistema de arquivos utilizado e das condições do mesmo – se foi formatado ou não.

2 | RECUPERAÇÃO DE DADOS

Na maioria dos sistemas atuais, quando um arquivo é apagado, o que o sistema operacional faz é excluir a referência do arquivo na tabela de alocação de arquivos e liberar os blocos ocupados pelo mesmo para a gravação de novos. Portanto, um arquivo mesmo excluído há anos, pode ser recuperado, se os blocos por ele ocupados ainda não tiverem sido sobrescritos por outros arquivos (COSTA; FAVA, 2004).

Neste trabalho será tratada somente recuperação de dados em alto nível de software, devido principalmente à falta de material relacionado a este assunto e ao foco do trabalho. Os programas utilizados na recuperação de dados trabalham a nível de sistemas de arquivos. Normalmente, nesse tipo de *software*, deve-se indicar o sistema de arquivo utilizado na partição que serão recuperados os dados, ou usar o *software* específico para àquele sistema de arquivo (COSTA; FAVA, 2004).

3 | SISTEMAS DE ARQUIVOS

Para que os dados processados pelo computador possam ser utilizados novamente, após o desligamento do sistema, eles precisam ser armazenados pelos dispositivos de armazenamento permanente. Mas para isso, eles precisam de: uma forma de agrupar os dados e um mecanismo de acesso.

Essa forma de agrupar os dados é chamada de “arquivo”. Pode-se definir arquivos, como uma forma de abstração de dados, encapsulando os bits em recipientes, podendo então salvá-los no dispositivo de armazenamento e recuperar novamente as informações mais tarde. O responsável por definir o mecanismo de acesso e a estrutura desses arquivos é o sistema de arquivos. Ele é responsável em lidar com os arquivos (TANENBAUM; WOODHULL, 2000). Geralmente, cada sistema operacional utiliza o seu próprio sistema de arquivos.

3.1 Sistema de arquivos ext3

O sistema de arquivos ext3 foi baseado no ext2 e desenvolvido por Stephen C. Tweedie. Seu grande objetivo não era ser o mais eficiente sistema de arquivos, mas sim, reduzir os problemas causados por desligamentos inesperados. O principal recurso implementado neste sistema foi o *journaling*, que funciona como um *log* de todas as transações realizadas pelo sistema de arquivos, possibilitando a recuperação de um estado anterior caso ocorra alguma falha (PETERSEN, 2008).

Os limites para o sistema de arquivos ext3 são de 16TiB para partições e 2TiB para arquivos. Mas as quatro mais importantes características do sistema de arquivos ext3 são: disponibilidade, integridade dos dados, velocidade e transição fácil do ext2 para ext3 (RED HAT, 2003). Tudo isso implementado pelo *journaling*.

3.2 Sistema de arquivos ext4

O desenvolvimento do sistema de arquivos ext4 foi uma bifurcação do projeto do ext3 (que é chamado na computação de *fork*). O intuito era prover algumas melhorias, sem abrir mão da compatibilidade com seu antecessor, que já era um sistema de arquivos bastante robusto e complexo.

Vários aperfeiçoamentos foram feitos no ext4 em relação ao seu antecessor, como aperfeiçoamentos no *journaling* e a capacidade de trabalhar com partições de até 1EiB e arquivos de até 16TiB (KUMAR, 2008).

De todas as melhorias implementadas, o recurso chamado *extents* foi a principal e mais significativa delas, pois otimiza a gravação dos dados no disco e reduz a fragmentação.

4 | FERRAMENTAS PARA RECUPERAÇÃO DE DADOS

No mercado existem vários sistemas para recuperação de dados, principalmente para ambientes Microsoft Windows. Estas ferramentas possibilitam a recuperação de arquivos apagados seja por descuido, de forma intencional ou por ataques de *malwares*. Até mesmo discos formatados podem ter seus arquivos recuperados por essas ferramentas.

O princípio básico de funcionamento destas ferramentas consiste na varredura do disco por arquivos deletados e sua posterior recuperação (CROSS, 2008). Serão apresentadas algumas ferramentas de recuperação de dados para sistemas de arquivos GNU/Linux, que é foco deste trabalho.

Em Linux, um sistema bastante antigo de recuperação de arquivos deletados é o e2undel. O mesmo não será utilizado como ferramenta de teste deste trabalho, pois só permite a recuperação de arquivos apagados em partições ext2 (DIEDRICH, 2004).

Outro aplicativo que pode ser citado, mas que encontra-se descontinuado, é o ext3grep. O mesmo permitia a recuperação de arquivos em sistemas ext3. Nem mesmo no site do projeto são apresentadas mais informações relevantes sobre a ferramenta.

Para substituir o ext3grep, foi criado o extundelete, usando os mesmos princípios de seu antecessor. De acordo com o site do projeto, o extundelete funciona em sistemas de arquivos ext3 e ext4 e procura por informações contidas no *journal* da partição para tentar recuperar arquivos apagados.

Scalpel e Foremost são duas ferramentas similares para recuperação de arquivos. O Scalpel foi criado a partir da versão 0.69 do Foremost e ambas, apesar de antigas, ainda são bastante utilizadas em sistemas GNU/Linux. Estas ferramentas trabalham com recuperação de arquivos a partir de seus tipos ou assinaturas, como cabeçalhos e rodapés, processo conhecido como *data carving* (FICHERA; BOLT, 2013).

A mais atual das ferramentas de recuperação de dados e que se encontra em pleno desenvolvimento é o Test Disk & Photorec. De acordo com o site do projeto,

são duas ferramentas em um só pacote, onde o Test Disk é utilizado para recuperar partições perdidas ou apagadas e o Photorec utilizado na recuperação de arquivos (desenvolvido inicialmente para recuperação de imagens).

5 | METODOLOGIA

Foram realizados testes com três ferramentas de recuperação de dados. Utilizou-se de duas partições virtuais, uma com sistema de arquivos ext3 e outra com ext4. Arquivos de imagens escolhidos aleatoriamente foram utilizados nas duas partições. Os mesmos arquivos utilizados em uma foram também utilizados na outra.

Três grupos de testes foram realizados: **Testes Grupo 1**, com os arquivos apenas apagados; **Testes Grupo 2**, com as partições formatadas; **Testes Grupo 3**, com os arquivos apagados e as partições parcialmente sobrescritas por outros arquivos. Todas as ferramentas de recuperação realizaram os mesmos procedimentos.

Os arquivos de imagens para a recuperação foram escolhidos aleatoriamente e dispostos em três subdiretórios: JPG, PNG e GIF. Mais detalhes quanto ao tipo, tamanho e quantidade de arquivos, podem ser vistos na Tabela 1.

Tipos de arquivos	Quantidade	Tamanho
.GIF	7	775,7 KB
.JPG	465	804,2 MB
.PNG	48	32,9 MB
TOTAL	520	875,3 MB

Tabela 1: Disposição dos arquivos a serem recuperados

Para manter a confiabilidade dos testes propostos neste trabalho, quatro partições foram criadas em quatro arquivos distintos. Em cada arquivo foi criada uma partição de 2GiB, duas com o sistema de arquivos ext3 e as outras duas com ext4. Ambos os sistemas de arquivos utilizados (ext3 e ext4) foram formatados com as opções padrão do sistema operacional em uso (Ubuntu 14.04), neste caso utilizando o *journaling* no modo *ordered*.

Em seguida, foram copiados para cada uma das partições, os três diretórios contendo os 520 arquivos para, posteriormente, serem recuperados. Feito isso, as quatro partições foram desmontadas para se ter certeza que todos os arquivos foram copiados e não ficaram em *cache*.

As partições foram montadas novamente com seus respectivos sistemas de arquivos e os diretórios com os dados foram excluídos com o comando: **rm -rf ***. Por fim, todas as partições foram novamente desmontadas. Para garantir a integridade dos arquivos apagados nas próximas etapas de recuperação, todo o procedimento foi realizado com as partições desmontadas.

5.1 Realização dos testes

Os testes de recuperação de imagens foram divididos em três grupos, chamados de **Testes Grupo 1**, **Testes Grupo 2** e **Testes Grupo 3**:

- Os **Testes Grupo 1** foram realizados em uma partição ext3 e outra partição ext4 e iniciou-se logo após a remoção dos arquivos das mesmas (nenhuma alteração foi feita nas partições após a exclusão dos arquivos);
- Os **Testes Grupo 2** foram iniciados (nas mesmas partições do teste anterior) com as partições novamente formatadas com seus respectivos sistemas de arquivos (novamente, após a formatação, nenhuma outra alteração foi feita nas partições);
- Os **Testes Grupo 3** foram realizados em uma outra partição ext3 e ext4 (partição diferente dos testes anteriores).

Todas as ferramentas de recuperação propostas neste trabalho foram utilizadas nos três grupos de testes. As ferramentas utilizadas na recuperação foram executadas na seguinte ordem: extundelete, Foremost e Photorec. Nenhum ajuste adicional foi realizado nas ferramentas citadas.

6 | RESULTADOS

Esta seção apresenta os resultados obtidos neste trabalho, comparando as ferramentas testadas quanto a sua eficiência na recuperação e o tempo gasto.

Dos arquivos recuperados, nem todos estavam totalmente legíveis, ou seja, muitos foram parcialmente recuperados ou recuperados, mas com seu conteúdo ilegível. Daqui em diante, os arquivos com seus conteúdos totalmente recuperados serão chamados arquivos íntegros. Os arquivos parcialmente recuperados e corrompidos serão chamados de arquivos corrompidos.

6.1 Recuperação de imagens nos Testes Grupo 1, partição ext3

Após apagados os arquivos de imagens que estavam na partição ext3, os testes com as ferramentas de recuperação de dados foram realizadas e os resultados obtidos são mostrados a seguir:

Teste com a ferramenta extundelete: o extundelete levou 14 segundos no processo de recuperação. Conseguiu recuperar de forma íntegra todos os arquivos apagados, incluindo seus respectivos nomes;

Teste com a ferramenta Foremost: a ferramenta Foremost gastou 49 segundos para finalizar o processo. Conseguiu recuperar todos os sete arquivos gif, 376 arquivos jpg e apenas três arquivos png, totalizando 386 arquivos recuperados. Dos arquivos

recuperados, a maioria deles estavam corrompidos;

Teste com a ferramenta Photorec: com o aplicativo Photorec, o tempo gasto no processo foi de um minuto e um segundo. Recuperou todos os 7 arquivos gif, os 48 arquivos png e dos 465 arquivos jpg, recuperou 831. Isso porque ferramenta recuperou também as miniaturas dos arquivos jpg, sendo 402 arquivos de miniaturas de imagens e 429 arquivos de imagens propriamente ditos. Desprezando-se as miniaturas encontradas (daqui em diante, em todos os testes, todas as miniaturas encontradas pelo Photorec serão desprezadas), recuperou ao todo 484 arquivos, todos eles íntegros.

6.2 Recuperação de imagens nos Testes Grupo 1, partição ext4

Depois de apagados os arquivos de imagens que estavam na partição ext4, os testes com as ferramentas de recuperação de dados foram realizados e os resultados são apresentados a seguir:

Teste com a ferramenta extundelete: neste teste, o extundelete gastou também 14 segundos no processo de recuperação. Para este sistema de arquivos, conseguiu também recuperar de forma íntegra todos os arquivos apagados, incluindo seus nomes;

Teste com a ferramenta Foremost: a ferramenta Foremost gastou 49 segundos para finalizar o processo e foi superior neste sistema de arquivo em relação ao mesmo teste com o ext3. No total, 519 arquivos foram recuperados. Destes, somente quatro estavam corrompidos (todos do tipo gif);

Teste com a ferramenta Photorec: desta vez, com o aplicativo Photorec, o tempo gasto no processo foi de dois minutos e quatro segundos. Ele conseguiu recuperar todos os sete arquivos gif, os 48 arquivos png e dos 465 arquivos jpg, recuperou 438. Todos os arquivos encontrados estavam íntegros. A eficiência desta ferramenta também foi superior no ext4.

6.3 Recuperação de imagens nos Testes Grupo 2, partição ext3

Após apagados todos os arquivos e a partição ext3 novamente formatada, os testes com as ferramentas de recuperação de dados foram novamente realizados e os resultados mostrados a seguir:

Teste com a ferramenta extundelete: o extundelete levou apenas um segundo no processo de recuperação, mas não conseguiu nada;

Teste com a ferramenta Foremost: a ferramenta Foremost gastou 56 segundos para finalizar o processo. Conseguiu recuperar todos os sete arquivos gif, apenas três arquivos png e 375 arquivos jpg. No total, 385 arquivos foram recuperados. Destes, muitos estavam corrompidos;

Teste com a ferramenta Photorec: com o programa Photorec, o tempo gasto no processo foi de um minuto e nove segundos. Recuperou todos os sete arquivos gif, os 48 arquivos png e 429 arquivos jpg. Todos os arquivos recuperados estavam íntegros.

6.4 Recuperação de imagens nos Testes Grupo 2, partição ext4

Depois de apagados todos os arquivos e a partição ext4 novamente formatada, os testes com as ferramentas de recuperação de dados foram realizados e os resultados apresentados a seguir:

Teste com a ferramenta extundelete: neste teste, o extundelete levou menos de um segundo no processo de recuperação e, novamente, não conseguiu recuperar arquivo algum;

Teste com a ferramenta Foremost: a ferramenta Foremost gastou 54 segundos para finalizar o processo. Conseguiu recuperar todos os sete arquivos gif, os 48 arquivos png e 464 arquivos jpg. No total, 519 arquivos foram recuperados. Destes, somente quatro estavam corrompidos (todos do tipo gif);

Teste com a ferramenta Photorec: desta vez, com o aplicativo Photorec, o tempo gasto no processo foi de dois minutos e quatro segundos. Todos os sete arquivos gif, 48 arquivos png e 474 arquivos jpg, foram recuperados. Todos os arquivos encontrados estavam íntegros.

6.5 Recuperação de imagens nos Testes Grupo 3, partição ext3

Após apagados os arquivos de imagens que estavam na segunda partição ext4 e os novos arquivos copiados para a mesma, os testes com as ferramentas de recuperação de dados foram realizadas e os resultados obtidos são apresentados a seguir:

Teste com a ferramenta extundelete: o extundelete levou apenas um segundo no processo de recuperação e não conseguiu recuperar arquivo algum;

Teste com a ferramenta Foremost: o Foremost levou 49 segundos para finalizar o teste. Conseguiu recuperar os sete arquivos gif, apenas dois arquivos png e 212 arquivos jpg. No total, 221 arquivos foram recuperados. Destes, apenas 20 estavam íntegros;

Teste com a ferramenta Photorec: a ferramenta Photorec, desta vez, gastou 14 segundos e recuperou somente 24 arquivos, todos íntegros. Destes recuperados, cinco eram do tipo gif, três png e apenas 16 arquivos jpg.

6.6 Recuperação de imagens nos Testes Grupo 3, partição ext4

Depois de apagados os arquivos de imagens que estavam na segunda partição ext3 e os novos arquivos copiados para a mesma, os testes com as ferramentas de recuperação de dados foram realizados e os resultados obtidos são exibidos a seguir:

Teste com a ferramenta extundelete: o extundelete levou menos de um segundo no processo de recuperação, mas não conseguiu nada;

Teste com a ferramenta Foremost: o Foremost gastou um minuto para realizar

o teste. Conseguiu recuperar os sete arquivos gif, todos os 48 arquivos png de forma íntegra e 412 arquivos jpg. No total, 467 arquivos foram recuperados. Destes, apenas quatro estavam corrompidos, três do tipo gif e um do tipo jpg;

Teste com a ferramenta Photorec: o Photorec gastou 46 segundos e recuperou 189 arquivos. Novamente todos íntegros. Destes, sete eram do tipo gif, 48 png e 134 arquivos jpg.

6.7 Comparações dos resultados obtidos

De acordo com o ambiente utilizado e os resultados dos testes aqui apresentados, pode-se concluir que: **quanto ao tempo gasto**, a ferramenta mais rápida foi o extundelete e a mais lenta, o Photorec; para **recuperação de imagens apenas apagadas (Testes Grupo 1)**, a melhor ferramenta foi o extundelete, tanto para sistemas de arquivos ext3 quanto ext4, pois além de recuperar todos os arquivos íntegros, recuperou também seus respectivos nomes; para **partições que perderam seus dados após serem formatadas (Testes Grupo 2)**, seja em ext3 ou em ext4, em geral, o Photorec apresentou melhor eficiência, com a única exceção de arquivos de imagens do tipo jpg no sistema de arquivos ext4, onde o Foremost conseguiu recuperar 5,59% a mais de arquivos que o Photorec; já nas **partições que tiveram seus arquivos apagados e parcialmente sobrescritos**, em ext3 o Photorec foi melhor; em ext4 o Foremost foi superior.

Os Gráficos 1, 2 e 3 mostram o comparativo geral entre as ferramentas testadas quanto a eficiência das mesmas.

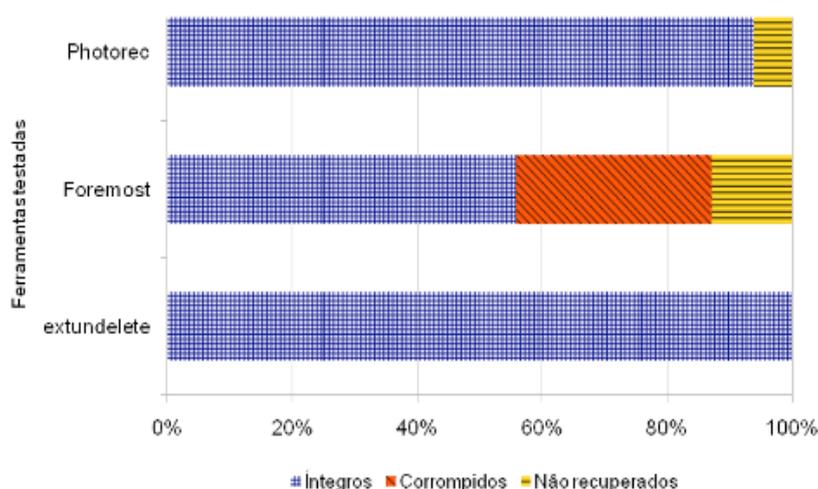


Gráfico 1: Comparativo geral entre as ferramentas nos Testes Grupo 1

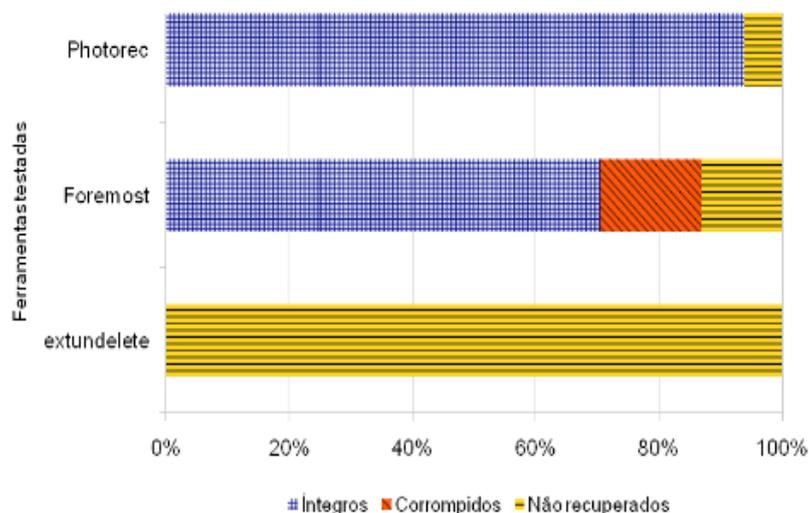


Gráfico 2: Comparativo geral entre as ferramentas nos Testes Grupo 2

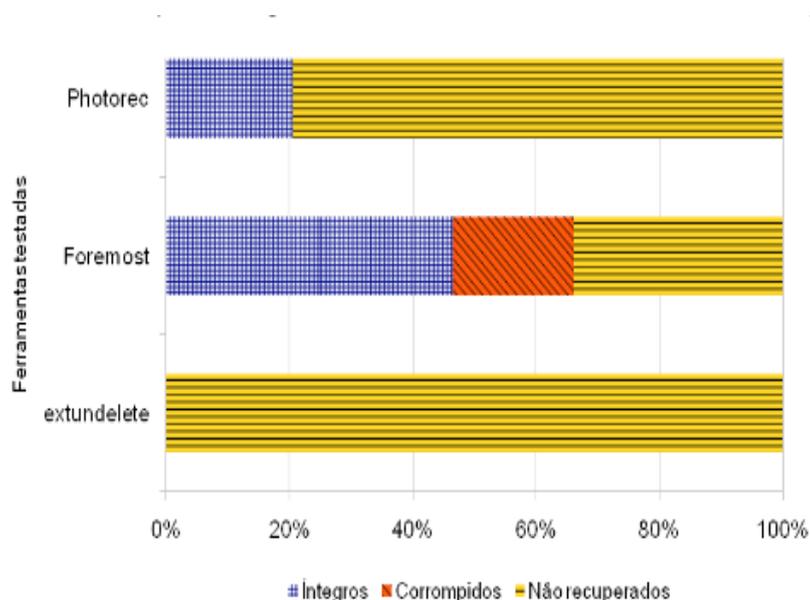


Gráfico 3: Comparativo geral entre as ferramentas nos Testes Grupo 3

6.8 Discussão dos resultados obtidos

Das três ferramentas testadas, duas tiveram resultados parecidos, enquanto uma delas, a extundelete, teve resultados bem diferentes. A diferença dos resultados encontrados com esta ferramenta em relação às demais se deve ao fato deste programa funcionar de forma diferente dos outros testados.

O extundelete trabalha tentando recuperar os arquivos com base nas informações do arquivo de *journal*. Com o *journal* íntegro e a partição “intocada” após a exclusão dos arquivos, todos os arquivos foram facilmente recuperados (inclusive com seus nomes) pela ferramenta. Após a formatação das partições (**Testes Grupo 2**), a ferramenta não contava mais com o arquivos de *journal*, não conseguindo portanto a recuperação dos arquivos. O mesmo ocorreu após a cópia de novos arquivos para as partições (**Testes Grupo 3**).

Portando para partições formatadas o extundelete não conseguiu recuperar arquivo algum, devido à criação de um novo *journal* neste processo. Porém, quanto aos **Testes Grupo 3**, outras simulações podem ser feitas utilizando-se, por exemplo, partições maiores e/ou copiando uma quantidade menor de arquivos para a mesma, a fim de testar o extundelete em outros cenários, avaliando assim sua eficiência em ambientes diferentes.

Já o Foremost e o Photorec trabalham de forma similar. Eles procuram por assinaturas de arquivos no disco (estruturas internas dos arquivos) e tentam recuperá-los, independente das informações contidas no *journal* (*data carving*).

Em todos os testes, o Photorec teve uma regularidade em seus resultados, tanto para sistemas de arquivos ext3 quanto ext4. Das imagens encontradas por esta ferramenta, todas estavam íntegras.

Já o Foremost, principalmente em partições ext3, encontrou muitos arquivos corrompidos. Contudo, em ext4 teve um resultado superior ao Photorec. Em geral, tanto o Foremost quanto o Photorec tiveram desempenho melhor quando utilizados em sistemas de arquivos ext4 em detrimento ao ext3.

7 | CONCLUSÃO

Este trabalho apresentou de forma prática os resultados dos testes entre três ferramentas de recuperação de dados, extundelete, Foremost e Photorec, para sistemas de arquivos ext3 e ext4. Para isso, três ambientes de teste foram criados em quatro partições virtuais, duas com sistema de arquivos ext3 e as outras duas com ext4.

O autor conseguiu cumprir com o objetivo inicial do trabalho, avaliando as três ferramentas no cenário utilizado e apresentado seus resultados. Conseguiu também, o autor, demonstrar que é possível, através de programas *open source*, e custo zero, ferramentas com resultados satisfatórios, seja para usuários ou mesmo profissionais na área.

Como trabalhos futuros, pode-se testar as mesmas ferramentas aqui demonstradas em outros sistemas de arquivos, como o recente Btrfs por exemplo. Como alternativa, pode-se também avaliar outros programas de recuperação de dados nas mesmas situações apresentados ou testar outros tipos de arquivos e não somente imagens.

REFERÊNCIAS

COSTA, R. P.; FAVA S. L. Informática forense: formato de imagens fotográficas digitais e seus reflexos na análise pericial. In: ICCyber – Conferencia internacional de perícias em crimes cibernético, 1, 2004, Brasília/DF - Brasil. **Anais...** SAIS Quadra 07, Lote 21, Ed. INC/DPF: Editora DPF – Departamento de Polícia Federal, 2004. p. 230-236.

CROSS, M. Acquiring Data, Duplicating Data, and Recovering Deleted Files. In: _____. **Scene of the Cybercrime**. 2ª Edição. Syngress Publishing, Inc, 2008, Burlington, MA, USA. cap. 7, p. 315.

DIEDRICH, O. the e2undel home page. Disponível em: <<http://e2undel.sourceforge.net/>>. Acesso em: 14 de jun. de 2014.

FICHERA, J.; BOLT, S. Volatile Data Analysis. In: _____. **Network Intrusion Analysis: Methodologies, Tools and Techniques for Incident Analysis and Response**. Syngress Publishing, Inc, 2013, 225 Wyman Street, Waltham, MA, USA. cap. 4, p. 95.

KUMAR, A.; CAO, M.; SANTOS, J.; DILGER, A. Ext4 block and inode allocator improvements. In: Proceedings of the 2008 Ottawa Linux Symposium, 2008, Ontario, Ottawa, Canadá . **Anais... Ottawa: Linux Symposium**. Disponível em: <<https://ols.fedoraproject.org/OLS/Reprints-2008/kumar-reprint.pdf>>. Acesso em: 14 de jun. de 2014.

PETERSEN, R. Part 3: Devices and File Systems. In: _____. **Red Hat Enterprise Linux 5: Administration, Security, Desktop**. Surfing Turtle Press. 2008. p. 312.

RED HAT, INC. **Red Hat Enterprise Linux 3: Guia de Administração do Sistema**. Disponível em: <http://web.mit.edu/rhel-doc/3/rhel-sag-pt_br-3/ch-ext3.html>. Acesso em: 14 de jun. de 2014.

TANENBAUM, A. S.; WOODHULL, A. S. Sistemas de arquivos. In: _____. **Sistemas operacionais projeto e implementação**. 2ª Edição. Porto Alegre/RS Brasil: ARTMED EDITORA S/A. 2000. cap. 5. p. 271-272.

REDIMENSIONAMENTO DO ENSINO DA COMPUTAÇÃO NA EDUCAÇÃO BÁSICA: O PENSAMENTO COMPUTACIONAL, O UNIVERSO E A CULTURA DIGITAL

Melquisedec Sampaio Leite

Universidade de Pernambuco – Campus
Garanhuns
Garanhuns - Pernambuco

Sônia Regina Fortes da Silva

Universidade de Pernambuco – Campus
Garanhuns
Garanhuns - Pernambuco

RESUMO: Este artigo teve como objetivo investigar como a Computação está sendo redimensionada em educação no mundo e no Brasil, através de análise dos eixos do Ensino de Computação para uma fundamentação de proposta de atividades educativas, que desenvolvam determinadas habilidades e competências, fortalecendo o pensamento computacional, o universo e a cultura digital de crianças e jovens na educação básica. Para isso, foram levantados dados na literatura sobre as questões políticas no Brasil, o pensamento computacional e o universo de cultura digital, relacionando estes dados às habilidades e competências propostas no SAEB e nas Diretrizes Curriculares do Ensino de Computação, da SBC para a educação básica.

PALAVRAS-CHAVE: Ensino de Computação. Pensamento Computacional. Educação Básica. Políticas Educacionais. Computação Educacional.

REDIMENSIONING COMPUTER EDUCATION IN BASIC EDUCATION: COMPUTATIONAL THINKING, THE UNIVERSE AND DIGITAL CULTURE

ABSTRACT: This article aimed to investigate how computing is being resized in education in the world and in Brazil, by analyzing the axes of Computer Teaching for a proposal of educational activities that develop certain skills and competences, strengthening computational thinking, the universe and digital culture of children and young people in basic education. For this, data were collected in the literature on political issues in Brazil, computational thinking and the universe of digital culture, relating these data to the skills and competences proposed in SAEB and in the Curriculum Guidelines of Computer Teaching, SBC for basic education.

KEYWORDS: Computer Teaching. Computational Thinking. Basic education. Educational Policies. Educational Computing.

1 | INTRODUÇÃO

Mudanças nos processos de comunicação e produção do conhecimento vêm ocorrendo nos últimos anos, impactando a educação escolar e motivando a necessidade de se adotar novas formas de ensino. Moran (2007) destaca que a educação não é um

processo apenas da escola, mas de toda a sociedade e que afeta todas as pessoas, seja através da transmissão de ideias, valores, informações dentre outros. Portanto, essa evolução vai além do ambiente institucional escolar, e indica que todos são educadores e aprendizes, pois transmitem e recebem conhecimento.

A popularização da internet vem caracterizando o século XXI, como a sociedade da informação, com avanços técnicos e científicos que provocam mudanças rápidas no modo de vida das pessoas. A internet hoje é indispensável às tarefas cotidianas e na escola. Nesse sentido, Moran (2007, p.9) coloca a seguinte ideia:

Escolas não conectadas são escolas incompletas (mesmo quando didaticamente avançadas). Alunos sem acesso contínuo às redes digitais estão excluídos de uma parte importante da aprendizagem atual: do acesso a informação variada e disponível on-line, da pesquisa rápida em bases de dados, bibliotecas digitais, portais educacionais; da participação em comunidades de interesse, nos debates e publicações on-line, enfim, da variada oferta de serviços digitais.

Nesse contexto, a escola sofre influência do processo de informatização, seja na utilização do computador como ferramenta de ensino-aprendizagem, nos debates sobre questões cotidianas dos alunos com os professores, na necessidade da formação de novos profissionais para novas profissões e até mesmo na gestão administrativa da escola.

Dispositivos computacionais podem ser utilizados no ensino de antigas habilidades que devem ser desenvolvidas na escola, como ler, escrever, realizar operações matemáticas, conhecer os aspectos geográficos de uma região, estudar a história de um lugar, atuar na preservação do meio ambiente e assim por diante. Situações cotidianas, envolvendo todas as ciências, podem ser trabalhadas em sala de aula, com o auxílio do computador o que leva a discussão sobre a inserção da ciência da computação e do pensamento computacional na educação, mais precisamente, na educação básica.

Atualmente, novas habilidades são exigidas dos educandos. Sendo estas: a capacidade de trabalhar em grupo, a capacidade de resolver um problema a partir de um conceito já construído, capacidade de resolver cálculos mentalmente, de associar os assuntos estudados nas disciplinas escolares com o contexto do mundo real. Habilidades e competências que tem relação com o desenvolvimento do pensamento computacional. Este refere-se a resolução de problemas presentes no dia a dia, das mais diversas áreas do conhecimento através de conceitos computacionais.

Busca-se nesta pesquisa investigar como a computação está sendo redimensionada em educação, no mundo, no Brasil e em Pernambuco, através de análise dos eixos do Ensino de Computação para uma fundamentação de proposta de atividades educativas, que desenvolvam determinadas habilidades e competências, fortalecendo o pensamento computacional, o universo e a cultura digital de crianças e jovens na educação básica.

2 | O PENSAMENTO COMPUTACIONAL NO MUNDO E NO BRASIL

Entende-se que o pensamento computacional é natural no ser humano manifestando-se desde a idade infantil. Portanto, desde as idades iniciais, a criança raciocina naturalmente de forma computacional a medida que realiza atividades que seriam algoritmos de forma simultânea mantendo o controle sobre essas tarefas. Entretanto se faz necessário amadurecer esse raciocínio e explorá-lo de forma contextualizada com a faixa etária e os interesses dos educandos na educação básica. Do contrário, com o decorrer do tempo, a criança que não tiver contato com este desenvolvimento do pensamento, terá maiores dificuldades em resolver problemas de maneira intuitiva no futuro.

O propósito do Pensamento Computacional na Educação é fornecer um método para solucionar problemas nas diversas áreas do conhecimento, usando conceitos computacionais. O pensamento computacional possibilita, através de ferramentas e técnicas da Ciência da Computação, meios para resolver diversos problemas do mundo real, o que sugere a formação em Computação de alunos da Educação Básica de acordo com uma abordagem que contempla conceitos, habilidades e novas competências.

Blikstein (2008) define que o pensamento computacional, como a capacidade de saber usar o computador como um instrumento de aumento do poder cognitivo e operacional humano, dessa forma, contribuindo com o aumento da produtividade, inventividade, e criatividade. Ainda afirma que a primeira etapa para se “pensar computacionalmente” é identificar as atividades cognitivas que podem ser feitas de forma mais rápida e eficiente por um computador. A segunda etapa seria transferir o que não é essencialmente humano para o computador, para que este realize as tarefas.

O raciocínio computacional vai muito além do simples uso do computador, ele expande a capacidade do pensamento humano para resolver problemas, mesmo sem a presença da máquina. Já existem trabalhos nesse sentido como, por exemplo, o projeto da Computação Desplugada de Bell e Witten, (1995). Este projeto apresenta alternativas de ensino da Computação por meio de atividades lúdicas e analogias do cotidiano, não utilizando recursos de hardware e/ou software. Este método de trabalho tem despertado interesse de pesquisadores e professores em diversos países. O ensino de Computação sem a necessidade de hardware e/ou software especializados possibilita a realização das práticas pedagógicas em diferentes estruturas escolares.

Segundo Yadav et. all. (2017), em vários países como a Austrália, EUA e Reino Unido, cresceu o entusiasmo nos últimos anos para o ensino de Computação. Em 2012, a Royal Society no Reino Unido colocou o seguinte: Toda criança deve ter a oportunidade de aprender conceitos e princípios da computação, incluindo a ciência da computação e as tecnologias da informação, desde o início do ensino primário em diante, e aos 14 anos podem escolher estudar para obter uma qualificação reconhecida

nessas áreas. Em 2016, o College Board nos EUA lançou um novo currículo de Ciência da Computação para escolas secundárias, chamado “Princípios da Ciência da Computação”, enfocando a exposição dos alunos ao pensamento computacional e práticas para ajudá-los a entender como a computação influencia o mundo.

Em países como Estados Unidos e Canadá, o modelo Model Curriculum for K–12 Computer Science (CSTA K-12) é adotado para o ensino de computação. Segundo as diretrizes do k-12, o aluno deve desenvolver, através do Pensamento Computacional, habilidades para resolver problemas numa forma que pode ser implementada em um computador, envolvendo um conjunto de conceitos, como abstração, recursão, iteração, dentre outras. As diretrizes CSTA para o ensino de computação da educação infantil ao ensino médio são baseadas em um modelo em três níveis (Figura 1). O nível 1 fornece os padrões de aprendizagem para os estudantes do infantil até ao sexto ano, o nível 2 fornece os padrões de aprendizagem para estudantes entre o sexto e o nono ano, e o nível 3 fornece os padrões de aprendizagem para os alunos do ensino médio (WANGENHEIM, Et. All., 2014). Esses limites especificados para cada nível podem variar de escola para escola.

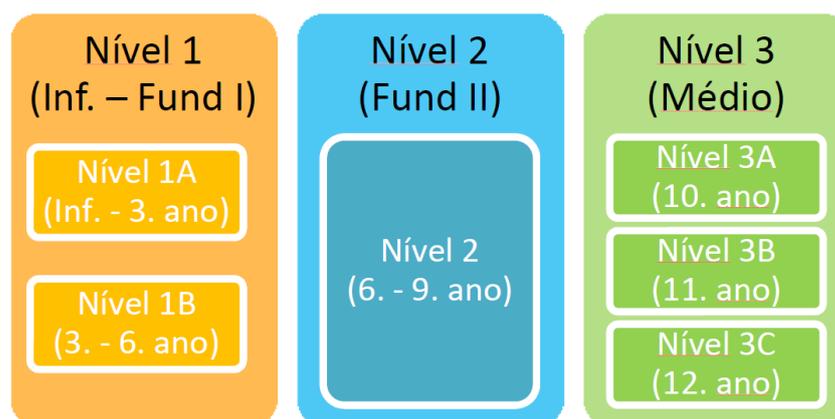


Figura 1: Níveis de Ensino de Computação

No Nível 1, são apresentados os conceitos fundamentais de Ciência da Computação a alunos do ensino fundamental. São incorporadas competências, como: criatividade e exploração, aplicadas a outras áreas curriculares, tais como: ciências sociais, língua, matemática e ciência. No Nível 2, o Pensamento Computacional começa a ser utilizado para resolução de problemas. No Nível 3, os alunos aprendem a dominar conceitos mais avançados de ciência da computação e aplicar esses conceitos para desenvolver artefatos virtuais e reais (WANGENHEIM, Et. All., 2014).

Atualmente as escolas brasileiras não dispõem, em seus currículos, de disciplinas que trabalham a Computação enquanto Ciência, apesar de proporcionarem contato com as máquinas e seus recursos. Entende-se que é de fundamental importância abordar, desde as primeiras séries da educação básica, os princípios da Ciência da Computação a fim de fundamentar a área em suas relações com o cotidiano escolar e da vida, colaborando para o exercício de habilidades cognitivas de raciocínio, abstração

e resolução de problemas e, como um possível campo de atuação docente. Isso não significa que os educandos terão que seguir essa área profissionalmente, mas que poderão utilizar os conhecimentos computacionais para auxiliá-los no exercício de qualquer área do conhecimento.

Além da iniciativa de informatizar as escolas, pode-se dizer que no Brasil os esforços para iniciar as discussões acerca do Pensamento Computacional na Educação Básica começaram com a oferta dos cursos de Licenciatura em Computação – LC. O licenciado em Computação é um professor da educação básica que tem como missão pensar o uso efetivo das tecnologias na escola, utilizando-se de ferramentas como as redes de informação, as tecnologias sociais e o conteúdo digital (CAMBRAIA e SCAICO, 2013).

Diante do exposto, pode-se levantar dois argumentos nesta pesquisa, diante alguns que justifiquem a necessidade de introduzir conceitos de Computação na educação básica. Em primeiro lugar, pelo seu caráter transversal, integrador e interdisciplinar as outras ciências, pois o Pensamento Computacional auxilia de forma significativa na resolução de problemas do dia a dia, relacionados a qualquer área do conhecimento. Outro fator, é que esse conhecimento é necessário para a inserção dos alunos no século XXI, já que independente de sua área de estudo ou da profissão escolhida, precisam entender computação para serem produtivos e competitivos em suas áreas. Além dos fatores já mencionados, pela enorme escassez de profissionais de TI, área que vem crescendo de forma substancial nos últimos anos.

3 | COMPUTAÇÃO NAS DIRETRIZES DA EDUCAÇÃO BÁSICA NO BRASIL

É inegável que a Computação faz parte da vida das pessoas. No contexto escolar, ela ensina métodos para resolver problemas que são relevantes em qualquer área do conhecimento, além de desenvolver o raciocínio lógico e computacional para o desenvolvimento de outras competências e habilidades das diversas áreas do conhecimento. Nos últimos anos, está em curso no Brasil uma reformulação dos currículos da Educação Básica, que engloba a educação infantil e os ensinos fundamental e médio.

O Curso de Licenciatura em Computação, atualmente, é a etapa da educação superior que habilita estudantes para a docência em Computação e Informática, para exercer a profissão para o ensino com Robótica e Computação na educação básica. É um Curso relativamente novo no Brasil, sendo expandido após a promulgação da Lei Nº 9.394, do CNE, em 1996, em decorrência da introdução de Informática e Computação, nos debates nacionais para a formação de competências e habilidades para um mercado em expansão, na década de 80. Inicia sua abrangência na educação em graduação, nos Cursos de Ciências Biológicas, Física e Matemática, bem como na educação básica, em sua última etapa, no ensino médio, no final da década de 90 e início deste século.

A Base Nacional Comum Curricular – BNCC (BRASIL, 2017), é uma exigência colocada para o Sistema Educacional Brasileiro a partir da Lei de Diretrizes e Bases da Educação Nacional (BRASIL, 1996; 2017), como também as Diretrizes Curriculares Nacionais Gerais da Educação Básica (BRASIL, 2009) e Plano Nacional de Educação (BRASIL, 2015), e deve se constituir como um avanço na construção da qualidade da educação.

A BNCC destaca que sua finalidade é orientar os sistemas na elaboração de suas propostas curriculares, e tem como fundamento o direito à aprendizagem e ao desenvolvimento, em conformidade ao que preceituam o Plano Nacional de Educação (PNE) e a Conferência Nacional de Educação (CONAE). A segunda versão, de abril de 2016, apresenta as áreas curriculares, divididas em:

- Linguagens, que se subdivide em Português, Língua Estrangeira, Arte e Educação Física;
- Matemática;
- Ciências da Natureza, que engloba Física, Química e Biologia;
- Ciências Humanas, onde estão História, Geografia, Filosofia e Sociologia;

Nesta versão, a Computação é citada no documento da Base Nacional Comum Curricular em seus temas integradores como “Culturas Digitais e Computação”, ressaltando a necessidade da escola em explorar as potencialidades dos recursos computacionais para o alcance de suas metas. Segundo a BNCC, a escola tem o importante papel de orientar aos jovens, que cada vez mais cedo passam a conviver com as Culturas Digitais, de utilizá-las de forma reflexiva e ética. Além disso, o documento coloca que o professor deve mediar essas atividades e contribuir para que o estudante possa lidar com essas informações.

O contexto atual apresenta aos professores um grande desafio que é utilizar as Tecnologias de Informação e Comunicação - TIC, como meio de construir e difundir o conhecimento. Nesse contexto, a escola deve assumir o papel de universalizar o conhecimento e, nessa perspectiva, as TIC passam a ter papel fundamental nesse processo. A medida que as tecnologias ganham espaço na escola, o professor se vê diante de novas possibilidades de acesso à informação e de abordagem aos conteúdos de forma dinâmica e contextualizada com a realidade. Para isso, é necessário que o professor desenvolva as habilidades necessárias para ser capaz de analisar os meios à sua disposição e proporcionar aos educandos os meios necessários a construção do conhecimento.

Nesse contexto, é indispensável ressaltar algumas competências gerais para o ensino de computação, contidas no SAEB (2011):

- Domínio do conhecimento da área e das ferramentas computacionais;

- Capacidade de realizar trabalhos vinculados com o projeto pedagógico da escola utilizando os conceitos da Ciência da Computação;
- Mostrar aos alunos como abstrair as informações mais importantes num determinado problema, de forma que essas informações sejam significativas para os alunos, permitindo que eles a compreendam e elaborem suas próprias soluções;
- Capacidade de coordenar atividades em que se organize a turma em grupos, desenvolvendo nos alunos a capacidade de trabalhar em cooperação uns com os outros, a capacidade de liderança, expressão, interação e troca de informações.

O SAEB é o Sistema de Avaliação da Educação Básica, conforme estabelece a Portaria N.º 931, de 21 de março de 2005, sendo composto por dois processos: a Avaliação Nacional da Educação Básica (Aneb) e a Avaliação Nacional do Rendimento Escolar (Anresc). Em novembro de 2015, o Instituto Nacional de Estudos e Pesquisas Educacionais (Inep) realizou a 13ª edição desse sistema de avaliação. Participaram do Saeb 2015 todas as escolas públicas brasileiras com, pelo menos, 20 estudantes matriculados no 5º ou 9º anos do Ensino Fundamental, de acordo com o Censo da Educação Básica 2015. Além desse conjunto de escolas, participou também uma amostra de escolas privadas com 10 ou mais estudantes matriculados no 5º ou 9º anos do Ensino Fundamental ou na 3ª série do Ensino Médio, bem como uma amostra de escolas públicas municipais e estaduais com 10 a 19 alunos matriculados no 5º e 9º anos do Ensino Fundamental e uma amostra de escolas públicas estaduais e municipais com 10 ou mais alunos matriculados na 3ª série do Ensino Médio.

Para a elaboração dos itens do SAEB, são associados os conteúdos da aprendizagem com as competências e habilidades utilizadas na construção do conhecimento. Sinteticamente, pode-se enumerar os seguintes:

- Capacidade de viver em sociedade, atuando, de maneira adequada e relevante, nas mais diversas situações sociais de comunicação;
- Capacidade de resolver um problema a partir da utilização/aplicação de um conceito já construído;
- Observação, estabelecimento de relações, comunicação (diferentes linguagens), argumentação e validação de processos, além de estimular formas de raciocínio como intuição, dedução e estimativa.

Diferente de anos atrás, hoje, os alunos têm acesso muito mais fácil às informações tornando as aulas expositivas menos interessantes. Portanto, a aprendizagem intermediada pelo computador proporciona mudanças nessa realidade, onde o conhecimento não é construído apenas pelo antigo modelo educacional, mas através da navegação em diferentes espaços de informação, troca de conhecimentos, trabalho em grupo e relação das disciplinas estudadas com a vida cotidiana.

4 | ESTRATÉGIA DE AÇÃO

Esse estudo foi realizado por meio de leituras em documentos e produções literárias, como também, através de um exercício realizado com uma turma do V Período de Licenciatura em Computação, em encontro presencial, em que se trabalhava o currículo e a análise da Proposta de Ensino de Computação da Sociedade Brasileira de Computação – SBC.

A estratégia foi a aplicação de um exercício sobre os eixos da Proposta da SBC, buscando saber a compreensão dos estudantes, futuros professores de Computação sobre as competências e habilidades que poderiam ser ensinadas na educação básica. Estiveram presentes 20 graduandos, que para esse estudo, foram divididos em 7 grupos, sendo 6 constituídos por 3 membros e um grupo com dois. Esses eixos, foram colocados aos graduandos de acordo com o quadro a seguir.

Pensamento Computacional	Ecossistema Digital	Cultura Digital
<p>Abstração: selecionar e utilizar modelos e representações adequadas para descrever informações e processos, e técnicas para construir soluções algorítmicas;</p> <p>Automação: descrever as soluções de forma que máquinas possam executar partes ou todo o algoritmo proposto, construir modelos computacionais de sistemas complexos;</p> <p>Análise: analisar criticamente os problemas e soluções para identificar não somente se existem que podem ser automatizadas, mas também avaliar a eficiência e a correção destas soluções.</p>	<p>Codificação: entender como informações podem ser descritas e armazenadas;</p> <p>Processamento: compreender como a informação é processada por computadores e os diferentes níveis de relação entre hardware e software;</p> <p>Distribuição: entender como se dá a comunicação entre diferentes dispositivos digitais, como os dados são transmitidos, como é garantida a integridade e segurança no mundo digital, entender a estrutura da internet.</p>	<p>Computação e Sociedade: compreender o impacto e decorrências da revolução digital e dos avanços do ecossistema digital na humanidade;</p> <p>Fluência tecnológica: utilizar de forma eficiente e crítica ferramentas que auxiliem a obter, analisar, sintetizar e comunicar informações de formatos e com fins diversos;</p> <p>Ética Digital: analisar de forma crítica, questões éticas e marais que surgiram com o ecossistema digital.</p>

Quadro 1. Propositivo ao Estudo – Eixos do Ensino de Computação da SBC

De posse dessa informação, os graduandos em LC elaboraram propostas e atividades para o ensino de Computação nas etapas da Educação Básica. Em seguida, foi feita a análise das respostas dadas pelos participantes, confrontadas com a proposta da SBC, destacando as aproximações e os afastamentos observados na compreensão dos estudantes.

5 | RESULTADOS E ANÁLISE

Habilidades e Competências Conforme os Eixos da Proposta da SBC

Durante a realização do exercício, a discussão pôde ser aberta ou fechada, tendo alguns grupos interagidos, apresentando seus pensamentos. A interação foi

como uma reunião de professores, sugerindo e discutindo as possibilidades e limites das atividades propostas.

Os participantes contribuíram para este estudo, num primeiro momento, com propostas de competências e habilidades para o Ensino de Computação, conforme os eixos das propostas da SBC:

Educação Básica	Pensamento Computacional	Ecossistema Digital	Cultura Digital
Educação Infantil	<p>A possibilidade de trabalhar com a abstração no processo de construção do conhecimento, pois entende-se que os alunos já são capazes de entender um processo, sem se aprofundar em como funciona;</p> <p>Aprendizagem através da análise e compreensão de problemas e criação de possíveis soluções (algoritmos simples).</p>	<p>Entender e vivenciar a Comunicação Digital;</p> <p>Entender e diferenciar hardware e software;</p> <p>Compreender como as informações são escritas e armazenadas;</p> <p>Interagir com dispositivos móveis e executar funções simples (jogar, digitar, etc.).</p>	<p>Entender a relação entre Computação e Sociedade através da comparação entre passado e presente por meio de imagens;</p> <p>Identificar os recursos digitais, suas finalidades e importância;</p> <p>Explorar o contexto histórico da evolução dos computadores.</p>
Ensino Fundamental I	<p>Elaborar, entender e resolver problemas através de ferramentas computacionais;</p> <p>Desenvolver a lógica;</p>	<p>Compreender as funcionalidades de hardware e software;</p> <p>Utilizar as funções básicas;</p> <p>Compreender como as informações são processadas pelos computadores.</p>	<p>Compreender os impactos atuais e futuros dos avanços tecnológicos;</p> <p>Debater de forma crítica as questões éticas que envolvem os avanços tecnológicos;</p> <p>Explorar o uso eficiente de ferramentas digitais.</p>
Ensino Fundamental II	<p>Trabalhar com abstração, automação, codificação e processamento;</p> <p>Analisar padrões e entender o processo de reconhecimento.</p>	<p>Compreender como as informações são processadas em diferentes níveis;</p> <p>Representar as atividades que hardware e software desempenham;</p> <p>Explorar a linguagem de máquina através do código binário.</p>	<p>Argumentar sobre questões éticas e morais através de exemplos de casos reais ocorridos na sociedade;</p>

<p>Ensino Médio Regular e Técnico</p>	<p>Realizar e estabelecer modelos e representações adequadas para descrever e construir algoritmos e suas soluções. No caso do EMT, adiciona-se a construção de software, aplicativos ou páginas web para representar estas soluções; Associar o Pensamento Computacional com outras ciências, como matemática, física, biologia, dentre outras; Utilizar o PC para resolver problemas reais de diversas áreas do conhecimento.</p>	<p>Compreender como funciona a comunicação entre dispositivos digitais, de forma a reconhecer que dispositivo utilizar para cada situação; Trabalhar com robótica, principalmente no EMT; Desenvolver softwares simples, com ferramentas lúdicas, a exemplo do Scratch, no caso do EMR e com ferramentas mais complexas e linguagens como Java, Prolog e outras no caso do EMT; Entender e realizar atividades sobre redes de computadores e segurança de sistemas de forma básica no EMR e aprofundada no EMT.</p>	<p>Discutir o impacto tecnológico na sociedade e fazer analogia entre passado, presente e perspectiva de futuro; Discutir a utilização ética das tecnologias e da internet;</p>
---------------------------------------	---	--	--

Quadro 2. Propostas de Competências e Habilidades

Considerou-se que os participantes apresentaram uma perspectiva relacional com o que se espera do ensino de computação, neste momento em que se discute a sua inserção não só no Curso Técnico, mas também nas demais etapas da educação básica.

As Redes Estaduais de Ensino Técnico orientam Cursos de Informática, de Redes de computadores, de Manutenção e Suporte de Informática, Hospedagem em redes de Computadores, Logística e Redes de Computadores, Informática para Internet, Programação de Jogos digitais e Multimídias, nas regiões do estado de Pernambuco. Nas Escolas Técnicas Federais, encontram-se o Curso de Informática e de manutenção e Suporte de Informática, Operador de computador, Computação Gráfica, Hospedagem, Informática para Internet, Manutenção e suporte em Informática, Redes de Computadores, com a malha curricular predominantemente de Computação e, em todos os Cursos a disciplina de Informática Básica.

Nos Estágios Obrigatórios constatou-se que os estudantes, ao escolherem

qualquer curso técnico, demonstra alta dificuldade em Informática Básica, necessitando de reforço ou nivelamento nesta área, a fim de que tenham êxito nos cursos. Isto fundamenta a necessidade deste conhecimento nas etapas da educação básica, anteriores ao ensino médio, e sua importância em todas as escolas públicas e privadas.

Atividades Relacionadas aos Eixos das Propostas da SBC

A segunda parte do estudo consistiu nas propostas de atividades, pelos mesmos graduandos de LC que sugeriram as competências. Essas atividades serão aqui enumeradas, de acordo com seu respectivo Eixo:

Eixos	Atividades
Pensamento Computacional	<p>Computação desplugada, respeitando as competências para cada nível da Educação Básica;</p> <p>Expor problemas do mundo real, abordando inclusive outras ciências, solicitando que os educandos abstraíam e apresentem soluções diversas;</p> <p>Apresentar problemas de acordo com os critérios do tópico anterior para serem resolvidos através de algoritmos;</p> <p>Por meio de Gamificação e PBL, elaborar problemas para serem resolvidos em equipes, de forma computacional e não computacional, gerando discussões e fortalecendo o trabalho em equipe, além de desenvolver o conceito de liderança;</p> <p>Realizar atividades lúdicas, como teatro por exemplo, representando processos de um programa, sistema operacional, protocolos de internet, dentre outros;</p> <p>Problemas matemáticos envolvendo lógica;</p>
Ecossistema Digital	<p>Atividade de debate e pesquisa sobre os meios de comunicação e transmissão existentes e a evolução dos mesmos;</p> <p>Atividade comparativa para fomentar o entendimento e exposição comparando hardware e software de maneira lúdica, inclusive levando para a sala de aula, componentes de hardware, como monitores, placas, memórias e outros artefatos, além de programas e aplicativos para representar os softwares;</p> <p>Caracterizar como funciona o sistema de redes de computadores, especificando a sua evolução;</p> <p>Simulação, através de software específico, para conexão de redes de computadores;</p> <p>Estudo prático utilizando estrutura de dados, onde o aluno terá de criar um relatório com os códigos e com suas palavras definir para que o código serve e comentar o mesmo;</p> <p>Em grupos, os alunos deverão criar textos, convertê-los para binários, trocar com outros grupos e realizar o processo inverso;</p>
Cultura Digital	<p>Fazer uma linha do tempo a partir da definição de ética digital mostrando a evolução digital;</p> <p>Mostrar, de forma lúdica, como utilizar internet de forma segura;</p> <p>Realizar análise de casos polêmicos, encontrados na internet, sobre ética digital;</p> <p>Criar uma tabela associativa entre crimes digitais e os do mundo físico, disponibilizando em redes sociais e realizando reflexões e debates sobre os mesmos;</p> <p>Atividades com softwares como Scratch, Robomind, para construção de jogos e outras soluções;</p> <p>Realizar pesquisa para conhecer softwares de automação.</p>

Quadro 3. Propostas de Atividades

Esse trabalho também tem o intuito de demonstrar que habilidades e

competências são necessárias para o ensino de Computação ou Computação Educacional. Habilidades estas que devem estar presentes no dia a dia dos professores, como a capacidade de realizar trabalhos vinculados com o projeto pedagógico da escola, utilizando os conceitos da Ciência da Computação, com a capacidade de demonstrar aos educandos como abstrair as informações mais importantes num problema e a capacidade de formar cidadãos que compreendam a realidade social em que vivem, desenvolvendo a capacidade de trabalhar em cooperação uns com os outros, a capacidade de liderança, expressão, interação e troca de informações.

6 | CONSIDERAÇÕES FINAIS

Diante de tão poucos trabalhos sobre esse assunto, considerou-se esta proposta mais um passo em direção a disseminação do Pensamento Computacional na Educação Básica, tornando-o acessível à comunidade escolar. O computador é necessário para qualquer pessoa e em todos os momentos da vida. Sua inserção na escola traz vários aspectos positivos e contribui para o processo de construção cognitiva do educando e o prepara para o atual e futuro mercado de trabalho.

Dentre as necessidades identificadas para trabalhos futuros, destaca-se a elaboração de método avaliativo para mensurar o nível de assimilação alcançado pelos educandos quando estiverem trabalhando o Pensamento Computacional.

Sabe-se que muitos dados ainda se necessita buscar na literatura internacional e nacional sobre o ensino de computação em educação básica, pois algumas já existem, mas de forma ampla e não direta. Não se pode negar que as tecnologias são prioridades no país, que possui suas políticas de educação e tecnologias. No entanto, reconhecer efetivamente, seria ter uma política de aumento do quadro docente das escolas, dando maior oportunidade a população de competir no mercado de trabalho, com um conhecimento que é eliminatório em atividades avaliativas de emprego.

Neste sentido, esta pesquisa deixa reflexões e contribui para demonstrar que o Ensino de Computação é uma realidade, caso deixem que isto aconteça, mas que se tem que lutar para a sua efetivação, seja estudante de LC, comunidade escolar, sociedade para que se alcance as políticas educacionais dos Estados.

REFERÊNCIAS

MATO GROSSO, Secretaria Estadual de Educação, Esporte e Lazer. (2016) **O professor frente às novas tecnologias de informação e comunicação**. <http://www.seduc.mt.gov.br/Paginas/O-professor-frente-%C3%A0s-novas-tecnologias-de-informa%C3%A7%C3%A3o-e-comunica%C3%A7%C3%A3o.aspx>, Setembro 2017.

KENSKI, Vani Moreira. (2013) **Tecnologias e tempo docente**. São Paulo: Papirus Editora.

MORAN, José Manuel. (2007) **A Educação que desejamos novos desafios e como chegar lá**. São Paulo: Papirus Editora.

BLIKSTEIN, Paulo. (2016) **O Pensamento Computacional e a reinvenção do computador na educação**.http://www.blikstein.com/paulo/documents/online/ol_pensamento_computacional.html, Setembro.

BRASIL. Ministério da Educação. (2016) **PDE : Plano de Desenvolvimento da Educação : SAEB : ensino médio : matrizes de referência, tópicos e descritores**. Brasília : MEC, SEB; Inep, 2008.

http://portal.mec.gov.br/dmdocuments/saeb_matriz2.pdf, Setembro 2017.

Ministério da Educação. (2017) **Sistema de avaliação da educação básica**. http://download.inep.gov.br/educacao_basica/saeb/2017/documentos/projeto_basico_SAEB_2017_V6.pdf, Março.

MEC. (2017) **Base Nacional Comum Curricular**. 3ª versão. Brasília. <http://basenacionalcomum.mec.gov.br/images/BNCCpublicacao.pdf>, Junho.

(2017) Lei Nº 9394. http://www.planalto.gov.br/ccivil_03/leis/L9394.htm, Maio.

(2017) **PNE**.http://pne.mec.gov.br/images/pdf/pne_conhecendo_20_metas.pdf, Junho 2017.

BELL, Tim; WITTEN, Ian H.; FELLOWS, Mike. (2016) **Computer Science Unplugged Ensinando Ciência da Computação sem o uso do computador**.<http://csunplugged.org/wp-content/uploads/2014/12/CSUnpluggedTeachers-portuguese-brazil-feb-2011.pdf>, Setembro.

GERHARDT, Tatiana Engel; SILVEIRA, Denise Tolfo (Orgs.). (2009) **Métodos de Pesquisa**. Porto Alegre: Editora da UFRGS.

YADAV, Aman. Et. All.. (2017) **Computational Thinking for Teacher Education. COMMUNICATIONS OF THE ACM**.

<https://cacm.acm.org/magazines/2017/4/215031-computational-thinking-for-teacher-education/fulltext>, Junho.

TUCKER, Allen, Et. All. (2017) **A Model Curriculum for K–12 Computer Science. CSTA**. https://www.acm.org/education/education/curric_vols/k12final1022.pdf, Junho.

RIBEIRO, Leila. (2016) **Computação na Educação Básica. Comissão de Computação na Educação Básica da SBC**. Porto Alegre: UFRGS.

WANGENHEIM, Christiane Gresse von, Et. All. (2014) **Resumo de Objetivos de Aprendizagem de Computação no Ensino Fundamental (Currículo de Referência CSTA/ACM K-12)**. Santa Catarina. UFSC.

CAMBRAIA, Adão C.; SCAICO, P. D. (2013) **Os desafios da Educação em Computação no Brasil: um relato de experiências com Projetos PIBID no Sul e Nordeste do país**. Maringá: Revista Espaço Acadêmico – Nº 148, Setembro.

REALIDADE VIRTUAL, UTILIZANDO DAS MELHORES PRÁTICAS DA INTERAÇÃO HUMANO COMPUTADOR

Bruno Moreira Batista

Universidade Estadual de Goiás (UEG)
Formosa – GO

Guiliano Rangel Alves

Universidade Estadual de Goiás (UEG) –
CCET
Anápolis – GO

Hellen Corrêa da Silva

Universidade Estadual de Goiás (UEG) –
CEAR
Anápolis – GO

Rhogério Correia de Souza Araújo

Universidade Estadual de Goiás (UEG) –
CEAR
Anápolis – GO

RESUMO: A realidade virtual refere-se à interação de seres humanos com sistemas computacionais como um espaço de comunicação, Interação Humano Computador (IHC). A interação é um sistema que engloba um conjunto de entidades que se comunicam, estas entidades podem ser um humano presente no sistema, uma interface promovendo uma função comunicativa emitindo signos para o observador, ou pode ser algo totalmente computacional, isto é, algum artefato cuja representação e comportamento são comandados pelo computador. Para entender a interface como um espaço de comunicação

utiliza-se da Semiótica, pois o homem interpreta o mundo através dos signos. O processo de estruturação e design de interfaces tem sido centrado no usuário, incorporando questões relacionadas diretamente a modelos cognitivos do processamento humano.

PALAVRAS-CHAVE: Semiótica, IHC, Designer.

ABSTRACT: Virtual reality refers to the interaction of human beings with computer systems as a communication space, Human Computer Interaction (HCI). The interaction is a system that encompasses a set of entities that communicate, these entities can be a human present in the system, an interface promoting a communicative function emitting signs for the observer, or it can be something totally computational, that is, some artifact whose representation and behavior are controlled by the computer. In order to understand the interface as a communication space, Semiotics is used because man interprets the world through signs. The process of designing and designing interfaces has been user-centered, incorporating issues directly related to cognitive models of human processing.

KEYWORDS: Semiotics, HCI, Designer.

1 | INTRODUÇÃO

Este trabalho apresenta o processo cognitivo e semiótica associada à Interação Humano Computador (IHC) e à Realidade Virtual. O mercado tecnológico exige novas tecnologias, pensadas e projetadas e que renovem a interação do homem com as máquinas. A realidade virtual está conquistando seu lugar neste mercado, facilitado pela semiótica que estuda os signos e como se relacionam.

Em tempos de mídias digitais interativas, em que há uma profusão de linguagens, tem-se a necessidade de verificar como ocorre o processo de criação de signos, capaz de gerar novas significações e, assim, a semiótica é essencial nesse processo.

O objetivo deste trabalho foi abordar a interação humana com o computador e como é influenciada pelo processo cognitivo dos fenômenos produtores de significado para esta interação com a realidade virtual.

2 | MATERIAL E MÉTODOS

A metodologia utilizada foi o método indutivo com pesquisa bibliográfica documental e mídias eletrônicas relacionada ao tema. Lakatos e Marconi (2006), afirmam que “a pesquisa bibliográfica consiste em um apanhado geral sobre trabalhos de grande importância já realizados e disponibilizados”. As etapas são: Introdução, Metodologia, Fundamentação Teórica, Considerações Finais e Referências Bibliográficas.

3 | RESULTADOS E DISCUSSÃO

Cada vez mais computadores estão no dia a dia das pessoas, utilizados não apenas como ferramenta de trabalho, mas também como meio de comunicação, de inclusão social e de planejamento familiar, assim a qualidade da interface é fundamental para que sistemas interativos possam ser utilizados com sucesso.

Semiótica é o estudo dos signos, que são criações que se constituem e se transformam como linguagens diversas ou meios de comunicação entre cognição e meio sociocultural e por isso, a semiótica é uma lógica que compreende, interpreta e configura conceitos, constitui significação e o sentido de todo e qualquer fenômeno. A compreensão do funcionamento de um processo de interpretação pressupõe o entendimento do que significa o signo – a unidade principal da semiótica.



Figura 1: Signos

Fonte: <https://newtrade.com.br/tecnologia/60045/>

A Engenharia Semiótica combina com a perspectiva do computador como mídia definida por Kammersgaard (1988), na qual “o computador é visto como um meio pelo qual, humanos se comunicam, encontrando a comunicação nessa perspectiva os níveis de expressão e significado”, os quais podem ser individualizados para cada sujeito, conforme o meio sociocultural, histórico e cognitivo deste.

Souza (2010) expõe que a distinção mais notável da Engenharia Semiótica comparada a outras teorias de IHC é desenvolver a sugestão que Kammersgaard propôs em 1988, de acordo com pressupostos por Souza (2010), “dois níveis de comunicação podem ser identificados neste processo, no nível 1, a interface é um meio de expressão unidirecional através do qual são enviadas mensagens do designer para o usuário, no nível 2, ocorre a troca de mensagens entre o usuário e a própria interface. Este nível caracteriza a interface como um artefato de intercomunicação, uma vez que ela própria será um emissor e receptor de mensagens”.

Para Andrade (2007), “a IHC envolve os processos relativos à comunicação entre usuários e computador, assistindo tanto os aspectos físicos, psíquicos, sociais, de trabalho, entre outros. A experiência de IHC é individual e única, cada pessoa é única em seu conhecimento e experiência”. Por isso é necessário que o designer tenha uma interpretação diferente do sistema quando comparada com a interpretação do usuário.

Comunicação designer-usuário, através do sistema

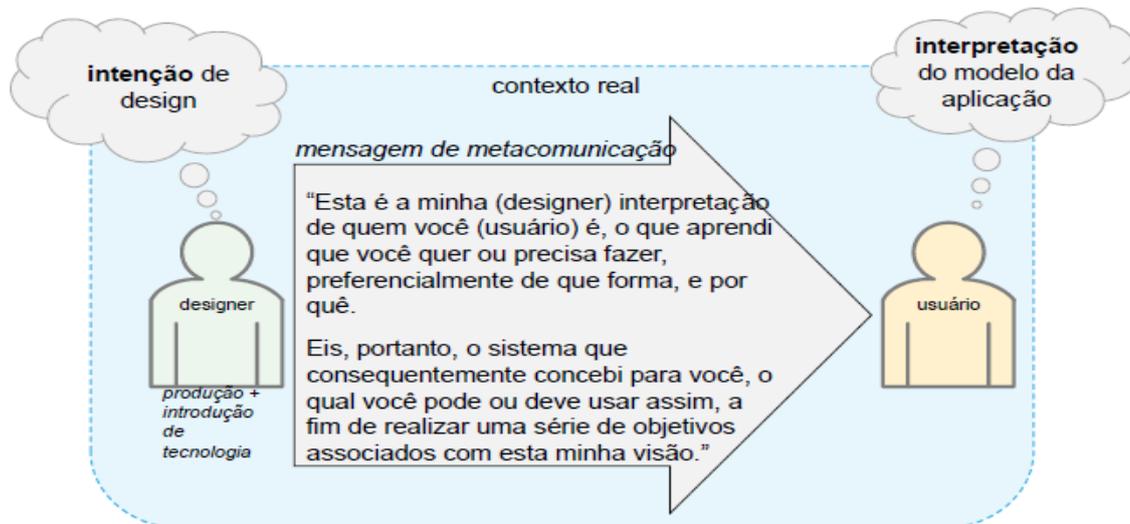


Figura 2: Engenharia Semiótica

Fonte: https://webserver2.tecgraf.puc-rio.br/~abraposo/inf1403/INF1403_10_MISeMAC.pdf

Conforme LABIUTIL (2013), “o sistema cognitivo humano é caracterizado por informações simbólicas, o indivíduo elabora e trabalha sobre a realidade por representações que monta a partir de uma realidade. Esses modelos, que condicionam seu comportamento constituem a sua visão da realidade, que é modificada pelo que é funcionalmente significativo para ele”.

Com a sofisticação da computação gráfica e da criação de um designer criou-se os ambientes virtuais que parecem reais promovendo assim uma realidade virtual consoante à subjetividade do indivíduo. No início os ambientes virtuais eram usados para aplicações científicas e a simuladores de alto custo, hoje são amplamente empregados em arquitetura, medicina, entretenimento, educação, psicoterapia, comunicação, etc., devido ao baixo custo dos recursos computacionais.

Ambientes virtuais conhecidos também por Realidade Virtual são ambientes computacionais de Interação Humano-Computador em que os signos aos quais a pessoa é exposta são gerados pelo computador de maneira relativa à sua consciência e à sua cognição e assim direciona a trajetória pelo seu pensar individual, ou seja, é um ambiente no qual os sentidos humanos são envolvidos em estímulos gerados pelo computador criando imagens e sons, os quais são exibidos sob o ponto de vista da pessoa, permitindo que se movimente neste ambiente como se movimentava em seu dia a dia.

A preocupação com designer da interação do ser humano em ambientes virtuais é relativamente recente. Tendo como pano de fundo a Semiótica que advém do entendimento destes ambientes, da adaptação ao design como signo considerando os aspectos psicológicos, físicos e cognitivos, não descuidando da comodidade, segurança e satisfação do utilizador diante do que é apresentado.



Figura 3: Angry Bird VR Rock in Rio

Fonte: <https://geracaogamer.com/2015/05/20/ignation-studios-desenvolve-angry-birds-em-realidade-virtual-para-rock-in-rio-usa/>

4 | CONSIDERAÇÕES FINAIS

O uso da semiótica pelos designers na criação de simbolismos (como caixas de texto, botões, menus, ícones, dentre outros) deve ser aprimorado a cada dia, a fim de atender a capacidade de interpretação do usuário, que por sua vez não é apenas uma pessoa que faz uso da interface é um elemento que participa do ambiente imerso em uma entidade. Nesse contexto, ressalta-se a importância das ações semióticas e as particularidades cognitivas e socioculturais deste sujeito evidenciando que conforme as necessidades evoluem, os usuários finais precisam entender e dominar práticas de computação. A utilização adequada quanto a cognição humana aumenta a eficácia e a eficiência de uma interface que pode representar melhor o mundo real.

REFERÊNCIAS

ANDRADE, A. L. **Usabilidade de Interfaces web. Avaliação Heurística no Jornalismo On-line.** Rio de Janeiro: E-pappers, 2007. 142p.

KAMMERSGAARD, John. **Four different perspectives on Human-Computer Interaction.** International Journal of Man-Machine Studies, 1988.

LABIUTIL. **Fundamentos Teóricos da Engenharia da Usabilidade.** Disponível em: <http://www.labiutil.inf.ufsc.br/hiperdocumento/unidade2_2_1.html>. Acesso: 04/06/2018.

LAKATOS, E. M.; MARCONI, M. A. **Fundamentos de Metodologia Científica.** 6ª ed., São Paulo: Atlas, 2006.

SOUZA, Clarisse Sieckenius de. **A Enciclopédia da Interação Humano-Computador.** 2ª Ed., 2010. Disponível em <<https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/semiotics>>. Acesso: 10/05/2018.

ORGANIZAÇÃO DO CONHECIMENTO PARA A MEMÓRIA EMPRESARIAL: UM RELATO TÉCNICO SOBRE A EXPERIÊNCIA DO SEBRAE/RJ

Leandro Pacheco de Melo

Mestre em Biblioteconomia – Universidade
Federal do Estado do RJ – UNIRIO

Gestor de Conhecimento, Informação e Memória
Organizacional do Sebrae/RJ

Rio de Janeiro – RJ

ORGANIZATION OF KNOWLEDGE FOR
BUSINESS MEMORY:

A TECHNICAL REPORT ON THE
EXPERIENCE OF SEBRAE/RJ

RESUMO: Este trabalho analisa o processo de organização e representação do conhecimento produzido pelo Sebrae/RJ com vistas à caracterização da sua memória empresarial. A busca por uma identidade representativa no âmbito das micro e pequenas empresas no Estado do Rio de Janeiro norteia o relato e discussão das práticas e experiências vividas na produção de conhecimento ao longo dos 40 anos de existência do Sebrae/RJ. Propõe a criação de um Centro de Memória do Sebrae/RJ estabelecendo parâmetros para a construção de um modelo de Sistema de Recuperação da Informação para a Memória Empresarial.

Palavras-chave: Gestão do Conhecimento, Representação do Conhecimento, Memória Empresarial, Centro de Documentação, Sebrae/RJ.

ABSTRACT: This paper analyzes the process of organization and representation of knowledge produced by Sebrae/RJ in order to characterize its corporate memory. The search for a representative identity in the context of micro and small businesses in the state of Rio de Janeiro guides the reporting and discussion of practices and experiences in the production of knowledge throughout the 40 years of the Sebrae/RJ. It proposes the creation of a Sebrae/RJ Memory Center establishing parameters for the construction of a model of Information Retrieval System for Corporate Memory.

KEYWORDS: Knowledge Management, Knowledge Representation, Corporate Memory, Documentation Center, Sebrae/RJ.

1 | INTRODUÇÃO

Segundo Cortez (1987), um Centro de Documentação Empresarial procura satisfazer as necessidades da empresa a que se vincula, considerando o seu campo de ação e consequentes objetivos. Ainda nas definições, o Centro de Documentação deve atuar como “coordenador das atividades de captação,

catalogação, absorção, recuperação e divulgação das informações estratégicas” (CORTEZ, 1987, p. 40).

Daí a importância de existirem os grandes Centros Informacionais. Procurando suprir suas necessidades de armazenamento adequado e para recuperação eficaz das suas informações decisórias, o Serviço Brasileiro de Apoio às Micro e Pequenas Empresas no Estado do Rio de Janeiro (Sebrae/RJ) inaugurou em 1998 o Centro de Documentação e Informação (CDI), cujo papel será aqui discutido, com vistas à ampliação de seu escopo, com a proposta de transformá-lo em um CDIM – Centro de Documentação, Informação e Memória.

Desde sua fundação em 1977, o Sebrae/RJ atua no desenvolvimento e na competitividade das micro e pequenas empresas no Estado do Rio de Janeiro, assessorando-as nas áreas de economia, administração, finanças, jurídica e tecnológica. As suas ações são executadas por programas e projetos em parceria com diversas instituições, programas e projetos que, por sua vez, geram documentos de conteúdo técnico.

Atualmente, grande parte desses documentos estão armazenados no CDI do Sebrae/RJ aguardando a definição de diretrizes e parâmetros. Outra parte ainda se encontra em poder de antigos funcionários e ex-funcionários – tanto documentos quanto recordações de experiências vividas na instituição.

Isso levanta uma questão. Como o conhecimento – formal e informal – pode ser preservado dentro das organizações? Para responder, buscaremos abordar o assunto tanto sob a perspectiva da Memória Empresarial, como da Gestão da Informação e do Conhecimento, e da Organização do Conhecimento. Alinharemos as técnicas da Biblioteconomia aos interesses da instituição para que o CDI se transforme em um centro de referência em empreendedorismo no Estado do Rio de Janeiro e agrupe, trate e divulgue todo o conhecimento gerado pela empresa nas últimas quatro décadas, criando assim o Centro de Documentação, Informação e Memória (CDIM).

O presente relato técnico tem caráter multiplicador tendo em vista a necessidade de disseminar o conceito de “memória empresarial”. A intenção de criação de um sistema de informação voltado para a organização da memória empresarial exige que se compreenda o que vem a ser “memória empresarial”, do ponto de vista não apenas da literatura, mas também da comunidade “sebraeana”, de instituições que desenvolvem projetos semelhantes e da sociedade. O resultado de tal experiência forneceu as bases para o delineamento de um modelo de sistema de organização da memória empresarial, incluindo critérios de seleção para a formação de acervos decorrentes de tais resultados.

2 | PARÂMETROS DA MEMÓRIA EMPRESARIAL

Projetos de memória empresarial devem ser entendidos dentro da premissa de que a história de uma empresa reflete a construção de conhecimento por um conjunto de pessoas e comunidades, fundamenta os valores corporativos e evidencia a reputação de empresas e atributos de marcas.

Para tanto, os projetos de memória empresarial devem estar apoiados em um trabalho sistemático de análise e organização, de modo a abranger os aspectos mais relevantes, tangíveis e intangíveis, da trajetória que se deseja narrar, divulgar e disseminar. Os produtos de memória empresarial, por sua vez, podem variar desde os mais evidentes, como publicações, filmes e sites, até programas completos de gestão de documentos e informações, como a implantação de Centros de Documentação e Memória. É preciso entender que a memória empresarial é um conceito de trabalho, do qual se extrai a matéria-prima para várias aplicações.

Podemos construir parâmetros observando exemplos de organizações que já estruturaram projetos e programas de memória empresarial e analisando as macro etapas que comumente são seguidas por estas organizações, chegamos ao seguinte retrato:

MACRO ETAPA	DESCRIÇÃO	RESULTADOS
Organização do conhecimento, gestão de documentos e informações.	Abrange tudo aquilo que se refere à construção e aplicação do conhecimento produzido na empresa. Projetos de memória empresarial podem envolver o estabelecimento de normas de produção, trâmite, arquivamento, temporalidade, taxonomia e acessibilidade de documentos e informações da empresa.	<ul style="list-style-type: none">• Arquitetura de portais, bancos de dados e de bancos de imagens;• Organização de arquivos físicos e eletrônicos;• Implantação de Centros de Documentação (CEDOC) setoriais ou corporativos;• Implantação de Gerenciamento Eletrônico de Documentos - GED;• Organização de acervos pessoais.

<p>Pesquisa histórica, memória oral e produção de conteúdos.</p>	<p>A história deve ser referência para o público interno e externo, instrumento de gestão de negócios e canal de comunicação. A compreensão dos fatos, estratégias e processos que marcam a trajetória de um empreendimento deve ser pautada pela análise ampla de todas as fontes disponíveis, como documentos, publicações e relatos orais.</p>	<ul style="list-style-type: none"> • Livros institucionais, biografias e outras publicações; • Exposições, <i>showrooms</i> históricos e museus empresariais; • Audiovisuais, <i>sites</i> e <i>hot sites</i>; • Coletâneas de entrevistas, bancos de memória oral e <i>Storytelling</i>; • Linhas do tempo e sínteses cronológicas; • Estudos de caso e históricos temáticos.
<p>Implantação e gestão de Centros de Documentação e Memória (CDM).</p>	<p>Os CDM são os produtos mais completos de memória empresarial, pois são setores responsáveis pela gestão da memória, ou seja, pela definição e aplicação de uma política que integre o trâmite, a preservação e a disseminação de documentos físicos e eletrônicos de caráter histórico, além de dados e informações relativos à trajetória da organização, no passado e no presente.</p>	<ul style="list-style-type: none"> • Diagnóstico e planejamento estratégico para a implantação de CDM; • Gestão de todas as etapas de implantação, da captação de acervos à arquitetura de portais históricos; • Gestão de CDM já implantados.

Quadro 1: Macro etapas de projetos e programas de memória empresarial.

Fonte: Memória & Identidade Consultoria Ltda.

Aniversários de fundação da empresa, inaugurações, homenagens... são situações em que o profissional da informação se vê diante da necessidade de iniciar projetos de memória empresarial e nem sempre consegue identificar as diferentes possibilidades existentes.

É comum que a solução encontrada seja paliativa. Uma editora competente pode fazer um belo livro, aquele funcionário antigo reúne antigos documentos e fotografias para uma exposição, gravam-se algumas entrevistas com a “prata da casa”, contam-se histórias, organiza-se uma grande festa. Passado o evento, é só esperar o próximo aniversário, inauguração, homenagem... para começar toda a correria novamente. E perde-se a oportunidade de iniciar um projeto consistente de memória empresarial.

O mesmo pode acontecer quando a principal demanda é a organização de documentos. A falta de controle sobre o que já foi produzido, mesmo que não tenha sido no passado mais remoto (*onde está aquela foto do relatório do ano passado? Onde está a autorização de uso da imagem?*) ou ainda o acúmulo excessivo de documentos, são situações que podem levar o profissional da informação a tomar decisões nem sempre acertadas. Afinal, comprar uma ferramenta de base de dados resolve só em parte o problema, já que para seu bom funcionamento é fundamental antes estruturar a informação. Caso contrário, é como comprar um armário sofisticado e simplesmente

empilhar tudo lá dentro.

Para avaliar os melhores caminhos para desenvolver um projeto de memória empresarial é preciso iniciar com uma bem fundamentada análise de demandas de curto, médio e longo prazo e entender que a dinâmica de um projeto deve considerar um plano geral de trabalho, com fases, metas e avaliações parciais. Ainda que a demanda indique, por exemplo, que o produto visado é um livro e não um projeto mais amplo, o processo de concebê-lo, a pesquisa, a mobilização do público interno e todos os outros aspectos envolvidos devem resultar em um ciclo virtuoso de reflexão sobre a trajetória do empreendimento. O bom livro será, então, apenas a parte visível de um processo de autoconhecimento.

Aspectos relacionados à hierarquia de papéis sociais e relações de poder tem forte influência, tanto na construção de sentido, quanto na dominação da informação, nos dando bons parâmetros na tarefa de construção de projetos que visem a memória empresarial. Assim, temos segurança para continuar a planejar ações que visem destacar os 40 anos do Sebrae/RJ.

3 | SISTEMA DE RECUPERAÇÃO DA INFORMAÇÃO PARA A MEMÓRIA EMPRESARIAL DO SEBRAE/RJ

Os sistemas de memória organizacional/empresarial baseiam-se na persistência do conhecimento para o uso deste em situações futuras. O objetivo é gerar uma base de conhecimentos relevantes para a empresa. Existem basicamente três processos, que funcionando de forma organizada, permitem que estes sistemas cumpram com seus objetivos.

Em primeiro lugar temos a aquisição e manutenção do conhecimento, que visa coletar o conhecimento a partir das diversas fontes de informação. No segundo lugar, a integração do conhecimento visa fazer persistente o conhecimento adquirido na memória organizacional. Finalmente o processo de recuperação do conhecimento, visa fornecer conhecimento relevante aos usuários para a realização de tarefas intensivas em conhecimento. Estes processos formam os principais componentes da arquitetura básica dos sistemas de memória organizacional (STEIN, 1995 apud Sasieta, 2011).



Figura 1: Arquitetura dos Sistemas de Informação da Memória Organizacional.

Fonte: Adaptado de Stein (1995) apud Sasieta (2011).

Uma das decisões mais importantes que uma empresa adota refere-se ao tipo de Sistema de Recuperação da Informação (SRI) que será implantado para desempenhar determinada função. No caso, o escopo aqui tratado é o conhecimento gerado ao longo de 40 anos pelo Sebrae/RJ, representado por fotografias (impressas e digitais), documentos (impressos e digitais) e artefatos tridimensionais (realia).

Em março de 2016 o Sebrae/NA liberou o **Módulo Memória** do sistema de gerenciamento de acervos e coleções *Chronus Web*, Sistema de Recuperação da Informação (SRI) adotado pelo Sistema Sebrae, que por sua vez, repassa a autorização de uso aos 26 estados do Sistema, incluindo o Sebrae/RJ. Assim, recentemente, o CDI do Sebrae/RJ passou a ter uma solução tecnológica para as questões do armazenamento e recuperação dos materiais identificados como itens de memória e que poderá ser utilizado no possível CDIM.

O *Chronus Web* é um *software* multilíngue que visa à gestão de acervos bibliográficos, arquivísticos e museológicos. Totalmente desenvolvido em linguagem Java 1.6, é um sistema de plataforma *Web*, que funciona através de *browser* e sistemas de navegação pela Internet, possibilitando então a integração de sistemas que compõem uma plataforma. Por tal motivo, a instalação é necessária em apenas um servidor e o acesso ao sistema dá-se por meio da distribuição de um link único, acessível por inúmeras máquinas ao mesmo tempo e de qualquer dispositivo móvel [...], além disso, o sistema permite aos seus usuários gerenciar documentos digitais, adicionar imagens de capas de livros, resumos, textos, e etc. De tal forma que em sua base de dados é criado um acervo digital. (CHRONUS WEB..., 2016, p. 2).

Para acessar o ambiente é necessário digitar o endereço, fornecido pelo administrador do sistema, em qualquer navegador (*Internet Explorer, Mozilla Firefox, Google Chrome* ou *Apple Safari*).

Entendemos que o SRI Módulo Memória do *Chronus Web* é apenas um ferramental no processo de implementação do Projeto Memória Sebrae/RJ. É necessário que ocorra antes a **pesquisa histórica**, o **levantamento detalhado de documentos** e só em seguida, a **sistematização das informações reunidas**.

A Universidade Corporativa Sebrae (UCSebrae), vinculada e mantida pelo Sebrae/NA, começou a estruturar em 2012 por ocasião do aniversário de 40 anos do Sebrae/NA o Projeto Memorial Sebrae, que tem como objetivo identificar, organizar e selecionar conteúdos e informações nos acervos das 26 unidades estaduais, que expressem os momentos mais significativos de sua história e que, além disso, demonstrem o papel relevante do Sebrae para o fomento do empreendedorismo e o desenvolvimento sustentável dos pequenos negócios em âmbito nacional.

O trabalho de coleta de dados foi iniciado em 2012, quando se deu a inauguração do Memorial físico em Brasília, na sede do Sebrae/NA, que fez parte das comemorações de 40 anos da instituição naquele ano. Desde então, os colaboradores responsáveis pelos CDI das unidades estaduais foram capacitados na metodologia Tecnologia Social

de Memória, desenvolvida pelo Sebrae/NA com o apoio de uma consultoria externa. A partir daí, coube aos gestores dos CDI definirem a linha do tempo (cronologia) do Sebrae dos seus estados, separando e digitalizando o acervo e catalogando as obras que julgavam ser itens de memória.

Fichas de inventário foram elaboradas e padronizadas para que o material fosse editado e encaminhado para exposição no Memorial Sebrae em Brasília. Além disso, todas as unidades realizaram entrevistas de resgate de memória oral que resultaram em vídeos para reprodução e que atualmente estão disponíveis no site do Memorial, acessíveis publicamente.

Para que o CDI do Sebrae/RJ atendesse da melhor forma possível a solicitação do Sebrae/NA quanto a seleção, tratamento e digitalização dos itens levantados como materiais de memória, o Núcleo de Gestão do Conhecimento da Gerência de Conhecimento e Competitividade contratou uma empresa para digitalizar os cerca de 1.200 itens do acervo do CDI selecionados como relevantes para evidenciar a trajetória do Sebrae/RJ, dentre livros, periódicos, relatórios, projetos, fotografias e outros. Os trabalhos ocorrem durante os anos de 2014 e 2015 e envolveu toda a equipe em uma força tarefa.

A consulta, possível a partir de qualquer equipamento conectado à Internet, é plena e avançada. Pode ser realizada por qualquer termo que o usuário imagina recuperar o documento. A busca por palavras é possível devido à utilização da tecnologia de Reconhecimento Óptico de Caracteres – *Optical Character Recognition* (OCR), que proporciona aos pesquisadores maior alcance na pesquisa textual. Outra vantagem do SRI da Memória Empresarial do Sebrae/RJ é que se pode também imprimir as páginas desejadas, assim como compartilhar o *link* do documento em *sites* e *blogs*.

Com a possível criação do Centro de Documentação, Informação e Memória (CDIM) será possível, através de recurso orçamentários, aprimorar o sistema de busca a utilização do Vocabulário Controlado do Sebrae (VCS) já construído e em uso nos principais sistemas da empresa, como por exemplo o *Chronus Web*. Isso permitirá a inclusão do endereço do Sistema de Recuperação da Informação da Memória Empresarial do Sebrae/RJ na Intranet e no portal do Sebrae/RJ.

O esforço de todos os envolvidos no projeto Memorial Sebrae coordenado pela UCSebrae do Sebrae/NA deve ser destacado, em especial todas as unidades estaduais que abraçaram o projeto de uma forma muito positiva. O trabalho do atual CDI do Sebrae/RJ está direcionado para estabelecer uma política de Memória Empresarial com direcionamento para que o processo de atualização dos dados históricos da empresa seja permanente. Destacamos que todos os trabalhos estão sendo coordenados por bibliotecários, tanto no Sebrae/RJ como no Sebrae/NA e que, inúmeras dificuldades estão sendo encontradas (políticas, administrativas, financeiras e outras), mas as propostas são consistentes com estratégias, missão e visão, encontrando assim, legitimidade para prosseguir no mapeamento de ritos, cultura e identidade dos “sebraeanos”.

4 | CONSIDERAÇÕES FINAIS

Os resultados que podemos projetar dessa experiência demonstram haver várias práticas de Organização do Conhecimento em funcionamento no Sebrae/RJ. No entanto, intrinsecamente a elas, estão acoplados grandes desafios a serem superados. Alguns deles foram apontados como prioridades estratégicas para os próximos anos (até 2019) na proposta de “Política e Diretrizes de Gestão do Conhecimento para o Sistema Sebrae” elaborada em 2017 pelo Sebrae/NA, porém ainda não aprovada pela Diretoria Executiva (DIREx). A proposta refere-se à implantação de soluções de gestão do conhecimento que deverão minimizar os problemas relacionados a:

1. Oferta desconectada da demanda;
2. Alta terceirização de conhecimentos;
3. Existência de ilhas de conhecimento;
4. Projetos que não interagem ou não se integram;
5. Tecnologias e sistemas pouco integrados;
6. Necessidade de alinhamento quanto aos conhecimentos críticos;
7. Necessidade de sistematização de processos de conhecimento.

Dada a importância e relevância de se organizar o conhecimento para o Sistema Sebrae, essa ação foi considerada uma estratégia nacional com interface, rebatimento e impacto em todas as áreas de atuação do Sebrae, devendo ser operacionalizada junto ao Sebrae/RJ e a outros estados, com expectativa de contribuição para o alcance dos objetivos estratégicos do Sistema para o período do Plano Pluri Anual (PPA) 2016/2019 e demais períodos vindouros.

Na proposta da política (POLÍTICA..., 2017, p. 21) foi informado que o diagnóstico de maturidade em gestão do conhecimento aplicado no Sebrae/RJ e em outros estados para a elaboração da mesma, identificou as principais dimensões em que as práticas de Gestão do Conhecimento são desenvolvidas. Especial atenção para os itens 2, 6 e 9 da citação:

1 **Aprendizagem organizacional:** amplo uso do conceito de aprendizagem organizacional que considere tanto o aprendizado dos indivíduos, grupos e equipes dentro da organização, bem como o aprendizado do próprio Sebrae (organizações que aprendem);

2 **Gestão da informação:** processos relacionados às técnicas de documentação, Biblioteconomia, Arquivologia e Museologia;

3 **Gestão de conteúdos:** processos de criação e disseminação de produtos de conhecimento, como elaboração de publicações impressas e eletrônicas e conteúdos para portais e bases de atendimento. Incluem desde a identificação de especialistas no tema, produção dos conteúdos, adequação de linguagem, inclusão em ferramentas apropriadas, disseminação em canais adequados, entre outros;

4 **Gestão da qualidade:** Modelo de Excelência na Gestão ou demais modelos, técnicas e certificações que objetivem a melhoria da qualidade dos produtos e serviços por meio de acompanhamento e aferição de resultados e processos;

5 **Gestão de Processos:** sequenciamento de atividades da organização que oportunizem aos melhores conhecimentos relacionarem-se como principais processos de negócios, já mapeados pelo Escritório de Processos, a fim de impactar no desempenho e nos resultados da organização;

6 **Inovação:** capacidade da organização de incentivar novas ideias, romper paradigmas e gerar inovação por meio de processos de conhecimentos sistemáticos;

7 **Inteligência:** desenvolvimento de sistemáticas que facilitem o uso do conhecimento para a tomada de decisão, permitindo que a organização tenha mais flexibilidade, atualização e adaptabilidade frente aos desafios, tendências e cenários de atuação;

8 **Projetos:** capacidade da organização em disponibilizar e utilizar as melhores informações obtidas e desenvolvidas em projetos;

9 **Tecnologia:** processos de gestão de dados e informações a serem inseridas e disponibilizadas em sistemas e ferramentas de aporte tecnológico;

10 **Atendimento e relacionamento:** processos relacionados às interações com clientes (atendimento) e outros *stakeholders*, tais como parceiros, credenciados, agentes e governos (relacionamento).

As práticas de Gestão do Conhecimento são “atividades de organização, representação e recuperação da informação” segundo Lima e Alvares (2012, p. 27) e podem ser desenvolvidas em diversos formatos. As formas de apresentação das práticas são como uma vitrine no qual os conhecimentos podem ser expostos, conhecidos, acessados e consumidos. As práticas mais comuns encontradas no Sistema Sebrae e apontadas na proposta da política como atividades prioritárias até 2019 são:

- Arquivos (que hoje estão descentralizados);
- **Bibliotecas (CDI);**
- Capacitações;
- Comunidades de prática;
- Conteúdos para uso externo;
- Conteúdos para uso interno;
- Desenvolvimento de produtos;
- Documentação (termos de referência, contratos, relatórios de atividades e desempenho etc.);
- Encontros e reuniões;
- Grupos de trabalho/comitês;
- Grupos focais;
- Lições aprendidas;

- *Mailing*;
- Melhores práticas (Estudos de casos);
- **Memória organizacional**;
- Missões, *benchmarking* e visitas técnicas;
- Multimídia (vídeos e *podcasts*);
- Palestras;
- Portais, *apps* e tecnologias;
- Ponteiros de expertise de competência (páginas amarelas de colaboradores);
- Publicações impressas e digitais (livros, revistas, cartilhas, guias, manuais e infográficos);
- Rodadas de negócio e outras metodologias similares de trocas de experiências;
- *Storytelling*.

Esperamos que com a aprovação da proposta de Política e Diretrizes de Gestão do Conhecimento para todo o Sistema Sebrae, o Sebrae/RJ assuma o *ranking* das unidades do Sistema que absorveram e valorizam práticas de Organização do Conhecimento, pois as vantagens são inúmeras, conforme registrado diversas vezes no decorrer desse relato técnico.

Outra consideração é a possibilidade do tema “Memória Empresarial” ser adotado como conteúdo de capacitações (curso, consultoria, clínica tecnológica) no portfólio de soluções do Sebrae/RJ para às micro e pequenas empresas fluminenses. Pois é certo que os conceitos e benefícios da temática são aplicados em qualquer porte de empresa, mas as micro e pequenas, devido suas peculiaridades, necessitam de tratamento diferenciado na abordagem do assunto.

REFERÊNCIAS

ALMEIDA, Mauricio Barcellos. **Um modelo baseado em ontologias para representação da memória organizacional**. 2006. Tese (Doutorado em Ciência da Informação)--Universidade Federal de Minas Gerais. Disponível em: <http://www.enancib.ppgci.ufba.br/premio/UFMG_Almeida.pdf>. Acesso em: 12 set. 2018.

ALVARES, Lillian (Org.). **Organização da informação e do conhecimento**: conceitos, subsídios interdisciplinares e aplicações. São Paulo: B4, 2012.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 10520**: informação e documentação: citações em documentos: apresentação. Rio de Janeiro, 2002.

_____. **NBR 6023**: informação e documentação: referência: elaboração. Rio de Janeiro, 2002.

CENTRO DE ASSISTÊNCIA GERENCIAL DO ESTADO DO RIO DE JANEIRO. **Ata de constituição**

da sociedade civil Centro de Assistência Gerencial do Estado do Rio de Janeiro – CEAG/RJ.
Rio de Janeiro, 12 de maio de 1977. (Documento manuscrito).

CHOO, Chun Wei. **A organização do conhecimento:** como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões. Tradução de Eliana Rocha. São Paulo: SENAC, 2003.

CHRONUS Web: tutoria. Brasília, DF, 2016.

COMO transformar a memória empresarial em ferramenta de gestão. Disponível em:
<http://www.aberje.com.br/novo/curso_detalhes.asp?id=12>. Acesso em: 02 set. 2018.

CORTEZ, Maria Tereza. **Centro de documentação.** 2. ed. São Paulo: M. T. Cortez, 1987.

COSTA, Carolina da Cruz. **O resgate da memória empresarial e seus impactos na imagem da empresa:** case Centro de Memória Bunge. 2006. 162 f. Monografia. (Aperfeiçoamento/Especialização em Gestcorp)--Universidade de São Paulo, São Paulo, 2006. Disponível em: <<http://sites.google.com/site/gestcorpespecializacao/lista-de-monografias>>. Acesso em: 12 set. 2018.

DRUZIANI, Cássio Frederico Moreira; CATAPAN, Araci Hack. A percepção da memória organizacional no setor público de tecnologia da informação. **Perspectivas em Gestão & Conhecimento**, [s.l.], v. 2, n. 2, p. 97-121, 2012.

FUJITA, Mariângela Spotti Lopes. **Organização e representação do conhecimento no Brasil:** análise de aspectos conceituais e da produção científica do ENANCIB no período de 2005 a 2007.

GAGETE, Élida; TOTINI, Beth. Memória empresarial: uma análise da sua evolução. In: NASSAR, Paulo (org.). Memória de empresa: história e comunicação de mãos dadas, a construir o futuro das organizações. São Paulo: Aberje, 2004.

GOMES, Hagar Espanha. **Tendências da pesquisa em organização do conhecimento.** Pesq. bras. Ci. Inf., Brasília, v.2, n.1, p.60-88, jan./dez. 2009.

HJORLAND, Birger. Nine principles of knowledge organization. In: ALBRECHTSEN, H. **Advances Knowledge Organization**, v.4, p. 91-100, 1994.

JULIANI, Douglas Paulesky et al. Modelo para construção de bases de conhecimentos sobre projetos suportados por ferramentas colaborativas. **Revista de Gestão e Projetos**, v. 3, n. 3, p. 277, 2013.

JUAÇABA, Daniele. Centros de Memória Corporativos e a democratização da informação. **Revista Eletrônica Patrimônio: Lazer & Turismo**, Santos-SP, nov. 2005. Disponível em:
<http://www.unisantos.br/pos/revistapatrimonio/artigos.php?cod=47&bibliografia=1&#bibliografia_ancora>. Acesso em: 12 set. 2018.

LIMA, José Leonardo Oliveira; ALVARES, Lillian. Organização e representação da informação e do conhecimento. In: ALVARES, Lillian (Org.). **Organização da informação e do conhecimento:** conceitos, subsídios interdisciplinares e aplicações. São Paulo: B4, 2012.

LÓPEZ-HUERTAS, María J. Some Current Research Questions in the Field of Knowledge Organization. **Knowledge Organization**, v. 35, n. 3/2, p. 113-136. 151 references.

MACHADO, Hilka Vier. A identidade e o contexto organizacional: perspectivas de análise. **Revista de Administração Contemporânea – RAC**, Curitiba, v. 7, n. spe, 2003. Disponível em: <http://www.scielo.br/scielo.php?pid=S1415-65552003000500004&script=sci_arttext>. Acesso em: 12 set. 2018.

MEMÓRIA E IDENTIDADE CONSULTORIA. **Sinopse das atividades.** São Paulo, out. 2010.

MIRANDA, M. L. C. de. A organização do conhecimento e seus paradigmas científicos: algumas questões epistemológicas. **Informare** – Cadernos de Programa de Pós-Graduação em Ciência da Informação, Rio de Janeiro, v. 5, n. 2, p. 64-77, jul./dez. 1999.

NASSAR, Paulo. **Reputação é memória**. 12 nov. 2006. Disponível em: <<http://terramagazine.terra.com.br/interna/0,,OI1243291-EI6786,00.html>>. Acesso em: 12 set. 2018.

_____. (Org.). **Memória de empresa: história e comunicação de mãos dadas, a construir o futuro das organizações**. São Paulo: ABERJE, 2004.

NORA, Pierre. Entre memória e história: a problemática dos lugares. **Projeto História, São Paulo, n. 10, p. 07-28, dez. 1993**.

OLIVEIRA, Simone Rosa de. Memória institucional: lugar de (Re) construção de uma memória coletiva? In: SOUTO, Leonardo Fernandes (Org.). **Gestão da informação e do conhecimento: práticas e reflexões**. Rio de Janeiro: Interciência, 2014. p. 255-264.

OROSCO, Norma T; COUTINHO, Odete C. A; MONTEIRO, Vânia S. **Preservando a memória organizacional da CNEN**. Disponível em: <<http://www.cinform.ufba.br/7cinform/soac/papers/d78556f901901fe00f4197ad4c3b.pdf>>. Acesso em: 11 set. 2018.

POLÍTICA e diretrizes de gestão do conhecimento no Sistema Sebrae (Proposta). Brasília, DF: Sebrae/NA, 2017.

ROSSATTO, Maria Antonieta. **Gestão do conhecimento: a busca da humanização, transparência, socialização e valorização do intangível**. Rio de Janeiro: Interciência, 2003.

SÁ FREIRE, Patrícia de et al. Memória organizacional e seu papel na gestão do conhecimento. **Ciências da Administração**, [s.l.], v. 14, n. 33, p. 41-51, 2012.

SANTOS, J. L. S. et al. Perfil das pesquisas acadêmico-científicas sobre memória organizacional. **Espacios**, [s.l.], v. 33, n. 12, p. 12, 2012.

SASIETA, Héctor Andrés Melgar; BEPLER, Fabiano Duarte; PACHECO, Roberto Carlos dos Santos. A memória organizacional no contexto da engenharia do conhecimento. **DataGramZero**, Rio de Janeiro, v. 12, n. 3, ago. 2011.

STAIR, Ralph M. **Princípios de sistemas de informação: uma abordagem gerencial**. 2. ed. Rio de Janeiro: LTC, 1998.

SERVIÇO DE APOIO ÀS MICRO E PEQUENAS EMPRESAS NO ESTADO DO RIO DE JANEIRO. **Estatuto social: nova redação aprovada pelo Conselho Deliberativo em 13 de outubro de 1997**. Rio de Janeiro, 1998.

SOUTO, Leonardo Fernandes (Org.). **Gestão da Informação e do conhecimento: práticas e reflexões**. Rio de Janeiro: Interciência, 2014.

SOUZA, Gisele Pereira; NASSAR, Paulo. Disseminação da informação em comunicação empresarial: o caso do Centro de Memória e Referência da Aberje. **Revista CRB-8 Digital**, São Paulo, v. 3, n. 2, p. 18-28, dez. 2010. Disponível em: <<http://revista.crb8.org.br/index.php/crb8digital/article/view/51/53>>. Acesso em: 14 set. 2018.

STEIN, Eric W. Organizational memory: review of concepts and recommendations for management. **International Journal of Information Management**, [s.l.], v. 15, n. 2, p. 17-32, 1995.

TEIXEIRA FILHO, Jayme. **Gerenciando conhecimento**: como a empresa pode usar a memória organizacional e a inteligência competitiva no desenvolvimento dos negócios. Rio de Janeiro: SENAC, 2000.

THIESEN, Icléia. **Memória institucional**: a construção conceitual numa abordagem teórico-metodológica. 1997. 169 f. Tese (Doutorado)--Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1997. Disponível em: <http://tede-dep.ibict.br/tde_arquivos/1/TDE-2008-02-15T15:39:41Z-24/Publico/icleiacosta1997.pdf>. Acesso em: 17 set. 2018.

VASCONCELOS, José Braga de; ROCHA, Álvaro; KIMBLE, Chris. **Sistemas de informação de memória organizacional**: uma abordagem ontológica para a definição de competências de grupo. Disponível em: <http://www.ime.uerj.br/~neide/Artigos/memoria_organizacional.pdf>. Acesso em: 11 set. 2018.

GERÊNCIA DE PROJETOS EM COMPUTAÇÃO VESTÍVEL: DIRETRIZES PARA O DESENVOLVIMENTO DE PRODUTOS VESTÍVEIS INTELIGENTES

Renan Gomes Barreto

Universidade Federal da Paraíba
Programa de Pós-Graduação em Informática
João Pessoa – Paraíba

Lucas Oliveira Costa Aversari

Universidade Federal da Paraíba
Programa de Pós-Graduação em Informática
João Pessoa – Paraíba

Renata Gomes Barreto

Universidade Federal da Paraíba
João Pessoa - Paraíba

Gabriela Ferreira Marinho Barreto

Faculdade Maurício de Nassau
João Pessoa – Paraíba

RESUMO: O ser humano em sua essência tem a necessidade de se adaptar ao meio através de ferramentas que possam facilitar sua vida cotidiana. Um estudo recente mostra que o mercado de dispositivos vestíveis cresce cerca de 10% ao ano e espera-se que o gasto anual com esses tipos de dispositivos seja por volta de 40 bilhões de dólares em 2022. Esta pesquisa visa propor diretrizes para o desenvolvimento de produtos vestíveis inteligentes a fim de auxiliar gerentes de projetos no processo de tomada de decisão. Durante o desenvolvimento desta pesquisa, foram selecionados produtos vestíveis inteligentes disponíveis no mercado durante os anos de

2016 a 2018 e, assim, foi feita uma avaliação heurística utilizando o método de inspeção e o framework DECIDE. Adicionalmente, personas primárias e secundárias foram definidas para cada dispositivo e, por fim, para cada categoria de dispositivo vestível, foram propostas diretrizes que poderão ser seguidas durante o desenvolvimento de um novo produto. O desenvolvimento desta pesquisa pode auxiliar gerentes e desenvolvedores na tomada de decisão durante a prototipação e desenvolvimento de um novo dispositivo vestível inteligente.

PALAVRAS-CHAVE: Computação Vestível, Interação Humano-Computador, Gerência de Projetos.

PROJECT MANAGEMENT IN WEARABLE COMPUTERS: GUIDELINES FOR INTELLIGENT PRODUCT DEVELOPMENT

ABSTRACT: The human being, in essence, has the need to adapt to the environment through tools that can facilitate his daily life. A recent study shows that the wearable devices market is growing at about 10% per annum, and annual spending on these devices is expected to be around \$ 40 billion by 2022. This research proposes guidelines for the development of intelligent wearable products to assist project managers in the decision-making process.

During the development of this research, we selected smart wearable products available in the market during the years 2016 to 2018 and, therefore, a heuristic evaluation was made using the inspection method and the DECIDE framework. Additionally, primary and secondary personas were defined for each device and, finally, for each category of wearable device, guidelines were proposed that could be followed during the development of a new product. The development of this research can help project managers and developers in decision making during the prototyping and development of a new smart wearable device.

KEYWORDS: Wearable Computer, Human-Computer Interaction, Project Management.

1 | INTRODUÇÃO

O ser humano em sua essência tem a necessidade de se adaptar ao meio através de ferramentas que possam facilitar sua vida cotidiana. O primeiro histórico datado de dispositivo vestível ocorreu por volta de 1700, na China, quando um ábaco foi miniaturizado em um anel e dado de presente à Qing durante sua dinastia (STEPHANIDIS, 2012). Assim como qualquer utensílio, os chamados wearables ou, dispositivos vestíveis, em geral, surgiram com essa proposta, baseados em duas vertentes: Mobilidade e Aumento da Percepção Sensorial ou Realidade Aumentada. Os wearables com foco em mobilidade baseiam-se na necessidade de estarem disponíveis onde o usuário esteja, por outro lado, os aparelhos baseados em realidade aumentada devem funcionar com o propósito de melhorar as atividades cotidianas sem interferir negativamente nela (CORSO, 2013).

Com o avanço do hardware, em 2004 a Microsoft retomou o conceito de smartwatch com o lançamento do Microsoft Spot, que integrava algumas funcionalidades como previsão do tempo local, cronômetro, alterações de layout do mostrador do relógio (MENTOR, 2013). Paralelamente a Apple, em parceria com a Nike, lançou um novo ramo de wearables, os trackers (MÜLLER, 2014). Em 2013 a Samsung retomou o conceito abandonado pela Microsoft e lançou o Galaxy Gear S3, um relógio que funcionava como segunda tela ao ser pareado com alguns aparelhos da linha Galaxy. O dispositivo era capaz de atender chamadas, responder mensagens, gravar áudios, vídeos curtos e tirar fotos (SAMSUNG, 2018). No mesmo ano a Google lança o Google Glass, uns óculos sem fio para o uso de realidade aumentada (LV et al., 2014)

Futuramente, com o avanço tecnológico, os wearables estarão cada vez mais presentes no nosso cotidiano. A necessidade de ter um dispositivo sempre ao alcance se tornará cada vez mais presente já que este serviria como uma extensão do corpo e, para acessá-lo, não seria preciso nada mais do que um simples gesto (JUNIPER, 2018).

Como objetivo geral, esta pesquisa visa propor diretrizes para o desenvolvimento de produtos vestíveis inteligentes a fim de auxiliar gerentes de projetos no processo de

tomada de decisão. Os objetivos específicos desta pesquisa podem ser identificados abaixo, em sequência, como:

- Avaliar os produtos existentes no mercado a fim de identificar suas principais características e padrões de design seguidos;
- Propor um conjunto de diretrizes que poderão ser utilizadas no desenvolvimento de novos produtos inteligentes vestíveis.

Este trabalho está dividido em seções com os seguintes tópicos: Introdução, Fundamentação Teórica, Diretrizes para o desenvolvimento de produtos vestíveis inteligentes e Considerações Finais. Na Seção 1 é apresentada a contextualização do problema, motivação, objetivos geral e específico. A Seção 2 apresenta os conceitos relevantes para o entendimento das diretrizes propostas. A Seção 3 apresenta a metodologia e alista de diretrizes propostas por esta pesquisa. Para finalizar, a Seção 4 resume a pesquisa realizada e apresenta suas contribuições.

2 | FUNDAMENTAÇÃO TEÓRICA

O termo computação vestível, do inglês *wearable computing*, foi definido pela primeira vez em 1998 por Steve Mann em seu artigo *Wearable computing as means for personal empowerment* (1998), onde se afirma que um *wearable* é um computador que está englobado no espaço pessoal do usuário e controlado por ele onde tem tanto constância operacional quanto interacional, ou seja, está sempre ligado e acessível. Mais notavelmente é um dispositivo que está sempre com o usuário, para o qual o utilizador pode sempre introduzir comandos e executar um conjunto de instruções dadas durante uma caminhada enquanto executa outras atividades.

O aspecto mais notável da computação, em geral, seja vestível ou não, é a sua reconfiguração e generalidade, por exemplo, sua função pode variar amplamente, dependendo das instruções fornecidas para a execução do programa. O computador vestível é mais do que apenas um relógio de pulso ou óculos normais: ele tem a funcionalidade completa de um sistema computacional, mas além de ser um computador com todos os recursos, ele também está intimamente entrelaçado com o usuário. Isto é o que define o computador vestível além de outros dispositivos portáteis, como relógios de pulso, óculos de grau, rádios portáteis, etc. Ao contrário dos demais dispositivos portáteis que não são programáveis (reconfiguráveis), o computador vestível é tão reconfigurável quanto um desktop familiar ou um computador mainframe (MANN, 1998). A Computação vestível é formalmente definida em seis atributos fundamentais:

- **Não restritivo:** O uso do equipamento não deverá restringir locomoção e visão do usuário.
- **Não monopoliza a atenção:** Deixa livre o usuário para se concentrar em

outras atividades.

- **Observável pelo usuário:** O dispositivo deve funcionar mesmo sem a atenção do usuário mas permite que o usuário o observe a qualquer momento.
- **Controlável:** Pode ser controlado pelo o usuário a qualquer momento
- **Atento ao Ambiente:** Permite ao usuário obter mais informações sobre o ambiente que o cerca.
- **Meio de Comunicação:** Aumenta a capacidade de comunicação do usuário.

Através dessas definições, também podem ser listadas outras características adicionais específicas dessa tecnologia:

- **Constante:** O aparelho não deve precisar ser ligado ou desligado. Ele deve estar sempre disponível para o uso, sempre ligado.
- **Pessoal:** Interconectado com seu usuário.
- **Privado:** Os dados devem estar disponíveis apenas para o usuário principal.
- **Protético:** Deve se comportar como como uma extensão do corpo do usuário.

A computação vestível possui um estado da arte em constante renovação. No que tange ao mercado dessa área de inovação, novas ideias e originalidades são emergentes em um fluxo contínuo e ininterrupto. Há novos produtos sempre surpreendentes, mas a maioria deles podem durar apenas um curto período de tempo, e logo ser substituído por um estado de arte superior. Neste sentido, pode-se citar dentro do mercado de computação vestível, produtos como os óculos inteligentes, relógios inteligentes e pulseiras inteligentes (CORSO, 2013).

Com o incessante progresso da ciência e da tecnologia, quando as pessoas estão familiarizadas aos conceitos de wearable, o mercado demanda de dispositivos inteligentes e móveis cada vez mais inovadores. É imaginável, com o avanço da tecnologia, como este aspecto da tecnologia tem se desenvolvido e popularizado, e tal influência positiva demonstra como a computação vestível oferece enorme oportunidades para o cotidiano (BILLINGHURST; STARNER, 1999).

O avanço tecnológico e o desenvolvimento de sensores sem fio com um baixo nível no consumo de baterias, corrobora para uma vasta produção de dispositivos aplicáveis em áreas diversas, como por exemplo, contribuindo nos serviços de saúde que buscam dispositivos vestíveis menos invasivos possibilitando um acompanhamento dos dados do paciente. Essa característica dos “wearables” auxilia principalmente os médicos, cuidadores e familiares no monitoramento dos idosos, que são o público-alvo de dispositivos vestíveis voltados para a saúde, permitindo através de sensores informações contendo os sinais vitais, fisiológicos, atividades diárias, dados médicos,

comunicação de emergência, ou seja, focados em apresentar um diagnóstico precoce do paciente, prevenindo possíveis imprevistos emergenciais (COPETTI, 2015).

Uma pesquisa recente realizada pela *Juniper Research* mostra que o mercado de dispositivos vestíveis cresce cerca de 10% ao ano e espera-se que o gasto anual com estes tipos de dispositivos seja por volta de 40 bilhões de dólares em 2022. Além disso, na área da saúde, a pesquisa prevê que 5 milhões de indivíduos utilizarão dispositivos vestíveis para o monitoramento de dados fisiológicos como batimento cardíaco e pressão arterial (JUNIPER, 2018).

No que diz respeito a privacidade, nos próximos anos, pesquisas deverão ser focadas na total transparência e padronização sobre quais informações estão sendo armazenadas por esses dispositivos, e como os usuários poderão manter o controle sobre sua segurança (BILLINGHURST; STARNER, 1999).

3 | DIRETRIZES PARA O DESENVOLVIMENTO DE PRODUTOS VESTÍVEIS INTELIGENTES

Durante o desenvolvimento desta pesquisa, foram selecionados produtos vestíveis inteligentes disponíveis no mercado durante os anos de 2016 a 2018, sendo eles: Apple Watch, Google Cardboard, Google Glass, Oculus Rift e Xiaomi MiBand.

A partir dos produtos foi feita uma avaliação heurística seguindo a metodologia proposta por BARBOSA e SILVA (2010), utilizando o método de inspeção e framework DECIDE, com o objetivo de coletar dados sobre a experiência de uso, detalhando como os usuários realizam suas atividades com ou sem o apoio de sistemas interativos.

Uma das etapas da avaliação é a definição das personas que utilizará o dispositivo. Uma persona é um personagem fictício, normalmente definidos por seus objetivos que representa um arquétipo hipotético de um grupo de usuários reais. Para cada dispositivo foi definido uma persona primária e uma secundária (BARBOSA; SILVA, 2010). A tabela 1 mostra uma persona primária criada para a avaliação do dispositivo Xiaomi MiBand.

Dispositivo	Xiaomi MiBand	Tipo	Primária / Usuário
Perfil: Amanda tem 20 anos e é estudante de Mídias Digitais, fala, lê e compreende Inglês razoavelmente bem. Gosta de música, cultura geek e como lazer costuma correr ou fazer trilhas nos finais de semana. É estagiária no departamento de comunicação de uma startup voltada para tecnologia, sua função é editar vídeos e criar conteúdo para o site da empresa e por conta disso, possui fácil acesso ao que há de novo no mercado tecnológico.			

Interesses:

Amanda tem interesse na área de tecnologia e pretende fazer mestrado nela. Pensa em fazer intercâmbio para o Reino Unido para agregar conhecimentos e aperfeiçoar suas habilidades linguísticas. Aspira se tornar funcionária efetiva como Gerente de Projetos na mesma startup que estagia.

Conhecimento em Informática:

Possui amplo conhecimento em informática. É frequentadora assídua de sites de tecnologia, fóruns, ferramentas de cooperação online e ferramentas, edição de vídeo e de mídias sociais.

Citações:

- “Faço uso da tecnologia no meu dia-dia para me manter informada instantaneamente sobre o que acontece nas minhas redes sociais”
- “Não conseguiria viver sem tecnologia pois esqueço meus compromissos com facilidade”
- “Gostaria que tecnologia fosse usada auxiliando seus usuários no acompanhamento de suas atividades físicas”

TABELA 1: Persona primária descrita durante a avaliação do dispositivo Xiaomi MiBand.

Cada dispositivo foi analisado segundo suas personas e cenários e cada heurística violada foi detalhada e classificada quanto sua severidade. A tabela 2 mostra a saída da avaliação heurística para o dispositivo Xiaomi MiBand.

Heurística Violada	Categoria	Severidade
Prover Feedback	Notificações de Status	Baixa
Descrição: O aplicativo oficial provido pela fabricante Xiaomi limita o número aplicativos que serão notificados pela pulseira e também não oferece um controle mais abrangente. Não é possível configurar a intensidade de vibração ou a cor dos LEDs de acordo com cada aplicativo, por exemplo. Dessa forma, ao ser notificado, não se sabe ao certo qual aplicativo está requisitando atenção do usuário, problema que poderia ser facilmente resolvido através de uma atualização, visto que o hardware da pulseira suporta essa funcionalidade.		
Heurística Violada	Categoria	Severidade
Liberdade do Usuário	Notificações de aplicativos	Média
Descrição: O dispositivo não dá feedback algum ao usuário de quando está em funcionamento ou com bateria fraca. Não é possível saber se o dispositivo está funcionando sem ter que abrir o aplicativo e as informações relacionadas à bateria fraca aparecem em forma de notificação somente quando a pulseira está conectada ao celular.		

TABELA 2: Heurísticas violadas pelo dispositivo vestível Xiaomi MiBand.

Por fim, de acordo com o documento resultante da etapa de avaliação heurística e a partir dos dados coletados dos produtos analisados, foi desenvolvido um conjunto de

diretrizes que poderá ser utilizado na criação de novos projetos de produtos vestíveis.

3.1 Diretrizes para SmartWatches

- a) O DESIGN DEVE SER COMPATÍVEL COM ALGO JÁ EXISTENTE: O público alvo dos SmartWatches são pessoas que normalmente já estão familiarizadas com smartphones e por isso é indicado a utilização de padrões de interfaces similares, facilitando o aprendizado e a memorização da mesma.
- b) O MATERIAL DEVE SER HIPOALERGÊNICO: O produto ficará em contato direto com a pele do usuário por longos períodos e não deve causar reações alérgicas.
- c) DEVE SER RESISTENTE A ÁGUA E POEIRA: Relógios entram em contato frequentemente com líquidos. Uma certificação IP67 ou similar é recomendável.
- d) O PESO DO CONJUNTO NÃO PODE SER INCÔMODO: Ao utilizar o produto, o usuário deve senti-lo como parte do seu corpo e até esquecer que o está usando. Um produto pesado pode ser um incômodo.
- e) MANTENHA A INTERFACE SIMPLES: Devido ao tamanho da tela, as informações apresentadas devem ser simples e diretas. Tente colocar as informações mais importantes no centro da tela. O uso efetivo de ícones pode reduzir significativamente a quantidade de texto.
- f) O CONTEÚDO DA TELA DEVE SER VISÍVEL SOB LUZ DO SOL: O dispositivo deve ser projetado pensando em seu uso ao ar livre. A utilização de fontes claras em fundos escuros melhora o contraste fazendo o conteúdo ser visível mesmo em contato direto com a luz do sol.
- g) UTILIZE AS BORDAS DO DISPOSITIVO COM SABEDORIA: As bordas do dispositivo podem ser utilizadas a seu favor aproveitando melhor o espaço.
- h) ESTRUTURA DE NAVEGAÇÃO CONSISTENTE E FÁCIL DE SER SEGUIDA: A utilização de uma estrutura de navegação consistente ajuda ao usuário entender melhor o funcionamento da aplicação. Tente diminuir os passos necessários para acessar cada funcionalidade do aplicativo. Identificar cada aplicativo por cor, tipo ou layout pode ajudar na navegação entre tarefas distintas.
- i) UTILIZE TRANSIÇÕES FLUIDAS: Transições fluidas entre tarefas ajudam aos usuários entenderem melhor a informação atual e seu contexto.
- j) INTERFACE RÁPIDA E RESPONSIVA: Tente diminuir o número de passos para realização de cada tarefa. Uma interface rápida e responsiva dará ao usuário a impressão que ele está no controle. Respostas imediatas dão ao usuário a certeza que suas ações foram processadas e aumentam a confiança que o mesmo tem no produto.

3.2 Diretrizes para SmartBands

- a) O DESIGN DEVE SER COMPATÍVEL COM ALGO JÁ EXISTENTE: Para manter a essência de uma pulseira, sua largura deve ser entre 12 mm e 22 mm, o comprimento deverá ser razoável, para ajustar-se a diferentes diâmetros de pulso.
- b) O MATERIAL DEVE SER HIPOALERGÊNICO E RESISTENTE: SmartBands dificilmente sairão do braço de seu usuário, seja durante a prática de alguma atividade física ou dormindo, devido a essa característica, elas devem ser resistentes a água e intempéries, é recomendável uma certificação IP67 ou similar.
- c) O PESO DO CONJUNTO NÃO PODE SER INCÔMODO: Ao utilizar o produto, o usuário com o passar do tempo deve senti-lo como parte do seu corpo e até esquecer que o está usando. Um produto pesado pode ser um incômodo e gerar futuras lesões.
- d) DISPOSIÇÃO DE BOTÕES PARA INTERAÇÃO: Se existir, ideal que tenha até dois botões bem localizados que independente do braço que o usuário preferir colocar, não terá dificuldade em acessá-los.
- e) INFORMAÇÕES VISUAIS NO DISPOSITIVO: Necessariamente uma SmartBand não exige um display, que facilmente pode ser substituído por painéis de LED, pois as informações apresentadas ao usuário no dispositivo são simples e devem ser intuitivas, mas se for o caso, deverá seguir um padrão médio de 1,4 polegadas.
- f) COMPATIBILIDADE E CONECTIVIDADE COM O SMARTPHONE: Uma conexão eficiente é primordial para que os dados do dispositivo não sejam perdidos ao saturar seu armazenamento interno, bluetooth ou NFC são recomendáveis. O sistema operacional onde irá funcionar a aplicação do gadget deverá ser de escolha do usuário, ou seja, o aplicativo deverá funcionar em sistemas como Windows phone, Android e IOS.
- g) EFICIÊNCIA ENERGÉTICA: A bateria deve manter o dispositivo em funcionamento a longo prazo visto que estes dispositivos são focados em monitoramento. O recomendável é que eles tenham carga para ao menos uma semana de uso sem interrupções.
- h) SISTEMA DE SENSORIAMENTO QUE FORNEÇA DADOS COM BAIXA MARGEM DE ERRO: Para um dispositivo voltado ao seguimento fitness espera-se que forneça informações claras e eficientes, isso irá ocorrer através de vários sensores bem localizados no corpo da SmartBands, como acelerômetros, sensor de batimentos cardíacos ou motor de vibrações cilíndricas.
- i) APLICAÇÕES COM INFORMAÇÕES CLARAS E OBJETIVAS: Ao acessar o aplicativo as transições devem ser de forma fluida, não exigindo grande

memorização para efetuar tarefas, gráficos com pouca complexidade na hora de passar as informações como as horas de sono dormidas ou o número de passos computados durante o dia.

j) **SISTEMAS DE NOTIFICAÇÃO:** Existem diversas formas para chamar a atenção do usuário sem incomodar ou prejudicar de forma invasiva outras atividades que ele pode estar realizando. Sistemas com LEDs ou através de vibrações já são bem recebidos e de fácil compreensão entre usuários de smartphones, ou seja, não exigirá um grande esforço para uma fácil associação na memória de longo prazo. Além disso, o usuário deve ter liberdade para configurar e gerenciar que aplicações deseja ser notificado.

3.3 Diretrizes para SmartGlasses

a) **O DESIGN DEVE SER COMPATÍVEL COM ALGO JÁ EXISTENTE:** Os SmartGlasses devem se ajustar aos diferentes tamanhos e formatos de rosto, além do foco da tela.

b) **DEVE SER RESISTENTE A ÁGUA E POEIRA:** Por ser um produto que ficará exposto boa parte do tempo, é de extrema importância que ele tenha resistência a água e poeira. De preferência com certificação IP67.

c) **O PESO DO CONJUNTO NÃO PODE SER INCÔMODO:** Ao utilizar o produto, o usuário deve senti-lo como parte do seu corpo e até esquecer que o está usando. Para um SmartGlass é recomendável que ele esteja entre 30g e 80g.

d) **NÃO SEPARE O PRODUTO DE SEU CONTROLADOR:** Os controladores do produto devem estar em uma única peça, tal como o Google Glass, que possui seus controladores na haste direita do dispositivo.

e) **CONEXÃO INDEPENDENTE DO SMARTPHONE E EFICIÊNCIA ENERGÉTICA:** Ao contrário dos SmartWatches, os SmartGlasses devem ser independentes dos smartphones podendo até mesmo substituí-los.

f) **ÁUDIO POR INDUÇÃO ÓSSEA:** Seguindo o conceito de vestíveis de realidade aumentada, a transmissão de áudio do dispositivo deve ser feita por indução óssea e caso não seja, faça uso de apenas um canal auditivo do usuário, procurando assim não o desligar da realidade

g) **INTERFACE NÃO POLUÍDA:** A interface não deve interferir nas atividades cotidianas, por isso, é preciso que ela seja mais limpa possível e não ocupe excessivamente o campo de visão do usuário.

h) **ESTRUTURA DE NAVEGAÇÃO CONSISTENTE E FÁCIL DE SER SEGUIDA:** A utilização de uma estrutura de navegação deve ajudar o usuário a entender melhor o funcionamento da aplicação. Tente diminuir os passos necessários para acessar cada funcionalidade do dispositivo e tornar seus menus padronizados dentro de cada aplicação.

i) **FUNCIONALIDADES SOMENTE SE NECESSÁRIAS:** Nada é mais frustrante que um dispositivo lento e sem resposta. Por isso é mais do que necessário que as funcionalidades sejam implementadas somente se houver necessidade de uso.

j) **NOTIFICAÇÕES NÃO INVASIVAS E CONSISTENTES:** As notificações não podem comprometer a visão do usuário. Devem estar dispostas de maneira discreta e durante intervalo que não ultrapasse 3 segundos. As notificações devem seguir o mesmo padrão e o usuário deve possuir a liberdade de escolher se quer expandi-la ou ignorá-la.

3.4 Diretrizes para SmartVR

a) **CONTROLADORES DE MOVIMENTO:** Os gadgets de realidade virtual em sua maioria são voltados para jogos, tendo assim a necessidade de controladores e para melhor conforto do usuário, é aconselhável que sejam sem fio.

b) **CONFORTO, QUALIDADE DE SOM E IMAGEM:** É de suma importância que o gadget seja confortável evitando que o usuário sinta náuseas. Para uma melhor experiência imersiva é essencial imagem e som em alta definição.

c) **TOTAL COMPATIBILIDADE:** Aplicações não devem necessitar do uso de acessórios externos como mouse para efetuar ações simples.

4 | CONSIDERAÇÕES FINAIS

O design, seja em hardware ou software sempre será uma barreira no ramo dos dispositivos vestíveis. Antes que o produto chegue no estado final é importante que a equipe esteja atenta a como o produto se comportará com seu usuário e no que está em sua volta. Os dispositivos, caso sejam voltados para o uso cotidiano, devem se camuflar ou substituir de maneira imperceptível o espaço de dispositivos já existentes ou similares.

O desenvolvimento desta pesquisa pode auxiliar gerentes e desenvolvedores na tomada de decisão durante a prototipação e desenvolvimento de um novo dispositivo vestível inteligente visto que ela propôs uma lista de diretrizes, categorizadas de acordo com o tipo de dispositivo, que levam em conta a usabilidade e possíveis interesses de futuros usuários do produto final.

REFERÊNCIAS

BARBOSA, Simone; SILVA, Bruno. **Interação humano-computador**. Elsevier Brasil, 2010.

BILLINGHURST, Mark; STARNER, Thad. **Wearable devices: new ways to manage information**. Computer, v. 32, n. 1, p. 57-64, 1999.

COPETTI, Luciele. Sistemas de aplicações ubíquos na comunicação da saúde: considerações sobre

práticas e procedimentos para o cuidado de si. In: **Anais do XXXVIII Congresso Brasileiro de Ciências da Comunicação**. 2015. p. 03-07.

CORSO, Aline. **Uma Breve Introdução aos Computadores Vestíveis: Corpo, Tecnologia e Ficção Científica**. 7º Simpósio Nacional da Associação Brasileira de Cibercultura, 2013.

JUNIPER, Research. **Digital Health: Disruptor Analysis, Country Readiness & Technology Forecasts 2018-2023**. 2018.

LV, Zhihan et al. Hand-free motion interaction on google glass. In: **SIGGRAPH Asia 2014 Mobile Graphics and Interactive Applications**. ACM, 2014. p. 21.

MANN, Steve. Wearable computing as means for personal empowerment. In: **Proc. 3rd Int. Conf. on Wearable Computing (ICWC)**. 1998. p. 51-59.

MENTOR, Jon. **2004 Microsoft SPOT Watch Smartwatch Review**, 2013. Disponível em: <<https://wear.guide/smartwatch-reviews/2004-microsoft-spot-watch-smartwatch/>>. Acesso em: 30 de abril de 2019.

MÜLLER, Leonardo. **Tudo sobre o Apple Watch, o tão esperado relógio inteligente da Maçã**, 2014. Disponível em: <<https://www.tecmundo.com.br/apple-watch/62795-tudo-apple-watch-o-tao-esperado-relogio-inteligente-maca.htm>>. Acesso em: 27 de abril de 2019.

SAMSUNG. **Samsung Gear S3 Classic Silver - SM-R770NZAZTO**, 2018. Disponível em: <<https://www.samsung.com/br/wearables/gear-s3-classic-r770/>>. Acesso em: 30 de abril de 2019.

STEPHANIDIS, Constantine. **The encyclopedia of human-computer interaction. The encyclopedia of human-computer interaction**, 2012.

EXPLORING *BIG DATA* CONTENT AND INFORMATION METRICS: INTERSECTIONS AND ANALYSIS TO SUPPORT DECISION-MAKING

Rafael Barcellos Gomes

(ORCID: <https://orcid.org/0000-0003-1179-2923>)

Universidade Federal do Rio de Janeiro – UFRJ
Rio de Janeiro – RJ

Vânia Lisboa da Silveira Guedes

Universidade Federal do Rio de Janeiro – UFRJ
Rio de Janeiro – RJ

ABSTRACT: It proposes to expose theoretical and practical aspects in metric studies of scientific information associating them with the exploration of *Big Data* to delineate tendencies and scenarios on Public Health in Brazil. All the while dissecting the contemporary concept of information and its strategic importance to support decision-making in organizations inserted in the context of the Information and Knowledge Society, examining the importance of scientometric studies and network produced data analysis for the development of valuable information discovery processes. To do this, it investigates possible relationships between data from *Big Data*, official data published by the Brazilian Ministry of Health, and technical data extracted from scientific publications indexed in research databases. Patterns of convergence among the observations were recorded, noting that some occurrences of endemic outbreak peaks are preceded by peaks of informational

demands through terms related to infectious agents involved in endemics. Finally, it highlights the relevance of analytical methodology - based on data visualization - for the development of studies that support decision-making processes and organizational strategies, outlining scenarios and trends to support Public Administration and private management.

KEYWORDS: *Big data*, data mining, scientometric analysis, strategic management, decision-making.

EXPLORANDO DADOS DE *BIG DATA* E MÉTRICAS DE INFORMAÇÃO: INTERSEÇÕES E ANÁLISES PARA SUPORTE A TOMADAS DE DECISÃO

RESUMO: Propõe expor aspectos teóricos e práticos em estudos métricos da informação científica, associando-os à exploração de dados de *Big Data* para deslindamento de tendências e cenários em Saúde Pública, no Brasil. Neste ínterim, diseca o conceito contemporâneo de informação e a sua importância estratégica para apoio a tomadas de decisão em organizações inseridas no contexto da Sociedade da Informação e do Conhecimento, examinando a importância dos estudos cienciométricos e da análise de dados produzidos em rede no desenvolvimento de processos de descoberta de informações de valor. Para tal, investiga

possíveis relações entre dados de *Big Data*, dados oficiais publicados pelo Ministério da Saúde do Brasil e órgãos correlatos e, dados técnicos extraídos de publicações científicas indexadas em bases de pesquisa. Registrou-se padrões de convergência entre as observações colhidas, constatando que alguns picos de ocorrência de surtos endêmicos são precedidos por picos de demandas informacionais por termos relacionados aos agentes infecciosos envolvidos nas endemias. Por fim, destaca a relevância da metodologia analítica – baseada na visualização de dados – para o desenvolvimento de estudos que apoiem processos decisórios e estratégias organizacionais em políticas de Saúde, delineando cenários e tendências para suporte à Administração pública e à gestão privada.

PALAVRAS-CHAVE: *Big data*, mineração de dados, análise cienciométrica, gestão estratégica, processo decisório.

1 | INTRODUCTION

Over the last four decades public and private administrations have been strongly impacted by new dynamics in the political, social, economic and technological fields, as a result of the accelerated changes that have taken place since the emergence of the so-called Information and Knowledge Society - in the early 1980s Brazil.

This moment was characterized above all by the migration of a characteristically industrial economic paradigm - based on manufacturing and industrial assets into a post-industrial one - characterized by an essentially tertiary economy, in which social, economic, and power relations settle based on information mediation, made possible by the development of new information and communication technologies.

The problem here approached is derived from the understanding that - in this new and challenging conjuncture - the survival and success of contemporary organizations depend on their adaptive capacity in face of the new configurations of demand and market, as well as the ability to distinguish the spatiotemporal context in which they are acting, transposing the challenges and taking advantage of opportunities (NEELY; GREGORY; PLATTS, 1995). These discernment and perception by the organization are born from the generation of knowledge, product of this element that moves the globalized world: information.

2 | CONCEPTUALIZATION AND CONTEXTUALIZATION

In the context of post-industrial society, information becomes more important in the process of capital generation, a fact by which it comes to be compared to an “[...] economic asset (commodity) [...]” and to a “[...] product/merchandise [...]” (TARAPANOFF, 2001, p.91), which are essential to the creation of knowledge and intelligence, whether it is for leisure, for work or for teaching, research and development. From this unusual socio-economic context there emerged a new understanding of effective business management, whose decision-making processes are based on the search for and use

of information resources. This was called Information Management (IM).

From a philosophical-epistemological point of view, Rascão (2006, p.33) understands “information” under four fundamental perspectives: information as a thing, as a process, as social construction, and information as “potentiality” (in the sense that a piece of data exists as a potential source of information and not as a piece of information itself).

Rascão then proposes to interweave the fundamental concepts of information to the concept of Information Management, noting that in the context of Information Society this management seeks “[...] to provide the right information (as a thing) to the right person at the right moment and in the right place [...]”, but also to preserve the information generated from the individual cognitive process, as well as information generated with “[...] the language, thought and dialogue [...]” (RASCÃO, 2006, p.51-52) of social groups.

From this understanding, it is possible to comprehend that the whole process of signification, valorization and the use of analyzed information will occur under cultural influence, which will also interfere in the profile of the strategies constructed for decision-making - both individual and organizational.

It is observed, therefore, that in Information Management in the context of the New Economy it is possible to recognize information under various aspects (KIRK, 1999, p.12). In this perspective, the concept of potential information - discussed by Rascão (2006), Shannon and Weaver (1948), among others - emerges as a particularly interesting aspect, especially with regard to the practice of Information Management in the last three decades.

Valentim (2010) calls this period New Economy, an expression that reinforces the perception of the concept of information as “commodity” (TARAPANOFF, 2001, p.91). In this model, social and economic links are each time less structured on tangible products (SCHWAB, 2016) and, increasingly, in the provision of services: online, on-time, full-time.

3 | DATA AND INFORMATION

Becker (2015) analyzes a series of statisticians and information theorists and evaluates that “data” correspond to the starting point in which to base an observation, allocating them at a primary level; the information in turn corresponds to the product of contextualized and interpreted data and is therefore classified at a secondary level - thus understanding that data when decontextualized have no meaning at all. Turning to the concepts proposed by Becker, and considering Wiener’s definitions, it is possible to trace important relations with the proposal of Rascão.

Wiener (1948) believed that data processed from an old piece of information could give rise to new information, which in turn, would have new values. This understanding that processed data can (have the potential to) give rise to precious information, is in

line with the idea initially exposed by Rascão when defining “information as potentiality” (2006, p.33), notable in the scope of the New Economy.

In the meantime, Ginman (1987) contributes to this observing that often a precious data is not valued for not being correctly and widely identified (visualized); this has prevented users and organizations from fully exploiting accessible and potentially valuable informational content. Some of these contents are generated from initiatives of digital users who, proceeding in search of information that meets their demand, leave informational “traces” often not perceived. This individual’s way of being, of acting or reacting - in the search for information - receives the name of informational behavior (SILVA; MOREIRA; SILVA, 2014).

Fialho and Andrade (2007) understand that informational behavior is a complex process, affected by external factors - environmental, demographic, economic, social, etc. It is observed, therefore, that data generated from actions of information users bring - implicitly - I get informational contents that reflect the context in which they were created. It is exactly this “potential information” load (RASCÃO, 2006, p.33) that gives the diluted data the probability/potential to generate new information.

4 | EXPLORING *BIG DATA* CONTENT

From the foregoing exposition, it is possible to understand, with greater clarity, the Market’s interest in unprecedented and timely information, treating it as a socio-economic element of powerful strategic value. In this context, many of the contents used have been obtained from the analysis of information in its grossest state: that of “data”.

However, the largest source of raw data for strategic information generation is non-standardized; a mass of virtual data currently estimated at more than 3 Zettabytes (3 billion Terabytes), an exponentially increasing volume - something in the order of 2.2 million Terabytes per day, worldwide (GARTNER, 2013). This immense informational content, available for exploration and analysis in digital networks, is called *Big Data*.

Big Data is the largest source of raw material for the generation of knowledge and innovation of the present time, formed by the profusion of structured and unstructured data available via electronic means and fed by the productions and publications of companies, governments and individuals, in the form of texts, photos, music, videos and social media contents.

According to Taurion (2014), the need for faster and faster data analysis tools is increasing, as the speed with which new data is created each day causes some data to become obsolete - for organizations - in a short time. Therefore, organizations have been keen to solve constant problems with storage space and data processing capacity; after all, for contemporary managers to acquire timely information is to acquire power (ROGERS, 2010).

Jewell et al (2014) classify the *Big Data* data into three fundamental types:

structured, semi-structured and unstructured. Currently, it is estimated that ninety percent of all data produced in the world is in unstructured form.

Big Data can also be defined through the so-called “V’s” - attributes that describe and qualify this virtual mega-mass. Initially, it was considered that the concept of *Big Data* rested on the traditional 3 V’s: *Volume*, *Velocity* and *Variety* (DUMBILL, 2012). However, two other attributes were suggested to integrate with the original ones: *Veracity* and *Value* (SAPORITO, 2013). *Validity* (KHAN, UDDIN, GUPTA, 2014, ELDER, 2014), *Visualization* (SEDDON; CURRIE, 2016), *Variability* and *Volatility* (OWAIS; HUSSEIN, 2016) appeared and were incorporated at the same time in which new observations emerged. Semantic studies (DUAN, 2015; SMITS, 2018) around data took into account the existence of another attribute little explored: *Vocabulary*. Finally, two other criteria were evaluated - *Venue* (BALAANAND et al 2017) and *Vulnerability* (ALLODI; MASSACCI, 2017; AYDINOGLU; BOVKIR, 2018) - concepts that, later, were integrated to the existing ones.

In a scenario where data is treated as the new natural resource of efficient organizations, to develop the ability to collect, store, classify, visualize and apply intelligence and innovation through analysis and interpretation is seen as a major organizational differential. This would be a reasonably simple task if the volume in *Big Data* were constant, uniform, clear, and debugged - which is not always the case.

Nowadays, the great difficulty for public and private organizations to adapt to this volatile and contingent scenario is precisely to deal with the characteristic attributes of Big Data. Paradoxically, these attributes - which express opportunities for organizations - also express major challenges.

For a better understanding of this concept (KHAN; UDDIN; GUPTA, 2014; OWAIS; HUSSEIN, 2016; BALAANAND et al., 2017; ALLODI; MASSACCI, 2017; AYDINOGLU; BOVKIR, 2018) each of the twelve attributes mentioned above will be listed and defined:

a) “*Volume*” attribute: it refers to the dimensional characteristic that *Big Data* occupies: a virtual mass of about 1 billion Terabytes composing the world Data platform. The challenge of contemporary organizations here is to store large volumes in limited virtual spaces;

b) “*Velocity*” attribute: it concerns both the speed with which data travels on this network and the speed with which new data are generated. The challenge of contemporary organizations here is to follow continuous and rapid production;

c) “*Variety*” attribute: should be appreciated as synonymous with heterogeneity, diversity: a whole formed by parts (data) of different natures, sizes, extensions and formats. The challenge of contemporary organizations here is to adapt to data processing with totally different natures;

d) “*Veracity*” attribute: refers to the importance of filtering the contents of *Big Data*, which are not always worthy of credibility. The challenge of contemporary

organizations here is to discern between what is true and what is false;

e) “*Value*” attribute: is related to the usefulness, the potential importance that *Big Data* data possess to add value to other contents. This attribute is directly related to the information concept conceptualized by Wiener (1948), which proposed that data processed from old information could be useful to give rise to new information of value. The challenge of contemporary organizations here is to evaluate between what can be useful and what can not;

f) “*Visualization*” attribute: also called “*Visibility*”, refers to the capacity that data - when properly processed - have to explain naturally hidden facts. This idea corroborates the epistemological concept of “information as potentiality” (RASCÃO, 2006, p.33), since it assumes that in *Big Data*, data - even if initially meaningless - constitute potential information. The challenge for contemporary organizations here is to have the right visualization tools for exploring raw content and thus to discover the hidden potential of the data;

g) “*Volatility*” attribute: unlike what used to happen two or three decades ago, today an organization can no longer keep the same data indefinitely; because they occupy a huge space, *Big Data* data need to be re-evaluated and discarded from time to time. Even if they are within “validity”, data tend to be discarded after they have been thoroughly analyzed and applied. The challenge of the organizations here is to assertively select the data without applicability to be discarded;

h) “*Vocabulary*” attribute: considered by some as “*Verbality*”, this attribute refers to the semantic-descriptive content stored in *Big Data* that, paradoxically, comes up against the difficulty of representation through metadata, indexing, ontologies and taxonomies. This thematic wealth of *Big Data* alternatively presents itself as an opportunity, and as a problem: an opportunity as it may consist of an immeasurable source of new ideas; a problem because there are enormous difficulties in defining and classifying unstructured data, which is why the concept of “*Vocabulary*” is sometimes referred to as “*Vagueness*”. The challenge for contemporary organizations here is to adequately label these hard-to-define content so they can store and retrieve them in the future;

i) “*Venue*” attribute: it talks about the situational nature of data in *Big Data*, which leads to reflect about the context of production, collection, analysis and application of these data. The challenge for contemporary organizations here is to understand the relationship between the scope of application of particular sets of data and the context from which they were drawn;

j) “*Variability*” attribute: also described as “*Viscosity*”, it stems from “*Velocity*”. It refers to the dynamic, fluid and unstable behavior of data; in *Big Data*, no perception is one hundred percent right, they are just odds. The challenge of contemporary organizations here is to manage contingencies and variations;

k) “*Validity*” attribute: is one of the most striking features in *Big Data*; it points to

the fact that the use of a given data has *timing*, that is, an optimal moment and a validity period. The challenge of contemporary organizations here is to seize opportunities in a timely manner;

1) “*Vulnerability*” attribute: while a culture based on data forecasting can be seen as an opportunity to add more security to decision making, it can be highly worrisome for managers to work with data that may be the gateway to contamination or breach of security protocols. The challenge of contemporary organizations here is to deal with all kinds of data without forgetting security.

It should be noted, therefore, that the complexity involving both the concept and the nature and applicability of *Big Data* often make it difficult to explore and apply. *Big Data* can be an excellent raw material for brainstorming, forecasting and intelligence generation, both in corporate environments and government agencies.

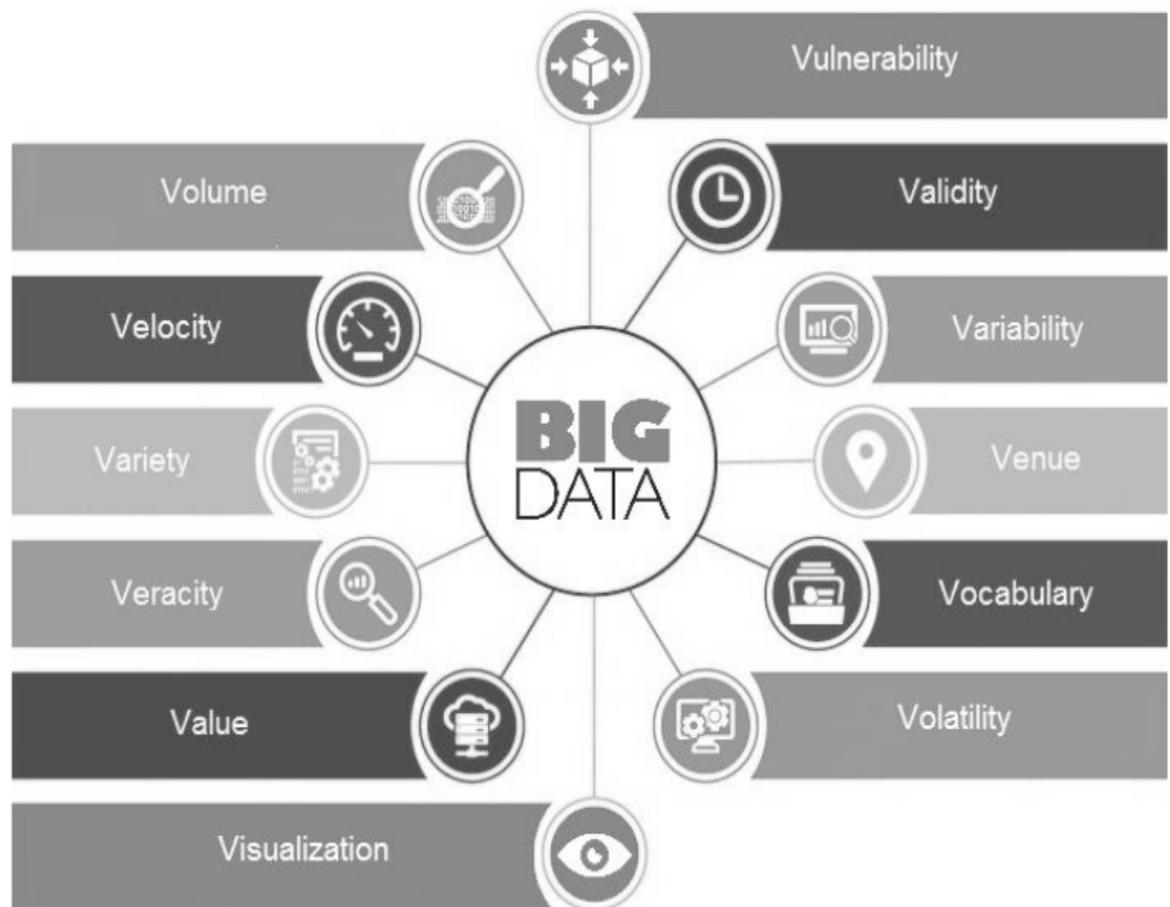


Figure 1 - The 12 V's in *Big Data*

Font: Adapted by author, 2018

However, for a successful application and exploitation of opportunities it is necessary – prior to giving any step related to the data - to have a clear idea about the questions that one wishes to answer. The visualization and analysis of data are intended to answer one or more problem questions, which must be well defined before the beginning of each process (TAURION, 2014).

In this way, there can be said to be four types of basic approaches *Big Data* can be worked with (MAFRA, 2017). They are: Descriptive, Diagnostic, Predictive and Prescriptive. Barga, Fontama and Hyong (2015) and Mafra (2017) argue that the Descriptive and Diagnostic approaches dissect the past, while the Predictive and Prescriptive ones project their gaze to the future.

Once the question (or questions) to be answered has been defined, the manager proceeds with the six steps for data analysis, to which Fayyad, Piatetsky-Shapiro and Smyth (1996) report: Selection and Extraction, Processing, Transformation, Data Mining (DM), Interpretation, Evaluation, and Application.

Therefore, it is observed that not only the origin of the data or the software technologies employed are important in the process of *Big Data* analysis, but also the methodology of study and the work of the researcher.

Finally, in order to obtain successful analyzes and responses from data, it is indispensable to define well what one wishes to respond to, to know the attributes of *Big Data* well, to understand the focuses of approach and to correctly apply the stages of evaluation and action.

5 | EXAMINING INFORMATION

For decades, there has been growing interest in content related to scientific and technological production - by researchers and those responsible for studying and creating public policies on security, development, health and education. Professionals related to research and development, both in the private and public sectors have sought in Information Science methods, tools and techniques to help them identify, select, process and extract scientific information for policies, academic assistance and decision support.

Thus universities, research centers and companies have focused on the study of productivity metrics, impact factor, collaboration studies and other investigations related to authors and contents in Science, using for this criteria and quali-quantitative processes explored in Scientometry.

It is called Scientometry the infometrics branch dedicated to the investigation of the various aspects related to the quantification and measurement of information in scientific production. Due to his contribution, Derek Solla Price was identified as the father of Scientometry. In his book *Little Science, Big Science*, published in 1963, Price had not yet coined the term "*Scientometry*", however, that work was of enormous importance for its development (GARFIELD, 2007). It is observed that much of the methodology and fundamentals adopted by Price, and later by Garfield, for the development of the scientometric field derived from bibliometric theoretical bases - notably older ones. For this reason, many authors today recognize Scientometry as a branch of Bibliometry, field of library science (HOOD, WINSON, 2001).

Scientometry, therefore - as a bibliometric branch - is part of the Social Sciences

group dedicated to supporting the development of scientific, education, technology and innovation policies (BUFREM and PRATES, 2005). Among its applications, scientometric studies focus: the occurrence analysis, frequency and impact of publications of scientific content, the mapping of productivity by author and by institution, and the exploration of complex networks involving terms and relations of collaboration in Sciences.

Surveys such as these enable development agencies, education and research institutions, and Public Administration itself to improve its policies and strategies for hiring staff, priorities establishment and resources allocation.

6 | METHODOLOGICAL PROCEDURES

Both the study of *Big Data* (related to Technological Sciences), and Scientometry (belonging to the Applied Social Sciences) work with data analysis and statistical methods in its methodology. For this reason Bibliometry - from which the foundations for the development of Scientometry originated - was initially called Statistical Bibliography (GUEDES; BORSCHIVER, 2005). Thus, it can be said that both procedures - with data - apply to the exercise of techniques for the visualization and analysis of information, in order to originate quali-quantitative studies of descriptive nature.

Therefore, this research sought to demonstrate how data produced from informational behavior of digital users in searches can be related to scientific productions data and - at the same time – to the interests of the Brazilian Ministry of Health. For this purpose, it was developed a scientometric study of terms in the Health field using the *Web of Science* and *Scopus* databases. The timeline cut was limited between May 2013 and June 2018 in Brazil.

The highlighted terms in the scientometrics study were then analyzed under the user's perspective in their electronic demands for information. For this, the *Google Trends*© - a tool for visualization of search trends in *Google*© which operates through the data mining engine using available data in *Big Data* - was selected. Online software *Google Trends*© operates by ranking the most used terms in searches made by users in *Google*©, the most used metasearcher in Brazil. According to studies, 94.31% of the online researches in Brasil are made through the *Google Brasil*© website, followed by 2.05% through *Google's* international website (SERASA EXPERIAN, 2015).

For experimentation and methodological demonstration, an exploratory-descriptive research was developed, using a quali-quantitative approach.

Finally, we will analyze the patterns of information evidenced among the three types of data sets - the first one, whose data were indexed from search attitudes in a popular search engine - the second, extracted from the two scientific databases, and the third, obtained from information sources of the Brazilian Government.

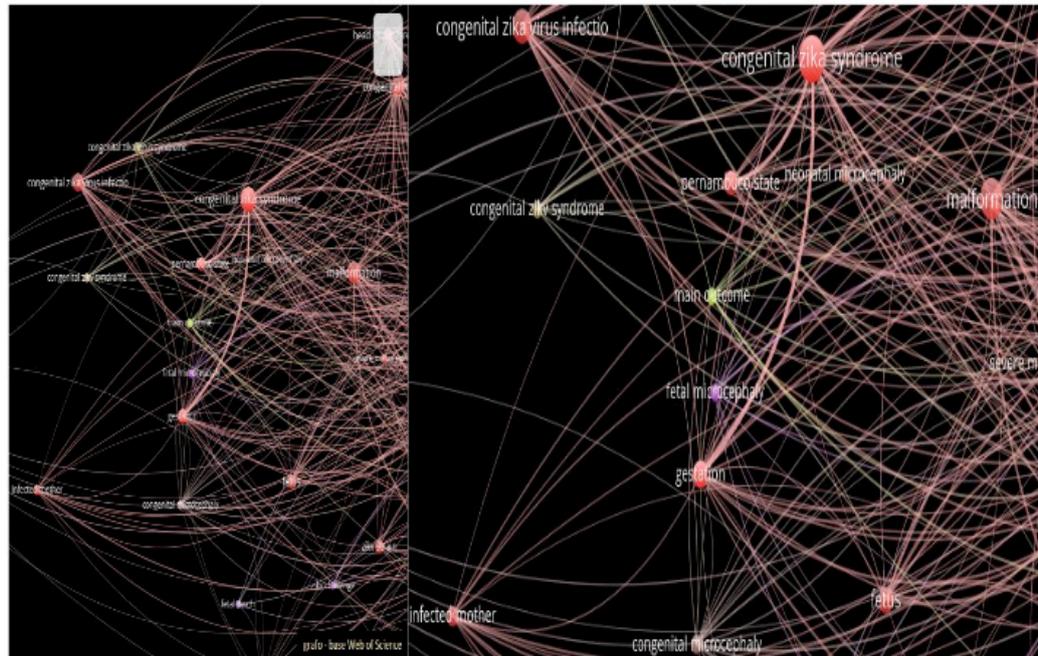


Figure 3 - VosVIEWER© software screenshot with *Web of Science* base data graph.

Font: Research data, 2018

It was observed that in the interstice from May 25 2013 to June 24 2018, the demands for information related to the terms “*Microcephaly*” (blue chart) and “*Zika*” (red chart) obtained maximum peaks of interest in exactly the same period (**Figure 4**): in the second week of December 2015 and in the second week of February 2016 in *Google Trends*©.

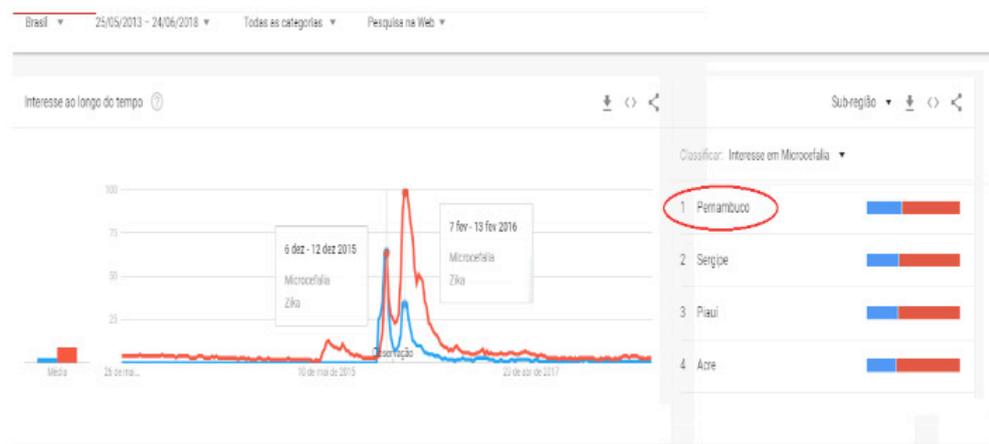


Figure 4 - Screenshot of the *Google Trends*© tool with analysis graphics for the *Big Data* content on the terms “*Zika*” and “*Microcephaly*” in the period from 05/25/2013 to 06/24/2018 in Brazil.

Font: Research data, 2018

It was verified that, among the sub-regions with the highest demand for search terms, the state of Pernambuco stands out leading the number of popular searches (**Figure 4**). It was also observed that, in the years 2015 and 2016, there were two peaks of searches related to the themes “*Microcephaly*” (blue line) and “*Zika*” (red line).

Associating this data with official data on cases of microcephaly, made available by the Ministry of Health in 2015 (**Table 1**), it was noted that Pernambuco stood out in this period as the state with the highest number of cases of microcephaly resulting from infection of newborns by Zika virus in the gestation period: 150 cases reported and 1,031 cases suspected only in this state. These numbers put Pernambuco with more than 37% of all cases registered in the country, and with more than twice as many cases as those reported in the state that appears in second place.

Distribuição dos casos suspeitos de microcefalia notificados à SVS/MS até a semana epidemiológica 50, por número de municípios e Unidade da Federação de residência, Brasil, 2015

Unidade da Federação	Total de municípios com casos notificados	Casos suspeitos de microcefalia relacionada ao vírus Zika		Óbitos suspeitos (n)
		n	%	
Centro-Oeste				
Distrito Federal	1	11	0,4	1
Goiás	12	40	1,44	0
Mato Grosso	10	78	2,8	0
Mato Grosso do Sul	2	3	0,11	0
Nordeste				
Alagoas	44	114	4,1	0
Bahia	64	271	9,74	10
Ceará	30	127	4,57	1
Maranhão	30	88	3,16	1
Paraíba	69	429	15,42	5
Pernambuco	150	1.031	37,06	3
Piauí	21	51	1,83	1
Rio Grande do Norte	42	154	5,54	10
Sergipe	40	136	4,89	5
Norte				
Pará	8	32	1,15	0
Tocantins	27	58	2,08	0
Sul				
Rio Grande do Sul	1	1	0,04	0
Sudeste				
Espírito Santo	10	18	0,65	0
São Paulo	6	6	0,22	0
Minas Gerais	33	52	1,87	1
Rio de Janeiro	18	82	2,95	2
Brasil	618	2.782	100	40

Table 1 - Reported and suspected cases of microcephaly in Brazil, in 2015.

Font: Brazil, 2015

It was also observed that, in the data mining research generated by searches for the term “*Microcephaly*”, in Brazil (**Figure 4**), the greater national demand for related information came from the population of Pernambuco.

Relating these cases of microcephaly to the official data of the Ministry of Health on notifications of contamination by the Zika virus between 2016 and 2018 (**Figure 5**), it is observed that the peak of cases in the three years occurs in the 7th Week of 2016, that is, around the third week of February 2016.

Attention is drawn to the fact that the second highest peak in information demand related to “*Zika*” and “*Microcephaly*” - recorded in the second week of February 2016 (**Figure 4**) - occurred about ten days before an outbreak of Zika, recorded in the middle of the third week of February 2016 (**Figure 5**).

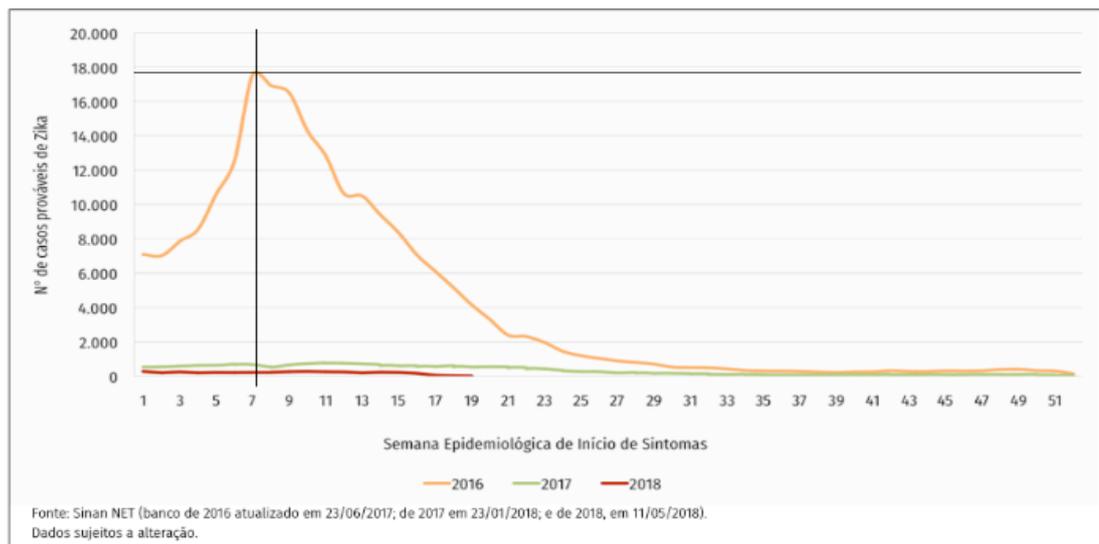


Figure 5 - Reported and suspected cases of acute illness by Zika virus, in Brazil, between 2016-2018.

Font: Sinan NET, 2018

It was observed, therefore, that there is a relation between peaks of volume of demand for information and peaks of endemic outbreaks. It was noted that the maximum peaks of information demand precede, in a short time, the maximum peaks of endemic crisis.

8 | CONSIDERATIONS

From the discussions developed, it is inferred that the existing synchronism among the peaks of data volumes in the *Google Trends*® chart, for the different terms, may be reflecting possible thematic associations established by digital users to define search criteria for information.

There was a strong relationship between the geographical location of the applicant user and the main geographic area from which the endemic outbreaks started, as well as a strong relationship of chronological proximity between the endemic outbreaks and the searches made for terms that represented them. These spatiotemporal relationships, which align informational demand behaviors (analyzed from data produced in *Big Data* environment) with official endemics data (published by the Brazilian Ministry of Health), suggest that the digital user acts - or reacts - driven by local and timely phenomena. In this context, aspects investigated here - such as locality (addressed under the term “*Venue*”) and timing (referred to as “*Validity*”), inherent in the processes of extraction and analysis of value data (referring to “*Value*”) for the discovery of useful information from *Big Data* - are remarkably instigating, revealing and relevant.

It was also evaluated that information provided by discursive genres related to the theme (scientific information offerings - dissected through scientometric study) derives from these perceptions of informational need; however, such studies are not always

timely in meeting population demands.

It was also observed that the peaks in demand for information occurred shortly before the endemic peaks recorded by the Ministry of Health and Secretariats. Thus, the peaks of searches for the term “Zika” with national demands headed by the state of Pernambuco (in the second week of December 2015 and in the second week of February 2016) were followed by peaks of Zika outbreak exactly in Pernambuco (in the third week of February 2016).

It was therefore assessed that informational behaviors of local users, manifested through informational attitudes with extraordinary volumes, seem to precede (and herald) critical peaks of impending social disorder.

It was found that data generated from informational “lanes” can reveal implicit information about the state, the location, and the political-economic-social moment in which the information user is. This occurs insofar as data generated, from an informational behavior, manifest the temporary and contextualized need of a specific group of regional users.

In the same way, it was possible to explore - methodologically - the intersections between scientometric research and data analysis from *Big Data* obtaining from them important contributions to the practice of extracting valuable information to assist organizations in the definition of scenarios and trends.

So, it is believed that judicious analyzes developed from metric studies - with data and information - can contribute to the construction of indicators that support research, policies and decision-making processes in organizations - public or private - inserted in the context of the Information and Knowledge Society.

REFERENCES

ALLODI, L; MASSACCI, F. Security events and vulnerability data for cybersecurity risk estimation. **Journal of Society for Risk Analysis**, n. 37, v. 8, Aug. 2017.

AYDINOGLU, A; BOVKIR, R. **Management of big geographic data for smart city applications**. FIG Congress, Istanbul, 2018.

BALANAND, M et al. A survey on *Big Data* with various v's on comparison of apache hadoop and apache spark. **Advances in Natural and Applied Sciences**, v.11, p. 362, 2017.

BARGA, R; FONTAMA, V; HYONG, T. **Predictive analytics with microsoft azure machine learning**. 2. ed. New York: Springer Sci + BM, New York, 2015.

BECKER, J. **Estatística básica: transformando dados**. Porto Alegre: Bookman, 2015.

BRASIL. **Sinan NET – Casos notificados e casos suspeitos de doença aguda por Zika vírus, no Brasil, entre 2016 e 2018**. Brasília, DF, 2018.

BRASIL, Secretaria de Vigilância em Saúde - Ministério da Saúde. **Boletim Epidemiológico**, Brasília, DF, v. 46, n° 6, 2015. ISSN 2358-9450.

BUFREM, L; PRATES, Y. O saber científico registrado e as práticas de mensuração da informação. **Ciência da Informação**, Brasília, v. 34, n. 2, p. 9-25, maio/ago. 2005.

DUAN, H et al. Distributed in-memory vocabulary tree for real-time retrieval of *Big Data* images. **Ad Hoc Networks**, v.35, p.137-148, dez. 2015.

DUMBILL, E. **Planning for Big Data**. Massachusetts: O'Reilly Media Inc., 2012.

ELDER, J. **Evaluate the validity of your discovery with target shuffling**. [S.l.]: ER, 2014.

FAYYAD, U, PIATETSKY-SHAPIRO, G; SMYTH. P. **From data mining to knowledge discovery in databases**. v.17, n.3, Providence, 1996.

FIALHO, J; ANDRADE, M. Comportamento informacional de crianças e adolescentes: uma revisão da literatura estrangeira. **Ciência da Informação**, Brasília, v. 36, n. 1, p. 20-34, 2007.

GARFIELD, E. From the science of science to scientometrics visualizing the history of science with HistCite Software. In: **ISSI**, 11., 2007, Madrid. Anais... Madrid, 25 jun. 2007.

GARTNER, Inc. **IT glossary**. New York: Gartner, [2013].

GINMAN, M. Information Culture and Business performance. **IATUL Quarterly**, v. 2, 1987.

GUEDES, V; BORSCHIVER, S. Bibliometria: uma ferramenta estatística para a gestão da informação e do conhecimento, em sistemas de informação, de comunicação e de avaliação científica e tecnológica. In: **ENANCIB**, 6., 2005, Salvador: UFBA, 2005.

HOOD, W; WILSON, C. The literature of bibliometrics, scientometrics, and informetrics. **Scientometrics**, v. 52, n. 5. 2001.

JEWELL, D. et al. **Performance and capacity implications for Big Data**. [S.l.: s.n.], 2014.

KHAN, Ali-ud-din; UDDIN, Muhammad Fahim; GUPTA, Navarun, 2014. **Seven v's of Big Data: understanding Big Data to extract value**. Conference of the American Society for Engineering Education, I, Bridgeport: IEEE, 2014.

MAFRA, F. **Business intelligence com sql server 2012 na prática**. [S.l.]: Alpha, 2017.

NEELY, A; GREGORY, M; PLATTS, K. Performance measurement system design: a literature review and research agenda. **IJO & PM**, [S.l.], v. 15, n. 4, p. 80–116. 1995.

OWAIS, S; HUSSEIN, N. Extract five categories CPIVW from the 9v's characteristics of the *Big Data*. **IJACSA**, v.7, n.3, 2016.

RASCÃO, J. **Da gestão estratégica à gestão estratégica da informação**: como aumentar o tempo disponível para a tomada de decisão estratégica. Rio de Janeiro: E-papers, 2006.

ROGERS, S. Information is power. **The Guardian**, London, 2010.

SAPORITO, P. The 5 v's of *Big Data*: value and veracity join three more crucial attributes that carriers should consider when developing a *Big Data* vision. **Best's Review**, v. 114, 2013.

SCHWAB, K. **A Quarta Revolução Industrial**. São Paulo: Edipro, 2016.

SEDDON, J; CURRIE, W. A model for unpacking *Big Data* analytics in high-frequency trading. **Journal of Business Research**, v. 70, p.300-307, 2016.

SERASA EXPERIAN. **Google Brasil tem 94,31% de participação nas buscas em dezembro**,

segundo Hitwise. São Carlos: Serasa Experian, 2015.

SHANNON, C; WAEVER, W. The mathematical theory of communication. **Bell System Technical Journal**. 1948. p.379-423.

SILVA, C; MOREIRA, J. **Comportamento informacional vs comunicação**: aplicação de modelos em contextos multidisciplinares. [S.l.: s.n.], 2014.

SMITS, G. et al. A soft computing approach to *Big Data* summarization. **FSets & S**, 2018.

TARAPANOFF, K. (Org.). **Inteligência organizacional e competitiva**. Brasília: UNB, 2001.

TAURION, C. Entre os v's do *Big Data*, velocidade cresce em importância. **CIO**, [S.l.], 2014.

VALENTIM, M. Informação e conhecimento no contexto de ambientes organizacionais. In: _____. **Gestão, mediação e uso da informação**. São Paulo: Cultura Acadêmica, 2010.

WIENER, N. **Cybernetics**: or the control and communication in the animal and the machine. Massachusetts: Institute of Technology, 1948.

DEMOCHAIN - FRAMEWORK DESTINADO A CRIAÇÃO DE REDES BLOCKCHAIN HÍBRIDAS PARA DISPOSITIVOS IOT

Lorenzo W. Freitas

Departamento de Engenharias e Ciência da
Computação
Universidade Regional Integrada do Alto Uruguai
e das Missões – Santo Ângelo, RS – Brazil
lorenzofreitas@aluno.santoangelo.uri.br

Carlos Oberdan Rolim

Departamento de Engenharias e Ciência da
Computação
Universidade Regional Integrada do Alto Uruguai
e das Missões – Santo Ângelo, RS – Brazil
ober@san.uri.br

RESUMO: O uso da tecnologia Blockchain no contexto da IoT (Internet of Things) está sendo cada vez mais explorado pela comunidade acadêmica e a indústria. No entanto, essa implantação pode ser custosa ou inviável pois a Blockchain pode exigir recursos computacionais que não são obtidos facilmente com o uso de dispositivos IoT. Assim, esse trabalho apresenta o framework Demochain cuja função é auxiliar no desenvolvimento de plataformas blockchains híbridas no contexto de IoT.

ABSTRACT: The use of Blockchain technology in the context of Internet of Things (IoT) is increasingly being explored by the academic community and industry. However, such deployment may be costly or impracticable

since Blockchain may require computational resources that are not easily obtained with the use of IoT devices. Thus, this work presents the Demochain framework whose function is to assist in the development of hybrid blockchains platforms in the context of IoT.

1 | INTRODUÇÃO

Uma importante área de pesquisa na Computação que está apresentando um rápido crescimento é a Internet das Coisas (IoT). Segundo Gartner (2017), o gasto total em dispositivos e serviços de IoT atingiu quase US \$ 2 trilhões em 2017, e haverá mais de 20 bilhões de “coisas” conectadas em todo o mundo até 2020.

Entretanto, apesar de possuir a natureza distribuída, a maioria das soluções IoT ainda dependem de arquiteturas que seguem o modelo cliente servidor. Embora esse modelo arquitetural possa funcionar hoje, o crescimento da IoT sugere que novas arquiteturas deverão ser propostas no futuro [Caramés e Lamas 2017]. Uma alternativa que vêm sendo explorada pela comunidade acadêmica e a indústria é o emprego de blockchains devido a sua, capacidade de manter os registros imutáveis sem perder segurança, com algoritmos que tratam nodos maliciosos.

Com esse cenário e a falta de padronização dos dispositivos IoT devido a seus diferentes objetivos existe uma dificuldade na implementação e modelagem de uma rede blockchain personalizada para dispositivos IoT [Conoscenti et al, 2016]. Assim, o presente trabalho apresenta o Demochain, um framework voltado para a criação de redes blockchain híbridas para dispositivos IoT. Um dos diferenciais do Demochain é a sua capacidade de oferecer opções para mesclar e combinar diferentes níveis na arquitetura e também funcionalidades da blockchain pura (totalmente descentralizada), variando seus protocolos e criptografias, construindo assim, uma blockchain híbrida. Além disso, pode-se ressaltar que a maior contribuição do Demochain é a possibilidade de facilitar a modelagem e a prototipação de novas redes blockchain em ambientes variados.

2 | FRAMEWORK PROPOSTO

O framework construído foi chamado de Demochain e foi pensado para diminuir a complexidade na conexão dos dispositivos. Portanto, vários conceitos utilizados em blockchains tradicionais foram simplificados. Ele foi desenvolvido com a linguagem Go [Google 2009] e foi utilizada uma abordagem de orientação a objetos com o padrão Decorator em sua implementação [Schmager, Frank 2010]. A arquitetura que foi implementada é multicamadas, baseada no trabalho de Li e Zhang (2017). Nessa arquitetura, cada nodo na rede pode assumir dois papéis: ser um nodo centralizador (Edge Node), que serve para controlar o acesso dos nodos abaixo na camada, ou ser um nodo de alto nível (High-Level Node), também chamado de nodo filho, onde são adicionado blocos e executado o protocolo de consenso.

Com isso tem-se um modelo descentralizado com alguns níveis de centralização, porém não totalmente distribuído. A Figura 1 apresenta a arquitetura do Demochain.

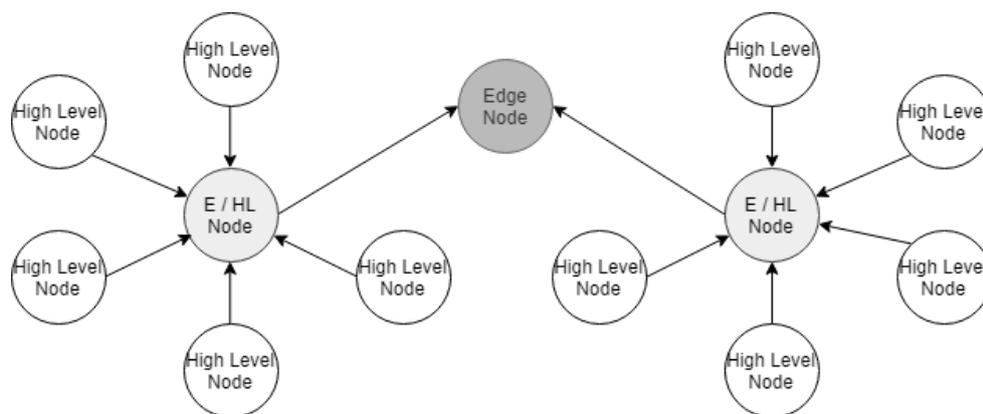


Figura 1. Arquitetura implementada.

Ao escolher o esquema de criptografia para o framework, foi levado em consideração não apenas a segurança fornecida de acordo com a carga computacional, mas também o consumo de energia, fazendo um trade-off entre segurança e consumo. Foram implementados quatro algoritmos de criptografia assimétrica: RSA, Ed25519,

Secp256k1 e ECDSA. A identificação dos nodos ocorre utilizando a pilha TCP em sua forma tradicional. Portanto, um mesmo dispositivo pode estar rodando várias instâncias do Demochain, desde que em portas distintas.

No Demochain a blockchain é replicada para todos os nós. Entretanto o diferencial deste framework é que a execução do consenso é por nível. Protocolo consenso consiste em um mecanismo que determina as condições a serem alcançadas para concluir que um acordo foi alcançado em relação às validações dos blocos para ser adicionado ao blockchain [Zheng et al. 2017]. Foram implementados três protocolos de consenso: PoW (Proof of Work), PoS (Proof of Stake) e PBFT (Practical Byzantine Fault Tolerance).

2.1 Proof of Work (PoW)

Algoritmos de consenso PoW, ou Prova de Trabalho baseiam-se no fato de que um nó malicioso tem que executar muito trabalho do ponto de vista computacional para atacar a rede, e por isso, é menos provável que ele vá querer atacar. O trabalho realizado geralmente envolve fazer alguns cálculos até que uma solução seja encontrada, um processo que é comumente conhecido como mineração. No caso da blockchain do Bitcoin, a mineração consiste em encontrar um número aleatório, chamado de número nonce que fará o hash SHA256 do cabeçalho do bloco para ter no início certo número de zeros. Portanto, os mineradores têm que demonstrar que eles realizaram certa quantidade de trabalho para resolver o problema. Uma vez resolvido o problema, é realmente fácil para outros nodos verificar se a resposta obtida é válida. No entanto, este processo de mineração faz a blockchain ineficiente em taxa de transferência, escalabilidade [M. Vukolić 2015], e também em termos de consumo de energia, o que não é desejável em uma IoT rede.

2.2 Proof of Stake (PoS)

PoS ou Prova de Participação é um mecanismo de consenso que requer menos recursos computacionais e potência do que PoW [BitFury Group 2015], por isso consome menos energia. Em uma blockchain baseado em PoS, assume-se que as entidades com mais participação na rede são os menos interessados em atacá-lo. Assim, os nodos precisam provar periodicamente que eles possuem certa quantidade de participação na rede (por exemplo, moeda ou quantidade de dados coletados a partir de sensores). Podemos comparar o PoS há uma loteria, onde é realizado um sorteio e escolhido um nodo que irá dizer se o bloco que está sendo enviado pela rede é válido ou não, por consequência, o nodo que tiver mais participação recebe mais “fichas” neste sorteio, e tem mais chances de validar o bloco. O nodo que realiza o sorteio pode ou não ser pré-definido.

2.3 Practical Byzantine Fault Tolerance (PBFT)

PBFT [M. Castro e B. Liskov 1999] ou Tolerância de Falhas Bizantinas Prática é um algoritmo de consenso que resolve o problema dos generais bizantinos, recomendado para ambientes assíncronos. PBFT assume que menos de um terço dos nós são maliciosos. Para cada bloco a ser adicionado à cadeia, um líder é selecionado para ser encarregado de encomendar a transação. Essa seleção deve ser apoiada por pelo menos $2/3$ de todos os nós, que devem ser conhecidos pela rede.

2.4 Implementação do consenso

No Demochain o PoW funciona da maneira clássica, é imposta uma dificuldade e é realizada uma tentativa aleatória de encontrar uma Hash utilizando SHA256 que comece com o número de zeros da dificuldade. Na figura 2 temos um exemplo de Hash's geradas de acordo com a dificuldade do PoW, onde é possível verificar o número de zeros no início da Hash.

Dificuldade 1

012b46ed9cf86cab3fd08708ec3ee868528044e87931b2d00dd21c4fbae5d919

Dificuldade 2

001ca89c82993d99e3477a11c08687caa5c8fb6760cd93ef25de45dc2746e8b4

Dificuldade 3

000ca89c82993d99e3477a11c08687caa5c8fb6760cd93ef25de45dc2746e8b4

Figura 2. Dificuldade do Proof of Work.

Como o PoW processa os blocos apenas com os próprios recursos, sem depender da aprovação dos demais nodos, no Demochain utilizando o PoW é possível realizar a validação e geração de novos blocos independente da conexão com o seu Edge Node, portanto mesmo um dispositivo estando desconectado da rede, ao se reconectar ele consegue sincronizar os blocos validados, isso se ele estiver executando a mesma instância, pois esses blocos ficam armazenados em memória, entretanto caso sua execução seja interrompida e ele não estiver conectado na rede os blocos validados são perdidos, já que foi implementado apenas um arquivo físico de Blockchain que mantém o conteúdo atualizado conforme o conteúdo atual que está circulando na rede. Uma implementação com dois arquivos de Blockchain, um atualizado conforme a rede e outro com o conteúdo "local" é uma possível melhoria, conforme é sugerido em [Liang et al. 2017].

A razão dessa funcionalidade é dar liberdade ao desenvolvedor dizer quando deve ser o envio dos blocos, não precisando ser necessariamente logo após a validação dos mesmos, o que em um cenário de concorrência de criptomoedas não faz sentido, mas em redes privadas pode fazer, dependendo dos objetivos de cada um.

A implementação proposta para o PoS é baseada no trabalho de Dan Larimer

(2014), também chamado de DpoS (Delegate Proof of Stake). A diferença do DPOs para o PoS “clássico” é que são delegados representantes para a realização do sorteio para a governança da rede com critérios predeterminados. Nesse caso a cada bloco gerado será solicitado um sorteio ao Edge Node, que irá direcionar o bloco ao vencedor para realizar a validação. Esse sorteio acontece no escopo de apenas um nível de rede, porém após ser validado o bloco é passado aos demais níveis. A figura 3 demonstra o funcionamento do PoS.

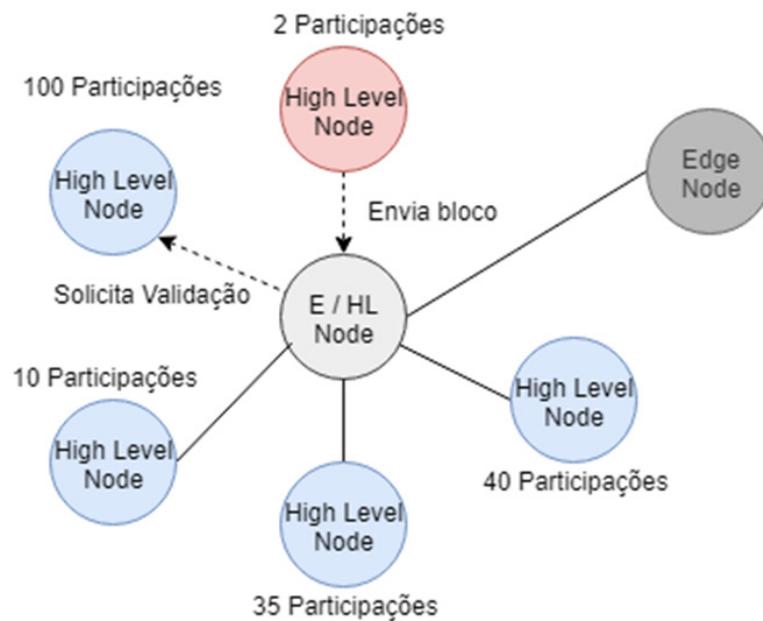


Figura 3. Funcionamento do PoS.

A implementação proposta para o PBFT também tem alterações. Assim como no PoS, toda validação deverá ser solicitado ao Edge Node da rede, porém ele irá enviar o bloco para todos os nodos do mesmo nível, ou seja, todos os High Level Nodes. Após o mesmo irá contabilizar quem aprovou e quem desaprovou o bloco. Se pelo menos 2/3 de seus High Level Nodes aprovarem o bloco, então é decretado que o bloco é válido. Assim como o PoS o protocolo executa em um nível da rede, porém ao ser validado o bloco é passado aos demais níveis. A figura 4 demonstra o funcionamento do PBFT.

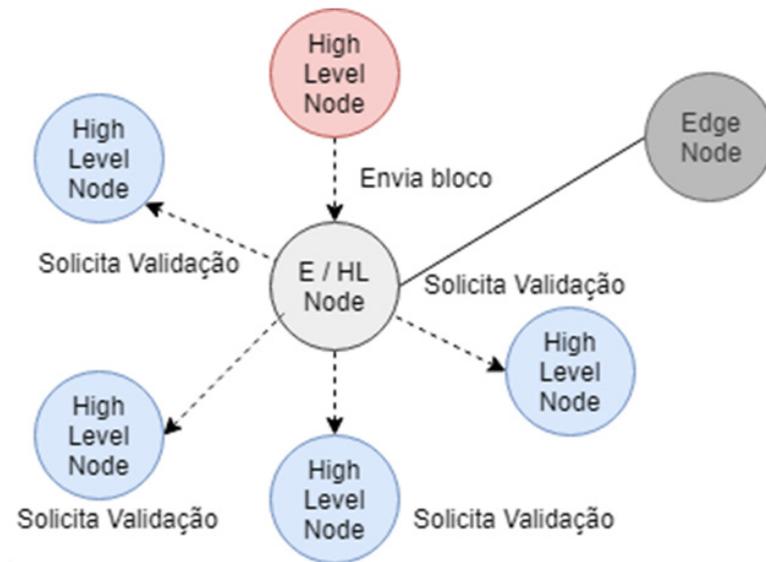


Figura 4. Funcionamento do PBFT.

Uma observação importante é que como os protocolos PoS e PBFT irão executar em apenas um nível da rede, uma preocupação que o desenvolvedor deve ter é a divisão dos Edge Nodes, pois quanto menos High Level Nodes vinculados a ele maior a chance de um nodo intruso realizar uma ação maliciosa, pois são protocolos probabilísticos.

Será suportado apenas um protocolo de consenso rodando em todos níveis de rede. Uma possível melhoria seria permitir cada High-Level Layer estar rodando um tipo de protocolo de consenso, ou mesmo não ter nenhum implementado, podendo apenas estar inserindo os blocos e os compartilhando sem protocolo de consenso, entretanto isso exigiria um esquema mais complexo de armazenamento e uma compressão dos dados validados de cada nível da rede, fazendo com que cada High-Level Layer tivesse uma blockchain diferente da outra em seu formato.

2.5 Mensagens

A troca de mensagens se dá de maneira simples, podendo ter um ou dois fluxos em cada conexão, um de recebimento e outro para envio. Caso o desenvolvedor defina uma rede não permissionada sempre será aberto os fluxos de recebimento e de envio, porém é possível definir também uma rede permissionada. Isso permite com que o desenvolvedor defina nodos apenas de backup das informações para eventuais falhas. Na figura 5 temos um exemplo ilustrativo de uma rede permissionada com o Demochain.

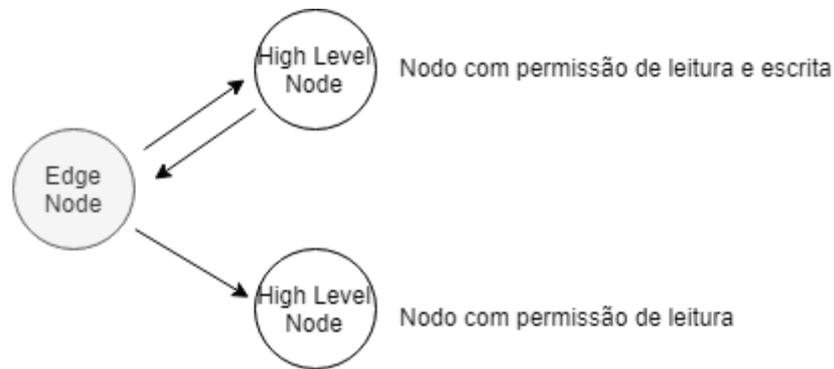


Figura 5. Rede permissionada com nodo de leitura.

Em uma primeira versão o demochain envia a Blockchain por inteira a cada mensagem trocada, o que a um longo prazo pode gerar gargalos na rede. Um controle mais complexo de troca de mensagens não foi implementado.

2.6 Estrutura dos blocos

A estrutura de blocos utilizadas pelo Demochain possui semelhanças e diferenças com as estruturas tradicionais utilizadas em criptomoedas. É importante ressaltar que normalmente em criptomoedas um bloco é formado de transações, sendo que toda transação tem um “recipient”, podendo ou não ter um “sender”. No Bitcoin por exemplo, em uma transação comum temos um “sender” que envia valor para o “recipient”, porém quando um novo bloco é minerado é chamado de “Coinbase”, pois não tem um “sender”. Como esse framework não é voltado para criptomoedas foi utilizado um conceito mais simples somente com blocos, pois a princípio a validação que será aplicada nos dados coletados não deverá ser passada para outros Nós. Abaixo temos a estrutura de bloco que o Demochain implementa:

- Index: Índice sequencial único que é incrementado a cada nova adição de blocos na blockchain.
- Timestamp: Carimbo de data e hora do momento da adição do bloco.
- Data: Dados que foram validados.
- Hash: Hash gerada para o bloco.
- PrevHash: Hash do bloco anterior.
- Consensus: Consenso utilizado para a validação do bloco.
- Nonce: Número Nonce é utilizado apenas para o consenso Proof of Work.
- Target: Target do nodo onde foi validado o bloco. Podemos dizer que corresponde ao “dono” do bloco.

2.7 Armazenamento

Para o armazenamento será gravado em um arquivo que servirá para manter os

blocos já validados por todos os nodos. O arquivo será gravado com extensão “.bc” com um Json não indentado com todos os dados da blockchain. Na figura 6 temos um exemplo de um bloco gerado com a estrutura de bloco do Demochain, e na figura 7 temos o mesmo bloco armazenado no arquivo “.bc”.

```
[
  {
    "Index":0,
    "Timestamp":"2018-11-17 19:02:25.834476 -0200 -02 m=+3.635126501",
    "Data":"0",
    "Hash":"f1534392279bddbf9d43dde8701cb5be14b82f76ec6607bf8d6ad557f60f304e",
    "PrevHash":"",
    "Consensus":{
      "TypeConsensus":1,
      "Difficulty":3
    },
    "Nonce":"",
    "Target":"QmbiS1uqyVvXzV2pYdife8vJeZZRtW4TQiNEP4Dj1ckNiG"
  }
]
```

Figura 6. Bloco json indentado.

```
[{"Index":0,"Timestamp":"2018-11-17 19:02:25.834476 -0200 -02 m=+3.635126501","Data":"0","Hash":"f1534392279bddbf9d43dde8701cb5be14b82f76ec6607bf8d6ad557f60f304e","PrevHash":"","Consensus":{"TypeConsensus":1,"Difficulty":3},"Nonce":"","Target":"QmbiS1uqyVvXzV2pYdife8vJeZZRtW4TQiNEP4Dj1ckNiG"}]
```

Figura 7. Bloco json não indentado.

A seguir serão apresentados os resultados de simulações de uso do framework com o objetivo de demonstrar a performance de cada protocolo de consenso sobre um mesmo desenho de arquitetura.

3 | RESULTADOS

Todos os códigos bem como testes realizados estão disponíveis em Freitas (2018). Os resultados foram obtidos com o uso de equipamentos para simular um ambiente IoT. A Tabela 1 demonstra a configuração do hardware utilizado para os testes do Demochain.

Código	Descrição	Processador	Memória RAM	Sistema Operacional
H01	Notebook	Intel Core i5 2x 2.20 GHz	4 Gigabytes	Windows 10 Pro
H02	Raspberry Pi	BCM2835	512 Megabytes	Raspbian Stretch Lite

Tabela 1. Tabela de hardware para experimentação.

Para testar o desempenho do framework foi utilizado as métricas descritas por S.

Angelis (2017), que são comumente usadas para medir aplicações descentralizadas. A Tabela 2 descreve as métricas utilizadas.

Código	Nome	Descrição
M01	Taxa de validação (Throughput)	Mede o número de blocos gerados com sucesso por segundo.
M02	Latência (Latency)	Mede o tempo entre envio e confirmação do bloco relativo em milissegundos.
M03	Escalabilidade (Scalability)	Média em milissegundos das mudanças de latência, conforme é aumentado o número de nodos na rede.
M04	Tempo Total (Total Time)	Mede o tempo total em minutos de execução do programa de teste.

Tabela 2. Tabela de métricas para experimentação.

A fim de simular dados reais de ambientes IoT foi utilizado um dataset público [Ortiz, J. e Gottschlich, N. 2016]. A Tabela 3 descreve o dataset utilizado.

Código	Descrição	Referência
D01	Dados de consumo de eletricidade doméstica com 60.640 medições coletadas entre janeiro de 2007 e junho de 2007 (6 meses).	[Ortiz, J. e Gottschlich, N. 2016]

Tabela 3. Tabela de datasets para experimentação.

Para testar o framework foi definido cenários de experimentação, no qual para cada cenário foi criado um aplicativo de testes que importa o Demochain. Os cenários possuem dois objetivos principais, demonstrar exemplos de como utilizar o framework e analisar o seu desempenho. Nas subseções seguintes está descrito cada cenário, seu objetivo e seus resultados.

3.1 Cenário 1

O Objetivo deste cenário é comparar desempenho de geração de blocos, utilizando as métrica M01 e M04. Para isso foi implementado um nodo único para testar o protocolo de consenso Proof of Work com diferentes níveis de dificuldade utilizado o dataset D01 como entrada de dados. A cada execução do Proof of Work foi incrementada a sua dificuldade, variando de 1 a 5. A Tabela 4 demonstra os resultados obtidos deste cenário, e as figuras 8 e 9 demonstram os resultados retirados da tabela 4

	Taxa de validação (blocos p/ segundo)	Tempo Total (minutos)
PoW 1	10 Blocos	434 min
PoW 2	8 Blocos	543 min
PoW 3	5 Blocos	869 min
PoW 4	3 Blocos	1.448 min
PoW 5	1 Bloco	4.344 min

Tabela 4. Resultados da validação do PoW aplicando as métricas M01 e M04.

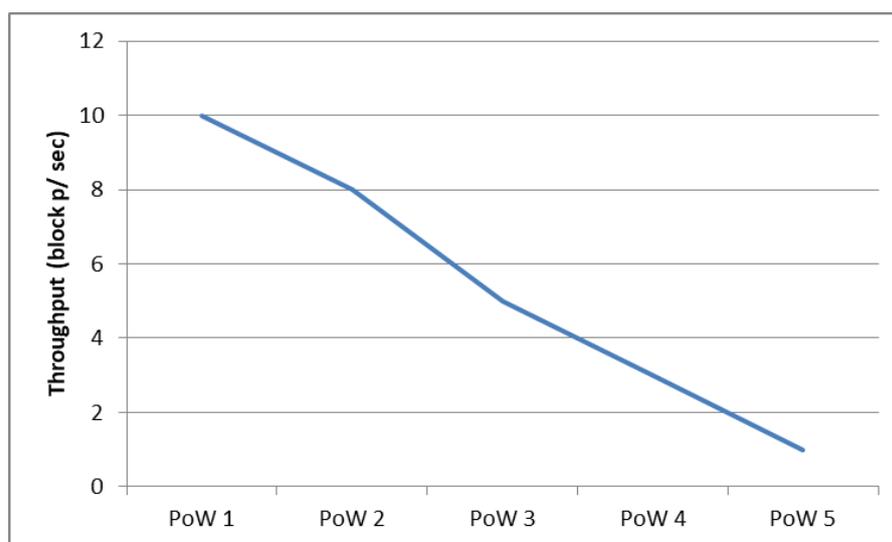


Figura 8. Aplicação da métrica M01 sobre o PoW.

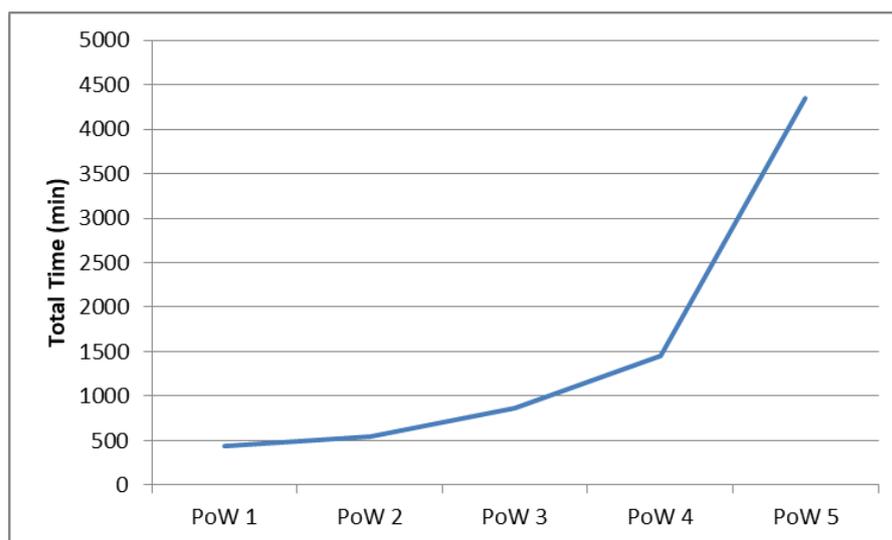


Figura 9. Aplicação da métrica M04 sobre o PoW.

Os resultados obtidos neste cenário demonstram que o Proof of Work tem um desempenho razoável na validação dos blocos até a dificuldade 3, portanto é uma boa alternativa para os nodos que não ficam conectados a todo momento, pois o seu

processamento não depende da rede.

3.2 Cenário 2

O objetivo deste cenário é comparar desempenho de cada protocolo de consenso, utilizando as métricas M01, M02, M03 e M04. Para isso foi implementado um rede com até 15 nodos dispostos nos hardwares H01 e H02, utilizando a criptografia RSA, com o dataset D01 como entrada de dados, sendo que foi utilizado o Proof of Work apenas com dificuldade 3. Todos os nodos estavam rodando sobre um mesmo nível de rede, ou seja, todos respondiam para o mesmo Edge Node.

A Tabela 5 apresenta a aplicação da métrica M01 sobre os três protocolos de consenso, considerando a variação do número de nodos na rede, e a figura 10 demonstra com um gráfico de linhas os resultados desta tabela.

	PoW 3	PoS	PBFT
1 Nodo	5 blocos	25 blocos	20 blocos
2 Nodos	4,7 blocos	21 blocos	15,7 blocos
5 Nodos	4,2 blocos	15 blocos	11 blocos
10 Nodos	3,5 blocos	10,5 blocos	6,5 blocos
15 Nodos	2,5 blocos	6,5 blocos	2,6 blocos

Tabela 5. Taxa de validação (Throughput) em blocos p/ Segundo.

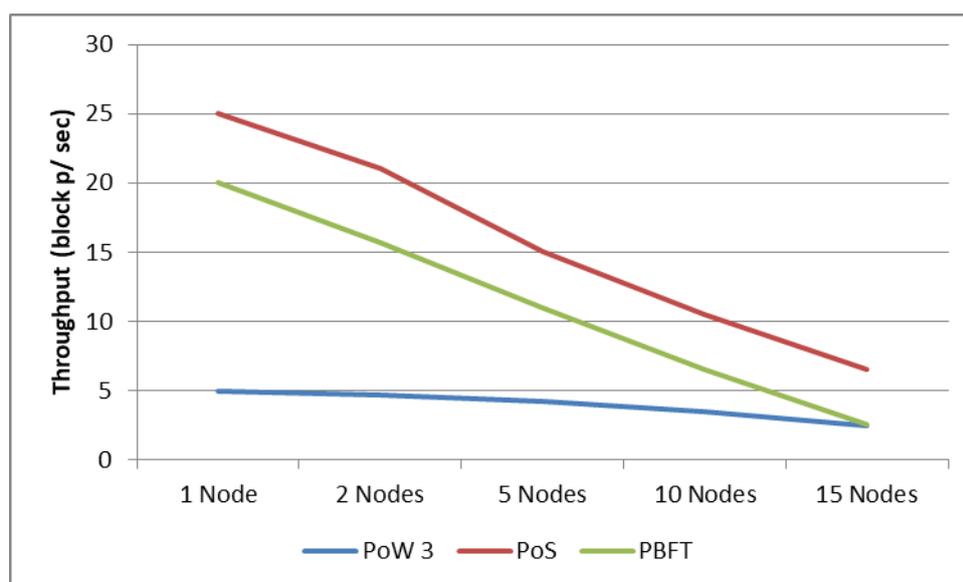


Figura 10. Aplicação da métrica M01.

A Tabela 6 apresenta a aplicação da métrica M02 sobre os três protocolos de consenso, considerando a variação do número de nodos na rede, e a figura 11 demonstra com um gráfico de linhas os resultados desta tabela.

	PoW 3	PoS	PBFT
1 Nodo	200 ms	40 ms	50 ms
2 Nodos	212 ms	47 ms	64 ms
5 Nodos	238 ms	66 ms	90 ms
10 Nodos	285 ms	96 ms	154 ms
15 Nodos	400 ms	153 ms	384 ms

Tabela 6. Latência (Latency) em milisegundos.

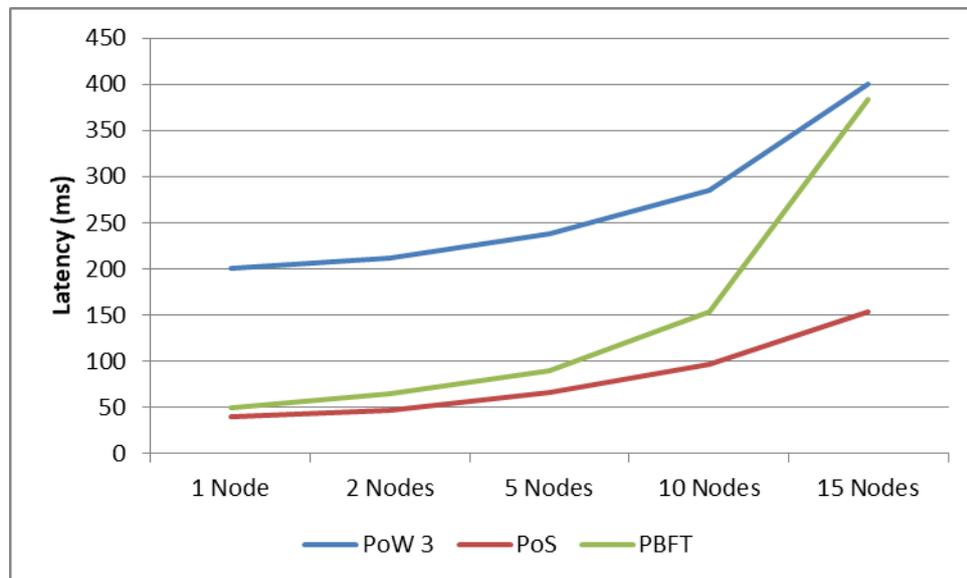


Figura 11. Aplicação da métrica M02.

A Tabela 7 apresenta a aplicação da métrica M03 sobre os três protocolos de consenso, considerando a variação do número de nós na rede, e a figura 12 demonstra com um gráfico de barras os resultados desta tabela.

	PoW 3	PoS	PBFT
Média de Latência p/ Nodo	13,33 ms	7,53 ms	22,26 ms

Tabela 7. Escalabilidade (milisegundos).

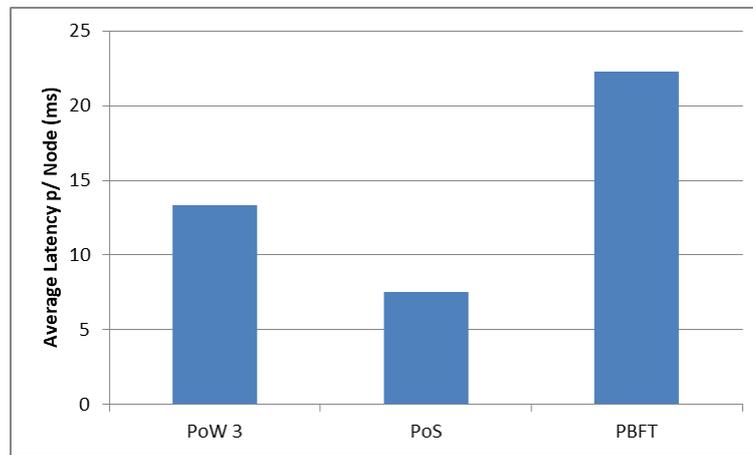


Figura 12. Aplicação da métrica M03.

A Tabela 8 apresenta a aplicação da métrica M04 sobre os três protocolos de consenso, considerando a variação do número de nodos na rede, e a figura 13 demonstra com um gráfico de linhas os resultados desta tabela.

	PoW 3	PoS	PBFT
1 Nodo	869 min	174 min	217 min
2 Nodos	462 min	103 min	138 min
5 Nodos	207 min	58 min	79 min
10 Nodos	124 min	41 min	67 min
15 Nodos	116 min	44 min	111 min

Tabela 8. Tempo total da execução.

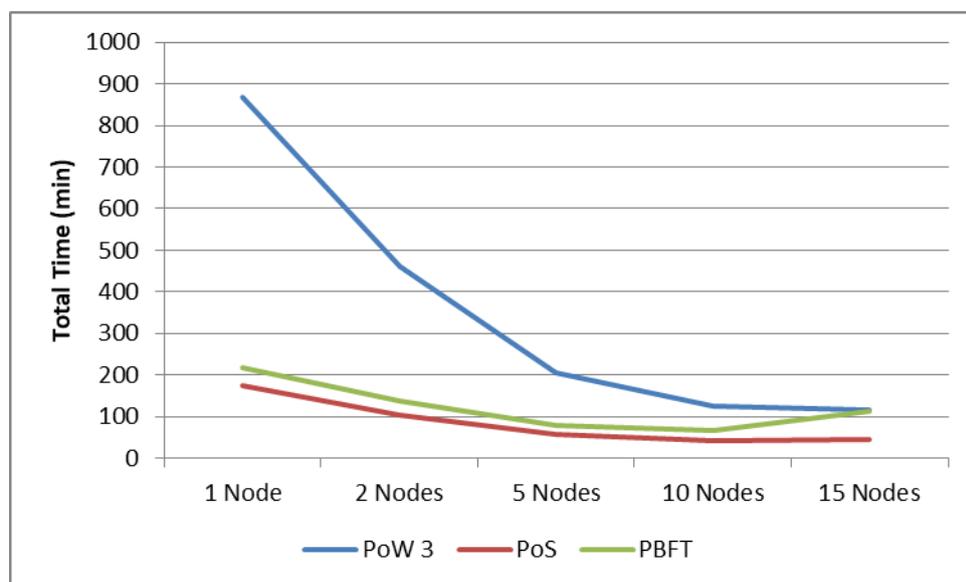


Figura 13. Aplicação da métrica M04.

Do ponto de vista de desempenho pode-se notar que o protocolo PoW é o menos performático, pois necessita de poder computacional para achar uma Hash específica com força bruta, porém é o mais escalável de todos já que a performance isolada de cada nodo não é influenciada pelos demais, apenas pelo seu próprio processamento, já para os protocolos PoS e PBFT a performance vai variar de acordo com a arquitetura modelada, já que a validação dos blocos é coletiva, depende da conexão com outros nodos, entretanto vale ressaltar que dependendo da disposição dos nodos, o nível de segurança varia, pois em um nível com apenas 2 nodos, por exemplo, caso um dos nodos se torne malicioso 50% do nível da rede está infectada, e como os outros níveis apenas “aceitam” os blocos deste nível infectado toda a rede pode ficar comprometida.

4 | CONCLUSÃO

Esse artigo teve por objetivo apresentar o Demochain, um framework para a criação de redes blockchain híbridas para dispositivos IoT. O trabalho realizado até o momento demonstra que o framework proposto é capaz de auxiliar no desenvolvimento de redes blockchain híbridas para ambientes IoT. O estudo realizado e o framework desenvolvido propiciaram a abstração de diversos conhecimentos, possibilitando assim que trabalhos relacionados possam ser desenvolvidos utilizando este como base em diversos aspectos.

Assim, pode-se concluir que não existe uma solução única se tratando de redes blockchain para dispositivos IoT. No entanto, a adoção de um framework para auxílio no desenvolvimento abre uma ampla possibilidade de novas soluções cada vez mais performáticas em diferentes contextos.

Como trabalhos futuros pode ser apontado a implementação de novos protocolos de consenso nesta arquitetura híbrida, com a possibilidade de cada camada estar rodando o seu próprio consenso, além de novos controles de armazenamento e envio da blockchain, e também utilizar outras métricas de desempenho e segurança para testes com este framework, em diferentes ambientes IoT.

REFÊRENCIAS

Bitfury Group (2015). “Proof of Stake versus Proof of Work. White Paper”. [LINK DE ACESSO](#). Último Acesso: 29/11/2018.

Castro, M. e Liskov, B. (1999). “Practical Byzantine fault tolerance”.

Conoscenti, M., Vetrò, A. e De Martin, J. C. (2016). “Blockchain for the Internet of Things: A systematic literature review”.

De Angelis, Stefano. (2017). “Assessing Security and Performances of Consensus algorithms for Permissioned Blockchains”.

Fernández Caramés, M., Fraga Lamas, D. P. (2017). “A Review on the Use of Blockchain for the Internet of Things”.

Freitas, L. (2018). Github com códigos do Demochain. Disponível em <https://github.com/LorenzoWF/Demochain>. Último acesso em 29/11/2018.

Gartner (2017). “Leading the IoT – Gartner Insights on how to lead in a connected world”.

Google (2009). Golang site oficial. Disponível em <https://golang.org>. Último Acesso: 29/11/2018.

Larimer, Dan (2014). “DPOS Description on Bitshares”.

Li, C. e Zhang, L. J. (2017). “A blockchain based new secure multi-layer network model for Internet of Things”.

Liang, X., Shetty, S., Tosh, D., Kamhoua, C., Kwiat, K. e Njilla, L. (2017). “Provchain: A blockchain-based data provenance architecture in cloud environment with enhanced privacy and availability”.

Ortiz, J. e Gottschlich, N. (2016). Base de dados “Household Power Consumption”. Disponível em <https://data.world/databeats/household-power-consumption>. Último Acesso: 29/11/2018.

Schmager, Frank (2010). “Evaluating the GO Programming Language with Design Patterns”.

Vukolić, M. (2018). “The Quest for Scalable Blockchain Fabric: Proof-of-Work vs. BFT Replication”.

Zheng, Z., Xie, S., Dai, H. e Wang, H. (2017). “An overview of blockchain technology: Architecture, consensus, and future trends”.

CONSUMO DO TRÁFEGO DE DADOS EM APLICAÇÕES DE VÍDEO SOB DEMANDA- YOUTUBE E NETFLIX

Patricia Emilly Nóbrega da Silva

IFPB campus Patos, Coordenação do curso técnico integrado de Manutenção e Suporte em Informática

Campina Grande - Paraíba

Éwerton Rômulo Silva Castro

IFPB campus Patos, Diretoria de Ensino

Campina Grande - Paraíba

RESUMO: O rápido crescimento da Internet e a dificuldade das empresas em oferecer serviços de qualidade, fez com que houvesse o interesse na cobrança por tráfego de dados nas conexões de rede fixas. Com isso, a medição do consumo de dados ficou mais evidente neste tipo de conexão. Ao analisar a utilização dos streamings de vídeo, YouTube e Netflix, busca-se esclarecer as dúvidas sobre o uso dessas conexões e seu consumo de dados. Nos resultados obtidos nesse trabalho, percebe-se o dispêndio desses serviços e compara-se com o tráfego dos jogos, apontado como aplicação de maior consumo.

PALAVRAS-CHAVE: Internet, monitoramento, consumo de dados.

CONSUMPTION OF DATA TRAFFIC IN VIDEO APPS UNDER DEMAND - YOUTUBE AND NETFLIX

ABSTRACT: Internet's fast growing rate and companies difficulties in offering quality services made the interest in charging for data traffic on broadband connections. Because of this, the measurement of the data consumption became more evident in this kind of connection. When analyzing the use of streaming services, YouTube and Netflix, the doubts about the use of these kinds of connections and consumer data are sought to be clarified. In the results obtained in this study, it is perceived the cost of these services and it compares to game traffic, pointed as the application with the biggest consumption.

KEYWORDS: Internet, monitoring, data consumption.

1 | INTRODUÇÃO

Nos últimos anos, a internet foi a rede de comunicação que mais cresceu no mundo. Apontando para o Brasil, o consumo dessa rede vem se popularizando dia após dia [1]. Atualmente o acesso à internet é obtido a partir de pacotes, em grande parte, com planos de dados ilimitados, disponibilizados por empresas de telecomunicações privadas. No entanto, o

expresidente da ANATEL (Agência Nacional de Telecomunicações), João Batista de Rezende, julgou como injusto a venda de pacotes de dados de internet ilimitados e pediu o reajuste dos mesmos que passariam a ser limitados com velocidade (taxa de transmissão) nula ou reduzida após o esgotamento do pacote. Semelhante ao acesso via rede celular. Isso viria acontecer devido ao rápido crescimento do uso de dados em plataformas como YouTube, Netflix e principalmente, jogos online, onde a rede não suportaria tal uso sem limitações [2]. Diante disso, surgiram indagações acerca desse uso, principalmente em relação aos jogos (gamers), que foram considerados culpados pelo consumo exacerbado. Para respondê-las serão medidos o consumo de dados usados nos streamings de vídeo, especificamente o YouTube e Netflix, e comparados com alguns valores que foram medidos pelo UOL Jogos [3], a fim de apresentar uma resposta clara e sensata não só para os internautas, que têm alto interesse no assunto, mas para a população em geral.

2 | METODOLOGIA

Para chegar ao objetivo desse estudo, foi feita uma análise quantitativa dos dados obtidos por meio do monitoramento dos streamings de vídeo YouTube e Netflix, utilizando o software GlassWire [4], que foi escolhido por apresentar uma interface de fácil acesso, podendo ser utilizado não só por indivíduos com conhecimentos técnicos, mas por qualquer um que não tenha conhecimento na área.

Os dados foram obtidos da seguinte maneira: o computador utilizado foi conectado à internet e abriu-se no navegador Google Chrome a plataforma de vídeos do YouTube, logo após o software utilizado para as medições foi colocado em funcionamento simultaneamente e os dados coletados foram inseridos em tabelas. Para as comparações dos dados foram avaliados dois vídeos de mesma duração e resolução no YouTube e na Netflix para coletar os dados de monitoramento de ambos e analisar quais os gastos do tráfego. Após o processo descrito acima, foi construída uma tabela de comparação. Os dados obtidos a partir desse trabalho foram para uso exclusivo de pesquisa.

3 | MEDIÇÕES E COLETA DE DADOS

Os dados apresentados nesta sessão foram coletados por meio do monitoramento com o auxílio do software GlassWire [4] nos streamings de vídeo YouTube e Netflix. A partir do monitoramento feito no YouTube, utilizando as resoluções disponibilizadas pela plataforma na reprodução de um vídeo de aproximadamente seis minutos de duração, foram coletados os seguintes dados (Tabela 1):

Resoluções	Download	Upload
144p	11,6 MB	127,8 KB
240p	16,4 MB	208,6 KB
360p	17,8 MB	500 KB
480p	26,7 MB	600 KB
720p (HD)	41,1 MB	831,3 KB

TABELA I – CONSUMO DO TRÁFEGO DE DADOS DO YOUTUBE

No monitoramento da Netflix foram seguidos os mesmos passos do monitoramento do YouTube, o vídeo foi reproduzido em diferentes resoluções (alta, média e baixa) com duração de aproximadamente uma hora e trinta minutos. É importante destacar que diferente do YouTube, a Netflix é um serviço de streaming de transmissão ao vivo e as resoluções variam de acordo com a velocidade da internet do usuário. A seguir serão apresentados os dados (Tabela 2):

Resoluções	Download	Upload
Baixa	317,8 MB	2,8 MB
Média	495,9 MB	3,2 MB
Alta	875,3 MB	3,4 MB

TABELA II – CONSUMO DO TRÁFEGO DE DADOS DA NETFLIX

Analisando as tabelas anteriores (Tabela 1 e Tabela 2) é possível perceber um aumento gradual no consumo de dados de acordo com o aumento da qualidade da resolução. A Tabela 3 foi retirada de uma reportagem feita pelo site UOL Jogos [3] para comparar o consumo de dados em alguns jogos online.

Jogo Online	Download
League Of Legends	30 a 50 MB
Counter-Strike: Global Offensive	100 a 130 MB
Destiny	150 a 300 MB
Diablo III: Reaper of Souls	8 MB
Dota2	50 MB
Fifa 16	20 MB
Guild Wars 2	20 a 30 MB
HearthStone	1 a 2 MB
Heroes of the Storm	50 MB
Overwatch (beta)	52 MB
Rainbow Six: Siege	34,5 MB
Smite	45 a 50 MB
Street Fighter V	21,4 MB

The Division	80 a 100 MB
World of Tanks	6 MB
World of Warcraft	30 a 40 MB

TABELA III – CONSUMO DO TRÁFEGO DE DADOS DE JOGOS ONLINES

4 | COMPARAÇÕES E RESULTADOS

Para este tipo de comparação (Tabela 4) foram utilizados os dados apresentados na sessão anterior do YouTube e Netflix, nas resoluções 480p e Baixa, respectivamente, medidos pela própria autora (Tabela 1 e Tabela 2) e os dados obtidos na página da UOL Jogos [3] (Tabela 3). Os dados dos jogos online foram medidos na duração de uma hora e os dados do YouTube foram medidos no tempo de seis minutos, por tanto é necessário multiplicar os dados do YouTube por 10 para chegar na duração equivalente.

Aplicação	Nome	Download
YouTube	Filme	267 MB
Netflix	Filme	317,8 MB
Jogo Online	League of Legends	30 a 50 MB
Jogo Online	Counter-Strike: Global Offensive	100 a 130 MB

TABELA IV – COMPARAÇÃO DOS DADOS

Nas análises feitas conclui-se que há uma diferença considerável do consumo de dados entre as plataformas, sendo assim, a escolha entre as mesmas para usuários de pacotes de internet limitados deve ser devidamente analisada para evitar o rápido esgotamento do plano. Também pode ser observado que ambas as plataformas têm maior consumo de dados que os Jogos Online. Com isso é possível retornar às palavras do expresidente da ANATEL, citadas na introdução deste trabalho, onde ele afirma que, a rede não suportaria o tal uso indiscriminado e refutá-las, pois, de acordo com os dados obtidos, enquanto alguns usuários assistem vídeos em um dos streamings apresentados e outros jogam online acontece na rede uma compensação onde um usa mais que o outro.

5 | CONCLUSÃO

Após a pesquisa, obtenção, análises e comparações dos dados, concluiu-se que as plataformas de vídeo analisadas tiveram diferentes desempenhos, no qual a Netflix teve o maior uso de dados quando em comparação com o YouTube e os Jogos Online. Quando comparados no tempo de uma hora, os streamings de vídeo consomem mais dados que jogos online. Ao contrário, do que foi divulgado na imprensa pela Anatel, o maior consumo de dados se concentra na utilização do Netflix. Além disso, o

crescimento desse serviço (vídeo sob demanda) evidencia ainda mais a tendência de crescimento do consumo de dados.

REFERÊNCIAS

GLASSWIRE. 1.2.79. Disponível em: <<https://www.glasswire.com>>

O ESTADÃO. **Presidente da Anatel diz que era da internet ilimitada chegou ao fim.** Brasil. 2016. Disponível em: <http://link.estadao.com.br/noticias/empresas,presidente-da-anatel-diz-que-era-da-internet-ilimitada-chegou-ao-fim,10000047861>

SECRETARIA DE COMUNICAÇÃO SOCIAL DA PRESIDÊNCIA DA REPÚBLICA. **Hábitos de consumo de mídia pela população brasileira.** Brasil. 2014. Disponível em: http://dotgroup.com.br/wp-content/uploads/2014/03/Pesquisa_BrasileiradeMidia20141.pdf

UOL. **Game “gasta” internet? Veja quanto consomem os principais jogos.** Brasil. 2016. Disponível em: <<https://jogos.uol.com.br/ultimas-noticias/2016/04/27/game-gasta-internet-confira-o-consumo-de-banda-dos-principais-jogos.htm>>

COMPUTAÇÃO NA ESCOLA: ABORDAGEM DESPLUGADA NA EDUCAÇÃO BÁSICA

Christian Puhmann Brackmann

Instituto Federal de Educação, Ciência e
Tecnologia Farroupilha (IFFAR)
Santa Maria, RS, Brasil

Marcos Román-González

Universidad Nacional de Educación a Distancia
(UNED)
Madrid, CM, Espanha

Rafael Marimon Boucinha

Universidade Federal do Rio Grande do Sul
(UFRGS)
Porto Alegre, RS, Brasil

Dante Augusto Couto Barone

Universidade Federal do Rio Grande do Sul
(UFRGS)
Porto Alegre, RS, Brasil

Ana Casali

Universidad Nacional de Rosario (UNR)
Rosario, SF, Argentina

Flávia Pereira da Silva

Universidade Federal do Rio Grande do Sul
(UFRGS)
Porto Alegre, RS, Brasil

RESUMO: A Computação na Escola, também conhecida como Pensamento Computacional (PC), vem gerando um novo foco educacional nas escolas mundiais como um conjunto de conhecimentos e habilidades para a solução de problemas. Entretanto, não há um consenso

de metodologia de ensino e disponibilidade de material para atender às expectativas dos professores. Para verificar a eficácia de aulas de Pensamento Computacional Desplugado, foi realizada uma avaliação de estudantes da educação primária espanhola com uma abordagem Quase-Experimental no intuito de ensinar os conceitos fundamentais da Computação às crianças em regiões/escolas onde não há dispositivos eletrônicos, Internet e até mesmo energia elétrica. Os resultados apresentaram relevância estatística comprovando melhoria significativa no desempenho dos estudantes que tiveram atividades de PC Desplugado.

PALAVRAS-CHAVE: Pensamento Computacional, Desplugado, Avaliação, Computação na Escola

COMPUTING AT SCHOOL: AN UNPLUGGED APPROACH FOR PRIMARY AND SECONDARY EDUCATION

ABSTRACT: Computing at School, also known as Computational Thinking (CT), has been placing the focus of educational innovation as a set of troubleshooting skills. Unfortunately, there is not a consensus if the teaching methodology and the available materials attend the expectations of the lecturers. To prove the impact that CT training has in primary school,

we attempted to evaluate Spanish students with a Quasi-Experimental approach to prove if the activities are effective for students who live in areas where there are no electronic devices, Internet or even electrical power can be also benefited. The results show statistically significant improvement. In this way, we are able to reinforce the claim that CS unplugged is an effective approach for students who live in unprivileged areas. **KEYWORDS:** Computational Thinking, Unplugged, Evaluation, Computers in Education

1 | INTRODUÇÃO

Nos tempos atuais, marcados pela fluidez da informação e pela valorização do conhecimento, tornou-se imprescindível dominar uma série de ferramentas e recursos tecnológicos que devem ser acessíveis a todos, sem distinção de qualquer natureza para lidar com a informação, processá-la e transformá-la em aptidões para a vida. Aos atuais usuários, é imposto o desafio de criar seus próprios sistemas (por exemplo, programas e jogos) ou modificar os existentes de acordo com sua necessidade pessoal. É neste contexto que surge a habilidade considerada como crucial no século XXI: o Pensamento Computacional (PC) (KOLOGESKI *et al.*, 2016). Devido a essa tendência, o PC tem sido adotado nas escolas da educação básica em diversos países (BRACKMANN *et al.*, 2016, p. 77).

Wing (2006) já define o PC como uma atividade mental para a formulação de um problema passível de ser resolvido computacionalmente, ou seja, trata-se de processos de pensamentos envolvidos na identificação de um problema e na expressão de sua solução de forma eficaz, de modo que tanto máquinas, quanto as pessoas possam executá-los. O PC utiliza quatro *dimensões* ou *pilares* para atingir seu principal objetivo: solução de problemas. Pesquisas lideradas pela Code.Org (2015), Liukas (2015) e BBC Learning (2015) mesclaram os elementos citados por Grover e Pea (2013) e resumiram nos chamados “Quatro Pilares do Pensamento Computacional”: *Decomposição, Reconhecimento de Padrões, Abstração e Algoritmos*. Tais pilares possuem relevada importância e são interdependentes durante o processo de formulação de soluções computacionalmente viáveis.

O Pensamento Computacional envolve identificar um problema complexo e dividi-lo em pedaços menores e mais fáceis de gerenciar (Decomposição). Cada um desses problemas menores poderá ser analisado individualmente com maior profundidade, de forma a identificar problemas parecidos já solucionados anteriormente (Reconhecimento de padrões), focando nos detalhes importantes e ignorando informações irrelevantes (Abstração). Por último, passos ou regras simples podem ser criados para resolver cada um dos subproblemas encontrados (Algoritmos). Ao propor regras ou passos utilizados para criar um código, o resultado se torna compreensível para uso em sistemas computacionais e, conseqüentemente, utilizado na resolução de problemas complexos de forma eficiente, independentemente da carreira profissional que o estudante deseja seguir.

Devido a diferentes razões, existem diversos esforços em tornar o Pensamento Computacional como uma habilidade básica para qualquer pessoa. Entende-se que no momento que temos conhecimento dos efeitos esperados com o desenvolvimento do Pensamento Computacional, pode-se definir o que é necessário para atingir os objetivos pretendidos, bem como, os métodos a serem adotados para alcançar as metas. Destacam-se no Quadro 1, os principais argumentos a respeito dos benefícios que o desenvolvimento do Pensamento Computacional pode trazer dividido em diferentes dimensões.

DIMENSÃO	DESCRIÇÃO
Empregos	Vivemos em tempos onde a criatividade do homem faz a diferença, pois a nova economia mundial não se baseia apenas em recursos naturais e matérias-primas, mas em conhecimento, fluxos de informação e habilidades em usá-los
Compreender o Mundo	As escolas ensinam química com a finalidade de que os estudantes consigam compreender o mundo em que vivem, ou seja, com uma diversidade de reações químicas em sua volta. Os estudantes também aprendem biologia, pois vivemos num mundo vivo. Da mesma forma, vive-se neste século um mundo computacional e a realidade da Computação irá impactar fortemente na rotina dos estudantes.
Transversalidade em diferentes áreas	A Ciência da Computação oferece não apenas software úteis e artefatos de hardware, mas também uma maneira diferenciada de pensar, e que todos, independentemente da área, podem se beneficiar ao pensar computacionalmente e descobrir novas ciências através da análise de uma quantidade gigantesca de dados ou fazer questionamentos que nunca foram cogitados ou ousados devido a sua escalabilidade, facilmente atendida pela Computação.
Alfabetização Digital	Com o uso cada vez mais massivo de computadores na sociedade, o uso de dispositivos digitais será mais uma forma de criar, se expressar e outra forma de apropriação de novos conhecimentos dentro de sua área profissional, pois muitas delas utilizam a Computação para realizar modelagens e simulações para entender melhor o mundo e de uma maneira diferente.
Produtividade	De acordo com McLuhan (1964), o homem deve utilizar as ferramentas como extensão do seu corpo para aumentar suas capacidades, assim como o martelo a força exercida por nossa mão e as bicicletas e os carros nossas pernas. Esses exemplos demonstram que as ferramentas fazem as mesmas ações que nossos membros exercem, porém de forma otimizada.
Ajuda no aprendizado de outras disciplinas	A literatura, através de pesquisas em educação computacional, sugere que a programação pode constituir um método para ensino de Matemática e Ciências (p. 48). Outra pesquisa (TEW <i>et al.</i> , 2008) afirma que alunos que tinham dificuldades em outras disciplinas, começaram a correlacionar conceitos da Computação com os temas trabalhados nas demais disciplinas e tiveram um rendimento superior, inclusive comparável aos melhores alunos.
Inclusão de Minorias	Uma forma de incentivar estudantes a entrar no mercado de trabalho ocorre através do domínio da Computação, o que gera um impacto significativo em sua carreira. Quando os estudantes têm acesso a Computação desde a Educação Básica, essa aproximação é facilitada.
Diminuição nas limitações físicas	A Computação é capaz de expandir o horizonte dos estudantes pelas diversas técnicas de simulação que permitem desafiar a gravidade ou eliminar riscos, facilitando a capacidade de abstrair, criar e lidar com problemas variados e complexos, de forma crítica.
Trabalhar em equipe	Este benefício não é diretamente relacionado com a Computação, mas de acordo com Wing (2006), esta qualidade é essencial para que projetos tenham um bom resultado. Por ser um exercício de persistência, os estudantes, mesmo desenvolvendo individualmente seus projetos, são estimulados a compartilhar com os demais colegas seus erros e acertos e dessa forma trabalham juntos colaborativamente.

Quadro 1. Benefícios do desenvolvimento do Pensamento Computacional (GUZDIAL, 2016) (BRACKMANN, 2017).

Levando-se em consideração os fundamentos da Computação e seus benefícios, para a presente pesquisa, foram criadas e adaptadas diferentes atividades para professores e alunos de maneira que estes pudessem usar e replicar esse material em suas aulas sem a necessidade de equipamentos eletrônicos, Internet ou eletricidade possibilitando que crianças estudem conceitos da Computação em escolas sem equipamentos apropriados (*e.g.* estragados, ultrapassados ou na ausência destes) ou localizadas em áreas geograficamente distantes (*e.g.* áreas rurais ou florestais). Acredita-se que com o uso dessas atividades desplugadas (sem a necessidade de máquinas) é possível ensinar Pensamento Computacional de maneira mais acessível, ou seja, usando basicamente papel, tesoura, canetas, lápis de colorir, cola e demais materiais escolares de uso comum.

Na literatura especializada, há poucas pesquisas a respeito de aplicação e avaliação de estudantes de maneira desplugada. Buscando preencher tal lacuna, o presente trabalho apresenta uma pesquisa realizada em duas escolas primárias na Espanha, cujo objetivo foi o de verificar se as crianças apresentaram uma alteração no desempenho das habilidades relacionadas ao Pensamento Computacional através de atividades sem computadores. Sendo assim, esta pesquisa tem como objetivo verificar se aulas de Pensamento Computacional Desplugado na educação primária são eficazes através da aplicação de um pré e pós-teste, realizados antes e após as aulas de PC Desplugado.

2 | CONTEXTUALIZAÇÃO

Muitos tópicos importantes da Ciência da Computação (CC) podem ser ensinados sem o uso de computadores. A abordagem desplugada da CC introduz conceitos de hardware e software que impulsionam as tecnologias cotidianas até pessoas não-técnicas. Em vez de participar de uma aula expositiva, as atividades desplugadas ocorrem frequentemente por meio da aprendizagem cinestésica (*e.g.* movimentar-se, usar cartões, cortar, colar, desenhar, pintar, resolver enigmas, etc.) e os estudantes trabalham entre si para aprender conceitos da CC.

Os registros literários acerca do surgimento do Pensamento Computacional Desplugado são difusos, no entanto, sabe-se que a necessidade de abstração para a criação de qualquer software e hardware é parte essencial da Ciência da Computação (CC). É importante constar também que o uso de exemplos físicos e materiais escolares são comuns para simular o comportamento de máquinas até os dias atuais em cursos de graduação. Quando se trata de salas de aulas da educação básica, os primeiros registros são encontrados a partir de 1997 quando Bell *et al.* (1997) lançaram um rascunho de um livro em formato digital denominado “*Computer Science Unplugged... Off-line activities and games for all ages*” para professores de todos os níveis interessados em ministrar aulas com um diferencial a seus alunos com

os seguintes objetivos (TAUB; BEN-ARI; ARMONI, 2009):

- Aumentar o interesse dos estudantes em CC;
- Avaliar se os alunos perceberão a CC como algo mais desafiador, intelectualmente estimulante e cooperativo do que anteriormente;
- Conduzir os alunos para uma melhor compreensão do que é CC e evitar confundi-lo com a programação;
- Promover CC como uma carreira para mulheres.

A ideia foi bem recebida pelos demais professores, bem como pela própria Academia. Devido à qualidade do material publicado, a Association for Computing Machinery (ACM) recomendou que as atividades contidas no livro fizessem parte do currículo proposto pela Computer Science Teachers Association (CSTA) dos Estados Unidos da América (Association for Computing Machinery, 2003). Até a publicação deste artigo, o livro CS Unplugged se encontrava na versão 3.1, podendo ser acessado no site do projeto (BELL, 2015).

Embora o uso de atividades de programação de computadores seja a principal abordagem para ensinar habilidades de PC em escolas, educadores e pesquisadores também estão fazendo uso da abordagem desplugada, como afirmado em uma revisão de literatura sistemática que estudou 125 artigos focados em PC (KALELIOGLU *et al.*, 2016). Conclusões semelhantes são encontradas em um levantamento a respeito da metodologia de ensino da Computação. Nessa pesquisa, 13% dos 357 professores participantes afirmam que eles usam atividades desplugadas em suas aulas de Computação (SENTANCE e CSIZMADIA, 2015). No entanto, enquanto a eficácia da programação em computadores para promover o desenvolvimento de aptidões de PC está sendo amplamente investigada (LYE e KOH, 2014), o mesmo não ocorre integralmente com a abordagem desplugada.

Existem diversos estudos que investigam a eficiência de linguagens de programação (visual e código) com crianças (ROMÁN *et al.*, 2015; 2016; GROVER e BASU, 2017; FRANKLIN *et al.*, 2017), porém carecem de abordagens desplugadas. Outras pesquisas tentaram padronizar a avaliação e o ensino de atividade de PC Desplugado, como por exemplo Taub *et al.* (2012) que levou a abordagem desplugada para a sala de aula, mais especificamente as atividades propostas inicialmente por Bell *et al.* (1997) e após a aplicação, através do uso de questionários e entrevistas, investigou-se os efeitos dessas atividades na visão de alunos do ensino fundamental. Os resultados mostram que “embora a maior parte dos estudantes entenderem o que é a Computação, eles também compreendiam como se o computador fosse a essência da Computação e não como uma ferramenta, contrariamente à intenção das atividades”.

Com objetivos e resultados semelhantes, outro projeto de Computação

Desplugada foi implementado como parte de um programa de divulgação com duração de um ano para estudantes do ensino médio com o objetivo de “estimular a próxima geração de alunos de graduação em pesquisa na área da Computação” (FEASTER *et al.*, 2011). Os resultados mostram que o programa não teve impacto percebido sobre os alunos, nem uma melhor compreensão do conteúdo envolvido ou sobre as suas atitudes em relação à Computação. Resultados contrários, porém, são atingidos com um grupo de pesquisadores que visitou diversas turmas de quarto ano com o objetivo de aumentar o interesse dos estudantes na área da Computação com uso de atividades desplugadas. Os resultados, baseados em pré e pós-teste, mostram uma melhoria na confiança e interesse, tanto na área da Computação, como na de Matemática (LAMBERT e GUIFFRE, 2009).

O uso da abordagem desplugada para treinamento de professores também foi estudado por Curzon *et al.* (2014). A pesquisa se deu através de uma série de oficinas organizadas para explorar a eficácia dos métodos desplugados para introduzir os educadores aos tópicos da Computação. A avaliação, baseada em formulários, sugere que “as atividades desplugadas proporcionam uma sessão inspiradora e divertida para os professores e que eles também acham útil, interessante e melhora sua confiança” (p. 92).

Na mesma linha, Curzon (2013) descreve como as atividades desplugadas incorporadas em histórias podem ser usadas para ensinar PC a professores. Especificamente, o artigo apresenta dois exemplos, o primeiro é baseado no problema de ajudar as pessoas com Síndrome do Encarceramento a se comunicar e a segunda baseada em truques mágicos. Após uma oficina de desenvolvimento profissional de 2 horas para professores, 100% dos participantes afirmaram que o evento lhes deu ideias úteis e viáveis para a sala de aula. Um resultado similar foi relatado por Faber *et al.* (2017), onde, após um total de 14.040 horas-aula de ensino de PC em 26 escolas concluíram que a abordagem desplugada é uma valiosa alternativa em relação à programação em computadores.

Como pode ser percebido, a maioria das pesquisas anteriores se concentra em medir as percepções e o interesse dos participantes pela Computação, mas não há avaliação se os participantes desenvolvem suas habilidades de PC. A revisão da literatura também destaca que, há sim uma necessidade de pesquisas empíricas que forneçam evidências sobre a eficácia das atividades desplugadas no desenvolvimento das habilidades do PC, especialmente quando se trata de escolas primárias.

Já, a solução proposta por Rodriguez (2015), Rodriguez *et al.* (2016) e Rodriguez *et al.* (2017) procura avaliar estudantes com cinco atividades desplugadas utilizando uma adaptação dos níveis de pensamento de Bloom (Taxonomia de Bloom), ou seja, em três níveis (proficiente, parcialmente proficiente e insatisfatório). Ainda, Rodriguez utilizou um teste composto por perguntas abertas que não passou por um processo de validação e o aplicou em turmas de sétimo ano do EF, onde foi possível verificar que alunos aprenderam habilidades do PC com atividades desplugadas.

Em Campos *et al.* (2014) também houve a tentativa de adotar um teste, entretanto sem apresentar resultado satisfatório. Já Pasqueline *et al.* (2012) realizou uma avaliação de acertos de estudantes, sem o uso de pré e pós-teste para verificar alterações no desempenho dos mesmos.

Os trabalhos aqui citados são parte de um grande conjunto de pesquisas que tentam medir as habilidades do PC, porém não utilizam uma solução direta, de fácil aplicação e com um processo de validação formal para que se possa atingir um resultado mais preciso e confiável. Diferentemente da proposta de Román-González *et al.* (2015), o teste de PC passou por um rigoroso processo de validação e que culminou em uma tese de doutorado (ROMÁN-GONZÁLEZ, 2016).

A revisão de estudos que fornecem evidências sobre a utilidade da computação desplugada para desenvolver habilidades do PC mostra que ainda há necessidade de novas pesquisas empíricas sobre esse assunto, especialmente quando se trata de seu uso em escolas primárias. Consequentemente, com a investigação relatada neste artigo, tenta-se esclarecer um pouco mais a respeito desse assunto.

3 | MATERIAIS E MÉTODOS

Nesta seção apresenta-se o material desenvolvido e adaptado, o perfil dos participantes envolvidos e o processo de execução da pesquisa.

3.1 Participantes e Grupos de Teste

Os testes e as aulas de PC foram aplicados em um total de 72 alunos dos quintos e sextos anos do ensino fundamental, meninos e meninas da rede pública de ensino na cidade de Madrid, Espanha. Cada uma das escolas que participaram da pesquisa possuía duas turmas do mesmo ano (pré-requisito para aplicação das tarefas), sendo as turmas experimentais e de controle escolhidas aleatoriamente. Nenhum dos participantes possuía experiência formal de programação. A distribuição dos participantes por gênero, nível, idade e grupo são apresentados na Tabela 1.

	Ano	Idade	Grupo	Sexo		Subtotal
				Masculino	Feminino	
Escola A	5	10-11 anos	Controle	10	13	23
			Experimental	11	8	19
Escola B	6	11-12 anos	Controle	6	8	14
			Experimental	9	7	16
Total				36	36	72

Tabela 1. Distribuição dos sujeitos participantes da pesquisa

3.2 Instrumento Avaliativo: Teste de Pens. Computacional

O teste utilizado na pesquisa para medir o desenvolvimento das habilidades que compõem o Pensamento Computacional foi desenvolvido pelos pesquisadores espanhóis Román-González *et al.* (2015). Este teste busca identificar a habilidade de formação e solução de problemas com base nos conceitos fundamentais da Computação, além de utilizar sintaxes lógicas usadas nas linguagens de programação. É o único teste encontrado na literatura com essas características. As questões que compõem o instrumento avaliativo incluem conceitos dos quatro pilares do Pensamento Computacional: abstração, decomposição, reconhecimento de padrões e algoritmos. O Teste de Pensamento Computacional utilizado na pesquisa foi escolhido devido à sua abordagem quantitativa e porque já sofreu um rigoroso processo de validação, que comprovou a validade do conteúdo (ROMÁN-GONZÁLEZ, 2015), a validade dos critérios (ROMÁN-GONZÁLEZ *et al.*, 2017) e a validade convergente (ROMÁN-GONZÁLES *et al.*, 2017). De um modo geral, os estudos psicométricos do teste mostram ser confiáveis ($\alpha^1 \approx .80$) e compatíveis para avaliação do nível de PC em estudantes de 10 a 16 anos de idade.

O instrumento é composto por 28 questões de múltipla escolha, sendo que cada questão possui quatro alternativas de resposta e somente uma é válida. O teste se divide em, basicamente, três partes, a primeira utiliza flechas para deslocar o personagem, a segunda parte faz uma movimentação relativa à posição e direção do personagem utilizando blocos e a última utiliza um lápis para fazer desenhos também com o uso de blocos. Um exemplo de questão é apresentado na Figura 1. Sua aplicação ocorre em navegadores (*e.g.* Chrome, Firefox, Edge) e pode ser acessado de qualquer dispositivo. Nesta pesquisa foram utilizados exclusivamente os computadores do laboratório de informática das escolas.

3.3 Desenvolvimento das Atividades

A maior parte dos materiais desenvolvidos foram criados e os demais foram traduzidos para o Espanhol e adaptados de atividades pré-existentes, como por exemplo o livro “Hello Rubby” (LIUKAS, 2015) e o jogo de tabuleiro “Code Master”, criado por Engelberg *et al.* (2015). Na tentativa de atrair a atenção das crianças, foram também utilizados personagens populares brasileiros nas atividades. Como as atividades utilizadas por esta pesquisa adotam personagens que estão sob direitos autorais criados pelo desenhista Maurício de Souza, o autor obteve autorização de uso de imagem da Turma da Mônica do Maurício de Souza Produções nas atividades. Esses personagens foram utilizados com a finalidade de divulgar a cultura brasileira no exterior. Para que o leitor se familiarize com as atividades trabalhadas, são apresentadas algumas delas no Quadro 2, porém também estão disponíveis no site Pensamento Computacional Brasil². É importante citar que as atividades têm um

1 Alpha de Cronbach

2 Site Pensamento Computacional Brasil: www.computacional.com.br

objetivo principal de auxiliar no processo de ensino-aprendizagem dos conceitos da Computação para crianças da educação básica. Sua aplicação é melhor descrita na seção seguinte.

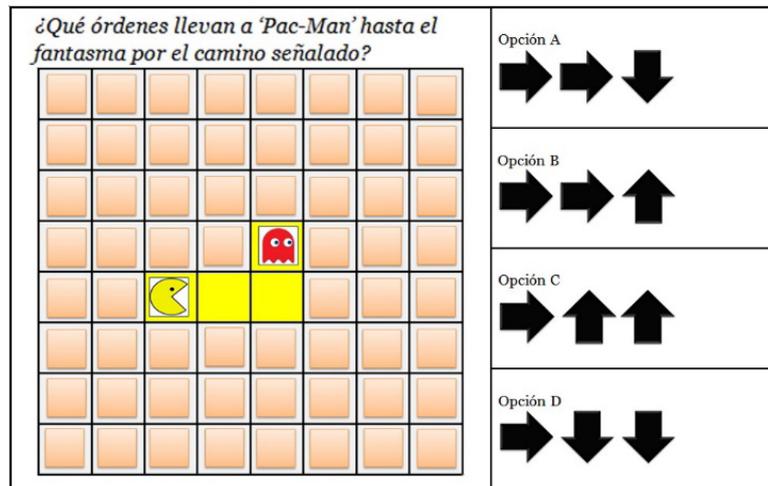
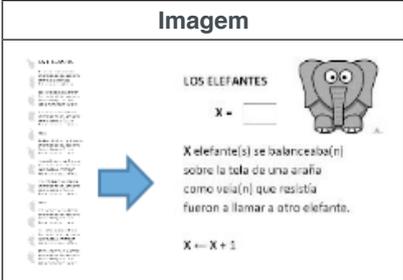
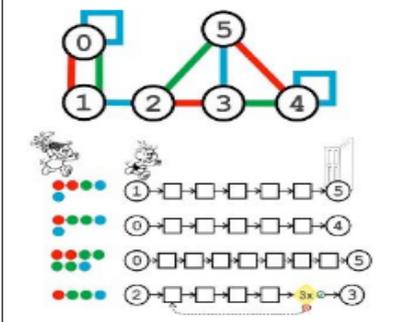


Figura 1. Exemplo de questão do Teste de Pensamento Computacional: “Qual sequência leva o ‘Pac-Man’ até o fantasma pelo caminho indicado?”

Imagem	Descrição da Atividade	Principais pilares envolvidos
	<p>Decomposição: estudantes deveriam especificar, por exemplo, os passos necessários para plantar uma árvore. A atividade é composta de seis situações.</p>	<p>Decomposição Algoritmos</p>
	<p>Mapa da Mônica: foi apresentado mapa com diversos personagens aos estudantes, que deveriam encontrar o caminho mais curto entre eles, usando setas (↑, ↓, ← e →). No segundo momento, deveriam encontrar os padrões e utilizar multiplicadores (e.g. ↑↑↑↑↑ = 5x ↑).</p>	<p>Reconhecimento de Padrão Algoritmos.</p>
	<p>Tetris1 (Instruções Simples): os estudantes foram dispostos em duplas. A um dos membros de cada dupla, foram apresentados alguns desenhos das peças do Tetris. O estudante que ficou com a parte superior do papel deveria passar instruções ao seu colega. Os comandos são limitados a: iniciar, para cima, para baixo, para esquerda, para a direita, para e multiplicadores (e.g. 3x para baixo). Nenhuma outra palavra poderia ser utilizada.</p>	<p>Abstração Algoritmos</p>
	<p>Tetris2 (Repetição): permite ao estudante compreender o uso de repetições para desenhar figuras. Nesta atividade o estudante deveria usar as instruções baseadas na perspectiva da direção e posição da seta e utilizar o maior número de multiplicadores possível. Diferentemente da atividade Tetris1, sua solução é individual e somente poderia usar os comandos <u>para frente</u>, <u>gire para direita</u> e <u>gire para esquerda</u>.</p>	<p>Decomposição Abstração Reconhecimento de Padrão Algoritmos</p>

Imagem	Descrição da Atividade	Principais pilares envolvidos
 <p>LOS ELEFANTES</p> <p>$X =$ <input type="text"/></p> <p>X elefante(s) se balanceaba(n) sobre la tala de una arafia como usia(n) que resistía fueron a llamar a otro elefante.</p> <p>$X \leftarrow X + 1$</p>	<p>Os Elefantes: a atividade utiliza uma música popular entre as crianças para exemplificar como uma canção pode se tornar um algoritmo. Nesta música em especial são trabalhados os conceitos de repetição, variáveis e condicionais.</p>	<p>Reconhecimento de Padrão Algoritmos</p>
 <p>Rota de fuga do Cebolinha e Cascão: uma versão mais simplificada do jogo de tabuleiro <i>Code Master</i> desenvolvido pela companhia <i>ThinkFun</i> (ENGELBERG <i>et al.</i>, 2015). Nesta atividade o estudante deveria encontrar uma rota entre dois nodos utilizando apenas as cores permitidas para cada caminho. A atividade é composta por quatro mapas e 16 desafios (rotas).</p>	<p>Decomposição Abstração Algoritmos</p>	

Quadro 2. Exemplos de atividades utilizadas nas intervenções

3.4 Procedimento

As escolas foram contatadas para apresentação do projeto com até um mês de antecedência do início dos testes e aulas. Os pesquisadores foram muito bem recebidos pelas instituições e todo o apoio necessário foi dado por elas. Cada escola possuía pelo menos duas turmas de um ano específico (Escola A: duas turmas do quinto ano e Escola B: duas turmas de sexto ano). Uma das turmas selecionadas de maneira aleatória foi considerada uma *Turma Experimental* e a outra *Turma de Controle*. Um resumo das etapas da pesquisa encontra-se na Figura 2. As aulas com atividades de PC Desplugado foram aplicadas na turma experimental após o pré-teste e antes do pós-teste, totalizando 10 horas aula. Para que a Turma de Controle tivesse a mesma oportunidade de se beneficiar das atividades, as aulas foram ministradas após o término do experimento, porém sem a geração de dados estatísticos.

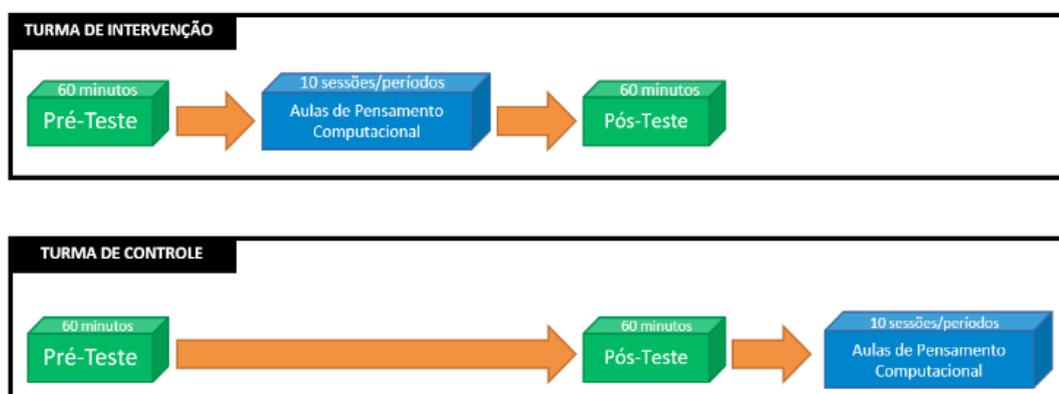


Figura 2. Etapas da pesquisa

Na primeira semana da pesquisa, os alunos das quatro turmas foram convidados a participar do experimento, como parte das suas aulas regulares durante o primeiro

semestre de 2017. Para aplicação dos pré-testes, os alunos foram acompanhados de seus professores ao laboratório de informática da escola, onde permaneceram por até 60 minutos para a realização de um teste individual de Pensamento Computacional desenvolvido por Román-González *et al.* (2015). Durante as cinco semanas seguintes, foram ministradas aproximadamente duas horas semanais de aula de PC Desplugado no grupo experimental. Alguns dos materiais utilizados durante as aulas são apresentados na Seção 3.3. Durante cada uma das sessões semanais, era possível trabalhar em média duas atividades. Na sétima semana, os estudantes de ambos os grupos (experimental e controle) foram novamente levados até o laboratório de informática para nova realização do teste. As respostas de todos os estudantes foram registradas na *Google Cloud* para que, posteriormente, pudessem ser visualizadas, recuperadas e convertidas. As respostas foram então tabuladas e analisadas estatisticamente. Os resultados e discussões sobre os dados coletados encontram-se disponíveis na próxima seção.

4 | RESULTADOS E DISCUSSÕES

Esta seção apresenta os resultados dos Testes de Pensamento Computacional e uma breve discussão dos mesmos. Lembra-se que nenhum dos alunos teve contato com aulas formais de programação. A pontuação dos testes de PC é calculada de acordo com a quantidade de questões respondidas corretamente, lembrando também que o teste é composto de 28 questões (pontuação máxima). O teste foi aplicado tanto no Grupo Experimental como no de Grupo Controle.

Na Tabela 2 são apresentados os resultados das médias e desvios padrão (DP) das duas turmas e na Figura 3 o gráfico compara as pontuações entre o pré e pós-teste.

CONTROLE					
	N	Pré-Teste Média (DP)	Pós-Teste Média (DP)	p-valor	Alteração de Desempenho (%)
Desempenho	37	10,27 ($\pm 3,26$)	10,84 ($\pm 3,62$)	0,267	+0,57 (5,53%)
EXPERIMENTAL					
	N	Pré-Teste Média (DP)	Pós-Teste Média (DP)	p-valor	Alteração de Desempenho (%)
Desempenho	35	11,46 ($\pm 2,98$)	13,83 ($\pm 3,01$)	< 0,001	+2,37 (20,70%)

Tabela 2. Resultados Estatísticos

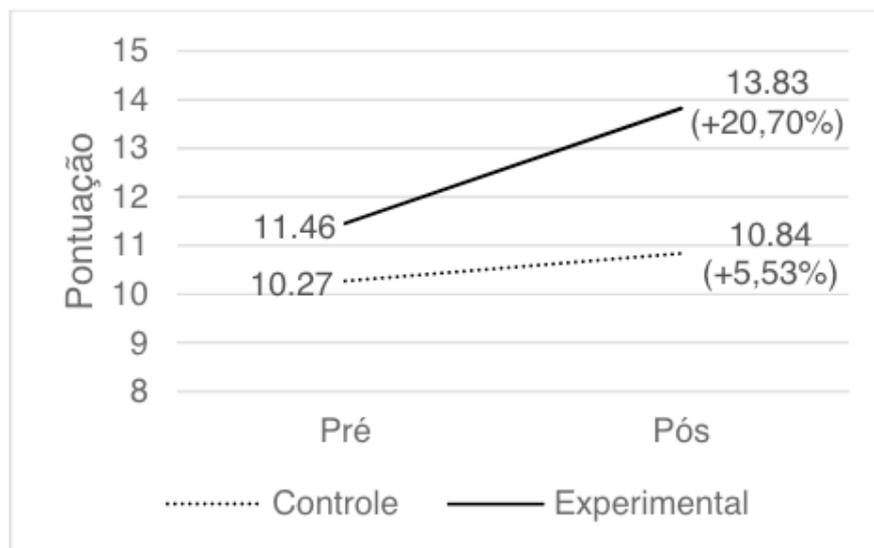


Figura 3. Gráfico comparativo dos resultados

Os resultados obtidos com este instrumento foram submetidos a procedimentos estatísticos a fim de testar a hipótese nula, ou seja, avaliar se houve diferença entre os resultados do pré-teste e pós-teste e se esta diferença era significativa. No primeiro momento, foi realizado um teste da equivalência dos grupos (Controle e Experimental) através do Teste T de Student Independente, onde se certificou que os grupos em questão possuem um mesmo perfil para que possam ser comparados estatisticamente.

Para verificar se a melhora da média é estatisticamente significativa ou ocorreu de forma aleatória, foi utilizado o Teste T para amostras pareadas, considerando o intervalo de confiança de 95%. No grupo onde houve intervenção (experimental), observou-se uma melhoria de 20,70% na quantidade de acertos e a análise estatística demonstrou um P ($T \leq t$) bicaudal $< 0,001$. Considerando que este valor é menor que 0,05, comprovou-se que há diferença significativa entre as médias, do ponto de vista estatístico. O mesmo tratamento de dados foi utilizado no grupo de controle, onde houve uma melhoria de apenas 0,53% de acertos e um P ($T \leq t$) bicaudal = 0,267, ou seja, neste grupo não houve melhora do desempenho. Este dado reforça a teoria que o incremento do Pensamento Computacional no Grupo Experimental foi de fato motivado pelas atividades de PC Desplugado. Sendo assim, os dados apresentados demonstraram que houve melhora no desempenho dos estudantes no grupo experimental e uma estagnação do grupo de controle, comprovando a hipótese inicial da pesquisa.

5 | CONCLUSÕES E TRABALHOS FUTUROS

Este artigo apresentou uma breve introdução do Pensamento Computacional, seu histórico e o estado-da-arte. Em seguida foram explanadas todas as etapas da pesquisa, os grupos participantes, o teste utilizado, os materiais desenvolvidos e os resultados.

Enquanto alguns países possuem em suas escolas praticamente um computador por aluno (*e.g.* Austrália e Nova Zelândia), outros países se destacam pela qualidade no ensino, mesmo compartilhando um único computador com até quatro estudantes (*e.g.* Suécia e Portugal) (OECD, 2015), comprovando que o uso massivo de dispositivos tecnológicos em sala de aula não garante uma melhor educação.

Baseado nesse princípio e na situação socioeconômica brasileira, optou-se pelo formato “desplugado” como uma alternativa para que todo o país, independentemente da situação da escola, possa usufruir dos benefícios que o Pensamento Computacional viabiliza. A presente pesquisa desenvolveu materiais e apontou evidências que o ensino do Pensamento Computacional através de atividades que não necessitam de máquinas é possível. Os resultados sugerem que ambos grupos experimentais tiveram uma melhoria do Pensamento Computacional significativa, independente do país, anos escolares investigados ou gênero. O mesmo não ocorreu no grupo de controle, comprovando o efeito positivo das intervenções nos grupos experimentais. Os dados positivos podem também ser compreendidos como uma variação muito pequena na melhoria de desempenho, porém é essencial considerar que ocorreram apenas 10 horas-aula de atividades desplugadas, porém pôde-se perceber grande entusiasmo e motivação das crianças durante as aulas de PC. Os professores também demonstraram muita satisfação com a oportunidade dada a seus alunos. De uma maneira geral, as escolas foram notavelmente receptivas com relação às aulas de PC, não criando qualquer barreira para o projeto.

Porém, considera-se importante salientar que atividades desplugadas não devem ser entendidas como uma solução completa de ensino. A abordagem desplugada não atende todos os fundamentos da Computação ou não proporciona uma prática plena. Por esse motivo recomenda-se que seja feito uso dela, sim, para a introdução do Pensamento Computacional pelo fato de “o trabalho com tecnologias digitais ter algo especial que não pode ser explorado por outros tipos de atividades” (NATIONAL US RESEARCH COUNCIL e COMMITTEE FOR THE WORKSHOP ON COMPUTATIONAL THINKING, 2010). A mesma recomendação pode ser encontrada em Grover e Pea (2013) ou em Aggarwal *et al.* (2017), em que foi realizado uma comparação entre o ensino do Pensamento Computacional com e sem o uso de computadores para a criação de um jogo. As atividades desplugadas, a partir de um certo ponto, podem não ser tão eficazes quanto esperado, por manter os alunos distantes de experiências com as tecnologias digitais. Isso poderá acarretar uma visão distorcida, como por exemplo, do que é a Computação, ou até mesmo distanciá-los dela, tendo em vista que programar no computador é uma experiência fundamental e praticamente única. O acesso às máquinas pelas crianças possibilita pôr em prática aquilo que foi trabalhado no formato desplugado, ampliando ainda mais o seu horizonte, a intimidade com uma linguagem de programação e possibilita solucionar problemas ainda mais complexos.

Atualmente percebe-se um grande interesse e apoio de diversas instituições (públicas e privadas) para a adoção do PC na Educação Básica, porém é essencial

divulgar sua importância, bem como formar novos professores ou possibilitar que os atuais se apropriem adequadamente da técnica, incentivando-os a adotar seu uso em sala de aula. Como trabalhos futuros, pretende-se aplicar as mesmas atividades e testes em escolas brasileiras.

6 | AGRADECIMENTOS

Esta pesquisa recebeu apoio financeiro da Agência Executiva de Educação, Audiovisual e Cultura da Comissão Europeia (EACEA/ Programa SMART²) e Ministério da Educação (MEC) no Brasil. Os pesquisadores agradecem os colégios CEIP República de Ecuador e CEIP Lope de Veja, ambos de Madri/Espanha, por permitir a execução da investigação em suas instituições. Agradecemos também aos Estúdios Maurício de Souza S. A. e Mark Engelman (Thinkfun, Inc.) em permitir formalmente o uso de suas criações nas atividades.

REFERÊNCIAS

AGGARWAL, A.; GARDNER-MCCUNE, C.; TOURETZKY, D. S. **Evaluating the Effect of Using Physical Manipulatives to Foster Computational Thinking in Elementary School**. ACM Press, 2017. Disponível em: <<http://dl.acm.org/citation.cfm?doid=3017680.3017791>>. Acesso em: 12 set. 2017

ASSOCIATION FOR COMPUTING MACHINERY. **A Model Curriculum for K-12 Computer Science: Final Report of the ACM K-12 Task Force Curriculum Committee**. New York: ACM, 2003.

BBC LEARNING, B. What is computational thinking? 2015.

BELL, T.; WITTEN, I. H.; FELLOWS, M. **Computer Science Unplugged... - Off-line activities and games for all ages (draft)**. 1. ed. [s.l.: s.n.].

BELL, T.; WITTEN, I. H.; FELLOWS, M. **Computer Science Unplugged**, 2015. Disponível em: <http://csunplugged.org/wp-content/uploads/2015/03/CSUnplugged_OS_2015_v3.1.pdf>. Acesso em: 1 jan. 2015

BRACKMANN, C. *et al.* **Computational thinking: Panorama of the Americas**. IEEE, set. 2016. Disponível em: <<http://ieeexplore.ieee.org/document/7751839/>>. Acesso em: 21 maio. 2017.

BRACKMANN, C. **Desenvolvimento do Pensamento Computacional Através de Atividades Desplugadas na Educação Básica**. 226 p. - Universidade Federal do Rio Grande do Sul (UFRGS), 2017. Disponível em: <<http://hdl.handle.net/10183/172208>>.

CAMPOS, G. *et al.* Organização de Informações via Pensamento Computacional: Relato de Atividade Aplicada no Ensino Fundamental. **20o. Workshop de Informática na Escola (WIE 2014)**, Congresso Brasileiro de Informática na Educação - CBIE 2014. 2014.

CODE.ORG. **Instructor Handbook - Code Studio Lesson Plans for Courses One, Two, and Three**. [s.l.] CODE.ORG, 2015.

CURZON, P. **cs4fn and computational thinking unplugged**. ACM Press, 2013. Disponível em: <<http://dl.acm.org/citation.cfm?doid=2532748.2611263>>. Acesso em: 20 maio. 2017

CURZON, P. *et al.* **Introducing teachers to computational thinking using unplugged storytelling**. In: WIPSC. ACM Press, 2014. Disponível em: <<http://dl.acm.org/citation.cfm?doid=2670757.2670767>>.

Acesso em: 20 maio. 2017

ENGELBERG, M.; THINKFUN. **Code Master Programming Logic Game** Thinkfun, 2015. Disponível em: <<http://www.thinkfun.com/products/code-master/>>

FABER, H. *et al.* Teaching Computational Thinking to Primary School Students via Unplugged Programming Lessons. **Journal of the European Teacher Education Network**, v. 12, 2017.

FEASTER, Y. *et al.* **Teaching CS unplugged in the high school (with limited success)**. ACM Press, 2011. Disponível em: <<http://portal.acm.org/citation.cfm?doid=1999747.1999817>>. Acesso em: 20 maio. 2017

FRANKLIN, D. *et al.* **Using Upper-Elementary Student Performance to Understand Conceptual Sequencing in a Blocks-based Curriculum**. ACM Press, 2017 Disponível em: <<http://dl.acm.org/citation.cfm?doid=3017680.3017760>>. Acesso em: 16 abr. 2017

GROVER, S.; BASU, S. **Measuring Student Learning in Introductory Block-Based Programming: Examining Misconceptions of Loops, Variables, and Boolean Logic**. ACM Press, 2017 Disponível em: <<http://dl.acm.org/citation.cfm?doid=3017680.3017723>>. Acesso em: 12 jul. 2017

GROVER, S.; PEA, R. Computational Thinking in K-12: A Review of the State of the Field. **Educational Researcher**, v. 42, n. 1, p. 38–43, 1 jan. 2013.

GUZDIAL, M. **Learner-Centered Design of Computing Education - Research on Computing for Everyone**. [s.l.] Morgan & Claypool, 2016.

KALELIOGLU, F.; GÜLBAHAR, Y.; KUKUL, V. A Framework for Computational Thinking Based on a Systematic Research Review. **Baltic Journal of Modern Computing**, v. 4, n. 3, p. 583–596, 2016.

KOLOGESKI, A. *et al.* Desenvolvendo o Raciocínio Lógico e o Pensamento Computacional: Experiências no Contexto do Projeto Logicando. **Revista Novas Tecnologias na Educação (RENOTE)**, v. 14, n. 2, 2016.

LAMBERT, L.; GUIFFRE, H. Computer science outreach in an elementary school. **Journal of Computing Sciences in Colleges**, v. 24, n. 3, p. 118–124, jan. 2009.

LIUKAS, L. **Hello Ruby: adventures in coding**. [s.l.] Feiwei & Friends, 2015.

LYE, S. Y.; KOH, J. H. L. Review on teaching and learning of computational thinking through programming: What is next for K-12? **Computers in Human Behavior**, v. 41, p. 51–61, dez. 2014.

MCLUHAN, M. **Understanding media: the extensions of man**. 1st MIT Press ed ed. Cambridge, Mass: MIT Press, 1994.

NATIONAL RESEARCH COUNCIL (U.S.); COMMITTEE FOR THE WORKSHOPS ON COMPUTATIONAL THINKING. **Report of a workshop on the scope and nature of computational thinking**. Washington, D.C.: National Academies Press, 2010.

NISHIDA, T. *et al.* **A CS Unplugged Design Pattern**. New York, NY: ACM, 2009 Disponível em: <<http://dx.doi.org/10.1145/1508865>>. Acesso em: 18 maio. 2017

OECD. **Students Computers and Learning**. [s.l.] OECD, 2015.

RODRIGUEZ, B. **Assessing Computational Thinking in Computer Science Unplugged Activities**. Golden, Colorado, EUA: Colorado School of Mines, 2015. Disponível em: <<https://dspace.library.colostate.edu/handle/11124/169998>>.

RODRIGUEZ, B. *et al.* **Assessing Computational Thinking in CS Unplugged Activities.** Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education. **Anais...** In: TECHNICAL SYMPOSIUM ON COMPUTER SCIENCE EDUCATION. Seattle, Washington, USA: ACM, mar. 2017

RODRIGUEZ, B.; RADER, C.; CAMP, T. **Using Student Performance to Assess CS Unplugged Activities in a Classroom Environment.** ACM Press, 2016Disponível em: <<http://dl.acm.org/citation.cfm?doid=2899415.2899465>>. Acesso em: 6 jul. 2017

ROMÁN-GONZÁLEZ, M. **Computational Thinking Test: Design Guidelines and Content Validation.** Proceedings of the 7th Annual International Conference on Education and New Learning Technologies (EDULEARN 2015). **Anais...** In: EDULEARN15. Barcelona, Spain: IATED, jul. 2015

ROMÁN-GONZÁLEZ, M. **Códigoalfabetización y Pensamiento Computacional en Educación Primaria y Secundaria: Validación de un Instrumento y Evaluación de Programas.** Madrid, Spain: Universidad Nacional de Educación a Distancia, fev. 2016.

ROMÁN-GONZÁLEZ, M.; PÉREZ, J. C.; CARMEN JIMÉNEZ-FERNÁNDEZ. Test de Pensamiento Computacional: diseño y psicometría general. 2015.

SCAICO, P. *et al.* Um Relato de Experiências de Estagiários da Licenciatura em Computação com o Ensino de Computação para Crianças. **Revista Novas Tecnologias na Educação (RENOTE)**, v. 10, n. 3, 2012.

SENTANCE, S.; CSIZMADIA, A. **Teachers' perspectives on successful strategies for teaching Computing in school.** IFIP TCS 2015. **Anais...** In: IFIP TCS 2015. 2015Disponível em: <<http://community.computingatschool.org.uk/files/6769/original.pdf>>

TAUB, R.; ARMONI, M.; BEN-ARI, M. CS Unplugged and Middle-School Students' Views, Attitudes, and Intentions Regarding CS. **ACM Transactions on Computing Education**, v. 12, n. 2, p. 1–29, 1 abr. 2012.

TAUB, R.; BEN-ARI, M.; ARMONI, M. **The effect of CS unplugged on middle-school students' views of CS.** ACM Press, 2009Disponível em: <<http://portal.acm.org/citation.cfm?doid=1562877.1562912>>. Acesso em: 20 maio. 2017

TEW, A. E. *et al.* Context as Support for Learning Computer Organization. **Journal on Educational Resources in Computing**, v. 8, n. 3, p. 1–18, 1 out. 2008.

WING, J. M. Computational thinking. **Communications of the ACM**, v. 49, n. 3, p. 33, 1 mar. 2006.

COLETA DE DADOS E ANÁLISE DE SENTIMENTOS NAS REDE SOCIAIS ON LINE

Maurilio Alves Martins da Costa
Bruna Emidia de Assis Fraga

RESUMO: As redes sociais on line (RSO) são utilizadas por milhões de usuários para compartilharem pensamentos, acontecimentos e opiniões sobre fatos e produtos. Este trabalho se objetiva a utilizar a técnica de análise de sentimentos (AS) para extrair o sentimento contido nas mensagens publicadas pelos usuários de uma rede social. Utilizou-se a abordagem léxica com saída polarizada [Benevenuto, Ribeiro e Araujo 2015] aplicada à RSO Twitter. Os resultados das coletas refletiram de modo satisfatório ao sentimento que correspondiam ao evento estudado, confirmando a polaridade existente bem como os destaques ocorridos nos debates.

ABSTRACT: Online social networks (OSN) are used by millions of persons to share thoughts, events, and opinions about facts and products. This work aims to use the technique of sentiment analysis (SA) to extract the sentiment contained in messages published by social network's users. We used the lexical approach with polarized output [Benevenuto, Ribeiro and Araujo 2015] applied to Twitter. The results reflected satisfactorily the feeling that corresponded to the event studied, confirming

the existing polarity as well as the highlights of the presidential debates..

1 | INTRODUÇÃO

As redes sociais online (RSO) são um tipo específico de mídia social. De acordo com Kaplan e Haenlein (2010) as redes sociais online possuem a funcionalidade de compartilhar conteúdos e têm seu foco voltado para o relacionamento e o engajamento entre os usuários de forma que seja estabelecida uma interação virtual entre eles.

As redes sociais online são em sua maioria gratuitas e alcançam em torno de 3 bilhões de pessoas em todo o mundo. Elas foram desenvolvidas para que as pessoas criem perfis, estabeleçam uma rede de amigos e publiquem conteúdos. Estes conteúdos podem ser de diversos formatos como links, fotos e vídeos mas principalmente mensagens textuais.

O conteúdo produzido por um usuário é conhecido como *user-generated content* (UGC) e dada a riqueza de informações contidas nas publicações geradas sobre variados temas e assuntos Balazs e Velásquez (2016) afirmam que é humanamente impossível compreender toda essa informação em um período de tempo satisfatório, sendo necessários mecanismos para extrair e analisar esses conteúdos,

surgindo, então, o campo de pesquisa da análise de sentimento (AS) ou mineração de opinião (MO).

Segundo Liu (2012), a análise de sentimento permite que se tenha uma percepção geral da opinião dos usuários da rede em um dado momento sobre um assunto, ou ainda a captura da reação do público, seja ela positiva ou negativa, sobre algum tema. De acordo com os estudos realizados por Medhat, Hassan e Korashy (2014) a análise de sentimentos pode ser realizada a partir de duas abordagens principais: (i) uma baseada em aprendizado de máquina e outra baseada em análise léxica. O aprendizado de máquina pode ser definido como um conjunto de algoritmos capazes de reconhecer padrões a partir de um conjunto de dados e fazer uma previsão futura. (ii) A outra abordagem para a análise de sentimentos é a análise léxica, que segundo Benevenuto, Ribeiro e Araújo (2015), relaciona valores quantitativos ou qualitativos às palavras para atribuir-lhes sentimento. Assim, é formado um conjunto de palavras opinativas que compõe um dicionário léxico. Nesta linha de abordagem, podem ser utilizados também padrões sintáticos para reconhecer novas palavras além daquelas que já existem no conjunto [Medhat, Hassan e Korashy 2014]. Segundo esses autores, a abordagem léxica é vantajosa porque não necessita de classificação prévia de sentenças e nem de base de treino como ocorre nos métodos de aprendizagem de máquina, assim, ela fica livre de contexto por justamente não sofrer especialização durante o treinamento

Da análise de sentimentos obtêm-se como resultado uma classificação do sentimento que pode envolver três aspectos: polaridade, emoção e força do sentimento [Benevenuto, Ribeiro e Araújo 2015]. Segundo os autores, a saída comum dos métodos é a polarizada. A polaridade indica uma orientação que pode ser positiva, negativa ou neutra, assim, o texto analisado terá uma saída que pode ser binária (positiva ou negativa) ou terciária (positiva, negativa e neutra).

O objetivo deste estudo é descrever uma metodologia de coleta e extração do sentimento exposto nas mensagens textuais publicadas pelos usuários de uma rede social, utilizando um método de análise de sentimentos com a abordagem de análise léxica com saída polarizada.

2 | ESTADO DA ARTE

A ascensão dos problemas de análise de sentimentos levou à publicação de um grande volume de pesquisas. As áreas de destaque destas publicações foram: a de saúde [Chew e Eysenbach 2010] e [Vydiswaran, Zhai e Roth 2011]; área empresarial, com o monitoramento do sentimento do público em relação às marcas e produtos [Evangelista e Padilha 2013] e [Ghiassi, Skinner e Zimbra 2013], social [Nascimento et al. 2012] e [Malheiros e Lima 2013] e política [Stieglitz e Dang-Xuan, 2012] e [Malini, Ciarelli e Medeiros 2017].

Em Cavalin et al. (2014), foi desenvolvido um sistema de tempo real capaz de

analisar o sentimento das mensagens dos usuários do Twitter. Ao longo da pesquisa foram analisadas as mensagens durante os jogos da Copa das Conferências de 2013. Quanto à análise de sentimentos a abordagem de aprendizado de máquina foi utilizada e o classificador probabilístico de Naive Bayes polarizou as mensagens em sentimentos positivo, negativo e neutro.

Souza (2015) criou um método para identificar usuários influentes da rede social Twitter e classificar automaticamente a categoria na qual o usuário é influente, baseando-se nos conteúdos postados. Os conteúdos poderiam pertencer a três categorias: política, esporte ou outros assuntos. Para a classificação das postagens, utilizou-se técnicas de mineração de texto e o método de aprendizado supervisionado Naive Bayes.

Wang et al. (2012) desenvolveu um sistema em tempo real que monitorou as eleições presidenciais americanas de 2012 no Twitter. Em seu estudo foram considerados quatro tipos de sentimentos: negativo, positivo, neutro e incerto. Para formar as bases de treino e teste do modelo Naive Bayes (NB) unigrama, uma série de mensagens foram classificadas por voluntários.

3 | FUNDAMENTOS TEÓRICOS

3.1 Redes Sociais Online (RSO)

As RSOs podem ser expressas como serviços baseados na web no qual as pessoas podem criar perfis em um sistema, podendo ser eles públicos ou não [Boyd e Ellison 2007]. Ainda de acordo com o autor, através desse serviço os usuários se apresentam, conectam-se uns aos outros visualizando e explorando tanto à sua lista de conexões como as de outras pessoas. O principal objetivo das plataformas sociais é formar e tornar visíveis as conexões entre pessoas.

Apesar de possuírem um objetivo central em comum, cada RSO disponibiliza recursos e funcionalidades de acordo com o seu propósito. Por exemplo, o Facebook¹ é têm como propósito a criação de relacionamentos de amizade. Já o Pinterest² e o Instagram³ incentivam seus usuários à compartilharem imagens. Com foco nos negócios a rede LinkedIn⁴ permite que seus participantes se apresentem de forma profissional revelando aptdões e habilidades, compartilhem conteúdos relacionados à temas profissionais relacionados ao trabalho e favorece o networking. Em uma outra frente, têm-se o Twitter⁵, uma rede que foca em conteúdos e notícias rápidas.

As redes Facebook e Twitter receberão maior destaque ao longo deste estudo por se tratar de redes que possuem características relacionadas à mensagens textuais e público mais coerentes com os objetivos deste estudo.

- 1 <https://www.facebook.com>
- 2 <https://br.pinterest.com>
- 3 <https://www.instagram.com>
- 4 <https://br.linkedin.com>
- 5 <https://twitter.com>

O Facebook estabelece a conexão entre os usuários de forma bidirecional, assim um usuário solicita a amizade e outro deve confirmar. As amizades geralmente já existem no meio offline e posteriormente se tornam virtuais. No Twitter, ocorre a unidirecionalidade da conexão, em que só existe a opção seguir. Dessa forma, os usuários buscam seguir e receber atualizações somente daqueles que produzem conteúdos considerados para eles relevantes. Assim, o Facebook conecta pessoas, enquanto o Twitter conecta tópicos e idéias.

As duas redes possuem em comum recursos como hashtags, menções, tópicos e é possível pesquisar pessoas, negócios, tópicos e empresas. Ainda segundo o autor o Facebook ao disponibilizar mais recursos que o Twitter, se torna mais difícil de ser usado pelos usuários.

3.2 Extração de dados das Redes Sociais Online

As redes sociais em geral, disponibilizam interfaces para a interação ou extração de seus dados. Essas interfaces são conhecidas como Application Programming Interface (API) e para realizar a extração de dados é preciso conhecer suas especificidades.

A principal API da rede social Facebook é a Graph API, disponível em <https://developers.facebook.com/>. Segundo documentação disponível em <https://developers.facebook.com/docs/graph-api/>, todas as outras APIs existentes são expansões desta. A Graph API estabelece uma limitação de volume de requisições que podem ser feitas ao servidor dentro de um intervalo de tempo. Durante uma hora, uma aplicação pode realizar uma quantidade de chamadas igual a duzentas vezes o número de usuários que ela possui. Para realizar solicitações à Graph API são necessários tokens de acesso que fornecem autorização para tal. Uma outra API é a de Feed Público, que necessita de aprovação prévia do Facebook.

O Twitter fornece aos desenvolvedores uma grande variedade de APIs para a construção de aplicações ou soluções para a sua ferramenta. A API Standard, que será utilizada neste trabalho, é uma categoria de API gratuita, já as duas outras categorias, Premium e Enterprise, são pagas. A categoria Standard conta com dois módulos: (i) API REST que permite a realização de consultas no banco de dados do Twitter de dados publicados há no máximo sete dias e; (ii) API Streaming que permite aos desenvolvedores acesso às publicações em tempo real.

3.3 Mineração de Texto

A Mineração de textos (MT) é uma extensão da área de mineração de dados [Hearst 1999]. Esta subárea foi desenvolvida para que a partir de textos não estruturados ou semi-estruturados, fosse possível encontrar padrões relevantes e não triviais e ainda extrair novas informações [Kantardzic 2011]. Nesse processo, conteúdos textuais de diferentes fontes são extraídos automaticamente com o objetivo de revelar uma informação completamente nova.

A mineração de dados e a mineração de textos, se diferenciam na forma de reconhecer os padrões de seus objetos de análise. Para minerar uma forma textual, segundo Weiss et al. (2010), se faz necessária a transformação dos textos em uma representação numérica. Já na mineração de dados, as informações estão estruturadas dentro do banco de dados e são formadas por fatos, já que foram armazenadas automaticamente por sistemas [Hearst 2003].

Este trabalho utilizará a mineração de textos a fim de extrair os sentimentos dos documentos online produzidos por usuários das redes sociais. Para isto, foi escolhido o modelo de mineração textual proposto por [Aranha e Vellasco 2007]. Este modelo é formado por cinco etapas sequenciais. A primeira delas é a formação da base de documentos, cujo objetivo é a coleta de informações para a construção de um arcabouço de dados a serem explorados. Nesse trabalho a ênfase da coleta estará voltada para dados contidos na Internet.

A segunda etapa é o pré-processamento onde os dados são transformados para se adequarem a formatos passíveis de serem interpretados pelos algoritmos de extração automática de conhecimento. Esta etapa reduz o custo da etapa de mineração agindo na solução de problemas como dados ruidosos, faltantes e ou redundantes.

Posteriormente têm-se a fase da indexação que cria coleções de termos extraídos do conjunto de dados proporcionando otimização da velocidade e desempenho na busca realizada [Soares 2013].

A quarta fase é a da mineração que irá trabalhar em cima dos dados provenientes dos processos anteriores para a produção efetiva de novos conhecimentos úteis, aplicando sobre os dados, algoritmos pertencentes à diversas áreas do conhecimento.

A quinta fase é a da análise dos resultados obtidos, nessa fase ocorre a análise e interpretação dos resultados a partir dos dados obtidos, fazendo uma verificação para avaliar o classificador de acordo com algumas métricas que podem ser taxa de erro, tempo de CPU e complexidade do modelo ou ainda as métricas podem ser qualitativas, levando em consideração o conhecimento de especialistas do domínio.

4 | METODOLOGIA

O presente estudo é uma pesquisa caráter exploratório e descritivo. Cujo objetivo é conhecer o objeto de estudo de forma a compreender a sua contextualização e o seu significado. Este tipo de pesquisa ainda pode ser considerada uma “investigação em área onde há pouco conhecimento sistematizado, acumulado.” (Vergara 1990, p.4). Já a pesquisa descritiva, é definida por Vergara (1990) como sendo uma exposição de características de um fenômeno em que fatores e variáveis podem ser correlacionados.

Dentre as diversas redes sociais online disponíveis online, o Twitter foi escolhida como objeto de estudo por ter como principal objetivo a propagação de mensagens textuais em um espaço limitado, o que auxilia no processo de mineração do texto. Além do mais os perfis criados nessa rede social são definidas por padrão como públicas,

assim, todas as informações compartilhadas são visíveis à toda a rede. Por fim, o Twitter oferece uma interface mais acessível que as outras redes sociais online.

Como tema para estudo foi escolhido um assunto que tivesse o potencial de gerar discussões no Twitter, sendo escolhido assim a campanha presidencial de 2018. Para isto, foram escolhidos alguns eventos para o monitoramento das mensagens publicadas nas redes sociais, como mostra a Tabela 1.

Evento	Período da Coleta (2018)
Oficialização do Haddad como candidato	11 de Setembro
Debate no SBT (TV)	26 de Setembro
Debate na Record (TV)	30 de Setembro
Debate na Globo (TV)	04 de Outubro

Tabela 1 – Acontecimentos monitorados durante o período de eleições em 2018

Fonte: Autora

Esta pesquisa utilizou menções e hashtags como alvo dos algoritmos de busca por concentrarem as mensagens do tema em questão que foram: a oficialização da candidatura de Haddad e os debates nas emissoras SBT, Record e Globo. Ao final do período de coleta, a extração obteve como resultado quatro grandes arquivos com um total de 651.837 tweets. Além de informações referentes à própria mensagem como texto, menções e hashtags, é possível obter dados de usuário como: nome, quantidade de seguidores e seguidos, status de perfil, quantidade de curtidas bem como outras informações de perfil e configurações de conta. Os dados selecionados para serem minerados foram: o id do tweet – número que identifica a mensagem como única. o texto - a mensagem compartilhada na rede social. O nome dos usuários e das hashtags mencionados no tweet como também a data de publicação da mensagem.

Utilizou-se a ferramenta Sentiment140⁶ para a classificação dos tweets conseguindo por consequência a polarização dos sentimentos. Porém, como a ferramenta opera na língua inglesa toda a base de dados necessitava tradução. Para tanto utilizou-se o Google Tradutor e por isso, foi necessário dividir a base de dados de cada evento em fragmentos menores para serem submetidos à tradução. Logo, reduziu-se a base de dados à um total de 20.000 tweets, sendo recolhidos 5.000 tweets para cada um dos eventos monitorados.

A partir de então aplicou-se o modelo de mineração textual proposto por Aranha e Vellasco (2007).

5 | RESULTADOS E DISCUSSÕES

A política é um tema que gera muitas discussões muitas das vezes acaloradas. As pessoas defendem os seus pontos de vista, levantam os prós e contras tanto das propostas como das qualidades de seus candidatos, elas fazem das palavras a sua

6 <http://www.sentiment140.com/>

maior arma para que a mensagem propagada se dissemine e consiga convencer os demais. Porém, o resultado das classificações mostrou que durante os debates e o evento de candidatura do Fernando Haddad, a maioria das mensagens publicadas eram fatos ou opiniões neutras, sem sentimentos explícitos.

A polaridade neutra prevaleceu com resultado acima dos 70% em ambas as classificações. Uma mensagem pode ser considerada neutra, devido à ausência de sentimentos expressos, ou também à presença equilibrada de sentimentos positivos e negativos. E em alguns casos, a polaridade neutra é atribuída a mensagens que não conseguiram ser classificadas nos outros sentimentos.

Outro fator que impacta na classificação dos sentimentos é a forma de escrita da internet. Ela dificulta ainda mais o processo, devido à linguagem predominantemente informal, uso de gírias, expressões populares e abreviações. Dessa forma, algumas palavras não foram traduzidas pelo tradutor e nem reconhecidas pelo classificador, interferindo no resultado.

Os candidatos mais citados durante o debate do SBT, foram Cabo Daciolo e Ciro Gomes. E os candidatos que mais receberam comentários positivos, foram respectivamente: Ciro Gomes, Haddad e Cabo Daciolo. Daciolo, foi citado precisamente 795 vezes e ganhou grande repercussão neste dia após uma de suas declarações no debate. Ciro Gomes também teve muitos tweets positivos (118 ao total).

O debate da Record dentre os demais debates foi o que apresentou o menor número de citações aos candidatos e o que apresentou a menor quantidade de sentimentos classificados. Os candidatos mais citados neste evento foram Ciro Gomes, Bolsonaro e Cabo Daciolo. Ciro Gomes, porém não recebeu nenhuma citação positiva neste debate só tweets negativos e neutros. Bolsonaro recebeu apenas tweets neutros e apesar de não ter comparecido ao debate foi o segundo mais comentado no Twitter.

O Debate da Globo apresentou os sentimentos explícitos de forma mais distribuída. Todos os candidatos que foram citados possuíram sentimentos positivos e negativos. Os candidatos mais citados neste debate foram Guilherme Boulos e Ciro Gomes. O candidato mais citado em tweets positivos foi novamente Ciro e o mais citado em tweets negativos foi o Fernando Haddad.

Quando se analisou o resultado do monitoramento dos debates percebeu-se que o Cabo Daciolo foi muito citado no debate do SBT e na Record, porém, o mesmo não foi convidado para o debate da Globo provocando uma queda drástica com o número de suas citações únicas nas mensagens. Entretanto, o mesmo não ocorreu com Bolsonaro que esteve ausente em todos os debates, porém, a quantidade de mensagens que o citavam foram sempre significativas quando comparadas aos candidatos presentes.

Através deste estudo foi possível se confirmar que o resultado obtido através das mensagens da rede social online, se referem aos sentimentos dos usuários no momento dos eventos, já que foi possível captar a reação instantânea do público às frases e discursos dos candidatos durante o debate.

REFERENCES

- Aranha, C. N. e Vellasco, M. M. B. R. (2007). Uma abordagem de pré-processamento automático para mineração de textos em português: sob o enfoque da inteligência computacional. *Pontifícia Universidade Católica do Rio de Janeiro*, Rio de Janeiro, RJ.
- Balazs, J. A., e Velásquez, J. D. (2016). Opinion mining and information fusion: a survey. *Information Fusion*, 27, 95-110.
- Benevenuto, F., Ribeiro, F. e Araújo, M. (2015). Métodos para análise de sentimentos em mídias sociais. In *Brazilian Symposium on Multimedia and the Web (Webmedia)*, Manaus, Brasil.
- Boyd, D. M., e Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of computer-mediated Communication*, 13(1), 210-230.
- Cavalin, P. R., Gatti, M. D. C., dos Santos, C. N., e Pinhanez, C. (2014). Real-time sentiment analysis in social media streams: The 2013 confederation cup case. *Proceedings of BRACIS/ENIAC*.
- Chew, C., e Eysenbach, G. (2010). Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. *PloS one*, 5(11), e14118.
- Evangelista, T. R., e Padilha, T. P. P. (2013). Monitoramento de posts sobre empresas de ecommerce em redes sociais utilizando análise de sentimentos. In *Brazilian Workshop on Social Network Analysis and Mining (BraSNAM)*.
- Hearst, M. A. (1999, June). Untangling text data mining. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics* (pp. 3-10). Association for Computational Linguistics.
- Hearst, M. (2003). What is text mining. *SIMS*, UC Berkeley.
- Ghiassi, M., Skinner, J., e Zimbra, D. (2013). Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with applications*, 40(16), 6266-6282.
- Kantardzic, M. (2011). *Data mining: concepts, models, methods, and algorithms*. John Wiley & Sons.
- Kaplan, A. M. e Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. In *Business Horizons*, v. 53, n. 1, p. 59–68.
- Liu, B., e Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In *Mining text data* (pp. 415-463). Springer, Boston, MA.
- Malheiros, Y. U. R. I., e Lima, G. E. O. R. G. E. (2013). Uma ferramenta para análise de sentimentos em redes sociais utilizando o sentinet. *IX Simpósio Brasileiro de Sistemas de Informação (SBSI2013)*.
- Malini, F., Ciarelli, P., e Medeiros, J. (2017). O sentimento político em redes sociais: big data, algoritmos e as emoções nos tweets sobre o impeachment de Dilma Rousseffl Political sentiment in social networks: big data, algorithms and emotions in tweets about the impeachment of Dilma Rousseff. *Liinc em Revista*, 13(2).
- Medhat, W., Hassan, A., e Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, 5(4), 1093-1113.
- Nascimento, P., Aguas, R., De Lima, D., Kong, X., Osiek, B., Xexéo, G., e De Souza, J. (2012, July). Análise de sentimento de tweets com foco em notícias. In *Brazilian Workshop on Social Network Analysis and Mining*.

Soares, F. (2013). *Categorização Automática de Textos Baseada em Mineração de Textos* (Doctoral dissertation, PUC-Rio).

Souza, A. O. (2015) Identificação e classificação de usuários influentes no twitter por áreas de influência. TCC. Centro Federal de Educação Tecnológica de Minas Gerais.

Stieglitz, S., e Dang-Xuan, L. (2012, January). Political communication and influence through microblogging--An empirical analysis of sentiment in Twitter messages and retweet behavior. In *2012 45th Hawaii International Conference on System Sciences* (pp. 3500-3509). IEEE.

Vergara, S. C. (1990). *Tipos de pesquisa em administração*. Escola Brasileira de Administração Pública da FGV. Rio de Janeiro. RJ

Vydiswaran, V. G., Zhai, C., e Roth, D. (2011, August). Gauging the internet doctor: ranking medical claims based on community knowledge. In *Proceedings of the 2011 workshop on Data mining for medicine and healthcare* (pp. 42-51). ACM.

Wang, H., Can, D., Kazemzadeh, A., Bar, F., e Narayanan, S. (2012, July). A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In *Proceedings of the ACL 2012 System Demonstrations* (pp. 115-120). Association for Computational Linguistics.

Weiss, S. M., Indurkha, N., Zhang, T., e Damerau, F. (2010). *Text mining: predictive methods for analyzing unstructured information*. Springer Science & Business Media.

ANÁLISE DO IMPACTO DO CASCATEAMENTO DE FILTROS ÓPTICOS EM UM CENÁRIO DE REDES ÓPTICAS ELÁSTICAS

Gabriela Sobreira Dias de Carvalho

Universidade Federal de Campina Grande,
Programa de Pós-Graduação em Engenharia
Elétrica

Campina Grande - Paraíba

William Silva dos Santos

Faculdade Pitágoras

Imperatriz - Maranhão

Lucas Oliveira de Figueiredo

Universidade Federal de Campina Grande, Centro
de Engenharia Elétrica e Informática, Unidade
Acadêmica de Engenharia Elétrica

Campina Grande - Paraíba

Helder Alves Pereira

Universidade Federal de Campina Grande, Centro
de Engenharia Elétrica e Informática, Unidade
Acadêmica de Engenharia Elétrica

Campina Grande – Paraíba

RESUMO: Neste artigo, analisa-se o impacto do estreitamento da largura de banda causado pela concatenação de filtros ópticos. Considerou-se a restrição de que filtros ópticos devem se adequar às diversas larguras de banda solicitadas pelas requisições de chamadas, geradas com granularidade de 12,5 GHz. Comparou-se ainda duas modelagens analíticas, disponíveis na literatura, com o objetivo de se avaliar o impacto dessa penalidade em um cenário de redes ópticas elásticas.

PALAVRAS-CHAVE: Estreitamento da Largura de Banda, Filtros Flexíveis, Filtros Ópticos, Penalidades, Redes Ópticas Elásticas.

ANALYSIS OF THE IMPACT OF OPTICAL FILTERS CONCATENATION IN AN ELASTIC OPTICAL NETWORKS SCENARIO

ABSTRACT: In this article, we analyze the impact of the bandwidth narrowing caused by the concatenation of optical filters. It was considered the constraint that optical filters should fit different bandwidths requested by the call requests generated with 12.5 GHz granularity. Two analytical models were also compared, both available in the literature, in order to evaluate the impact of this penalty on an elastic optical networks scenario.

KEYWORDS: Bandwidth Narrowing, Flexible Filters, Optical Filters, Penalties, Optical Elastic Networks.

1 | INTRODUÇÃO

Para satisfazer as demandas de comunicações e promover máxima utilização dos recursos da rede, as redes ópticas elásticas e seus canais com largura de banda variável vêm ganhando espaço nos últimos anos (ZHANG et al., 2013). Nas redes ópticas elásticas, a comutação dos sinais ópticos é realizada por

meio de comutadores seletivos em frequência (WSS – *Wavelength Selective Switches*). Em WSSs flexíveis, elementos comutadores podem ser agrupados para gerar larguras de banda com maior granularidade (MAROM et al., 2012). As características da função de transferência programável destes dispositivos devem ser cuidadosamente modeladas para caracterizar fielmente suas penalidades (PULIKKASERIL, 2012). Para analisar o impacto que o formato do filtro óptico impõe ao desempenho da rede, é importante descrever o modelo espectral utilizado, especialmente quando formatos avançados de modulação são empregados para aumentar a eficiência espectral do sistema (PULIKKASERIL, 2012).

Neste artigo, analisa-se o impacto do estreitamento da largura de banda causado pela concatenação de filtros ópticos, comparando o modelo tradicional, que utiliza uma função gaussiana, e o modelo proposto por Pulikkaseril (2012), em um cenário de redes ópticas elásticas. O artigo está organizado da seguinte forma: na Seção II, apresentam-se os dois modelos analíticos de filtros ópticos flexíveis utilizados neste trabalho. Na Seção III, a análise da concatenação de filtros ópticos e os resultados obtidos são apresentados. Por fim, as conclusões e os trabalhos futuros são descritos na Seção IV.

2 | MODELAGEM ANALÍTICA DOS FILTROS ÓPTICOS

A equação de uma função gaussiana pode ser modelada da seguinte forma (CARVALHO et al., 2015):

$$S_{sg}(f) = Ae^{-\ln\sqrt{2}\left(\frac{2\Delta f}{BW_{3dB}}\right)^{2n}}, \quad (01)$$

em que A representa a amplitude em W/Hz, Δf a variação de frequência, a partir da frequência central, em Hz, BW_{3dB} a largura de banda a 3dB e n a ordem efetiva da função gaussiana. Strasser et al. (2010) mostraram que uma gaussiana de ordem 4 era suficiente para modelar transmissões em que a largura de banda do sinal transmitido não ultrapassava 65 % da largura de banda do canal, quando não mais que 32 filtros fossem cascadeados. Neste trabalho, além dessa consideração, assumiu-se a modelagem da função gaussiana com n=4 como referência.

Pulikkaseril (2012) propôs uma modelagem para filtros ópticos programáveis. Nessa modelagem, a amplitude do espectro de saída do dispositivo é presumida como a convolução de um filtro retangular ideal com uma função de transferência gaussiana. O resultado dessa operação faz uso da função erro $erf(x)$, conforme a seguinte equação (PULIKKASERIL, 2012):

$$S(f) = \sigma\sqrt{2\pi} \left[erf\left(\frac{B-\Delta f}{\sqrt{2}\sigma}\right) - erf\left(\frac{-B-\Delta f}{\sqrt{2}\sigma}\right) \right], \quad (02)$$

em que σ representa o desvio padrão da função transferência, sendo definido por (PULIKKASERIL, 2012):

$$\sigma = \frac{BW_{OTF}}{2\sqrt{2\ln 2}}, \quad (03)$$

em que BW_{OTF} representa uma medida da resolução do filtro óptico e caracteriza a agilidade da transição entre os estados de bloqueio e de passagem, bem como o formato de suas bordas. Neste artigo, assume-se que $BW_{OTF} = 10$ GHz.

3 | CONCATENAÇÃO DE FILTROS ÓPTICOS

Em uma rede óptica, um sinal óptico pode passar por vários nós antes de atingir o seu destino. Em cada nó da rede, o elemento de comutação pode atuar também como um filtro (CARVALHO et al., 2015). A função de transferência efetiva de um cascadeamento de WSSs é o produto das funções de transferência de seus respectivos filtros ópticos individuais (CARVALHO et al., 2015). Logo, a largura de banda resultante dos filtros ópticos concatenados pode ser menor do que a largura de banda de um filtro óptico individual (PULIKKASERIL, 2012). Esse efeito é conhecido como estreitamento da largura de banda e é tanto menor quanto maior for a região plana do topo da função transferência que modela o filtro óptico (PULIKKASERIL, 2012).

Na Fig. 1, é ilustrado o estreitamento da largura de banda que a concatenação de 32 filtros ópticos causa no espectro resultante da função de transferência de um caminho óptico (*lightpath*), considerando o modelo gaussiano e o modelo utilizado por Pulikkaseril (2012). Para canais de 50 GHz, como os que podem ser utilizados em redes multiplexadas por comprimento de onda (WDM – *Wavelength Division Multiplexing*), as modelagens apresentam curvas similares. No entanto, quando a largura de banda do canal varia, o impacto do cascadeamento é claramente dimensionado de maneira distinta pelos dois modelos. Para avaliar tal situação, foram calculadas as potências ópticas inicial e final de um sinal gaussiano, normalizado e sendo transmitido em cada caso. Considerou-se ainda que a largura de banda do sinal óptico correspondia a 60% da largura de banda do canal.

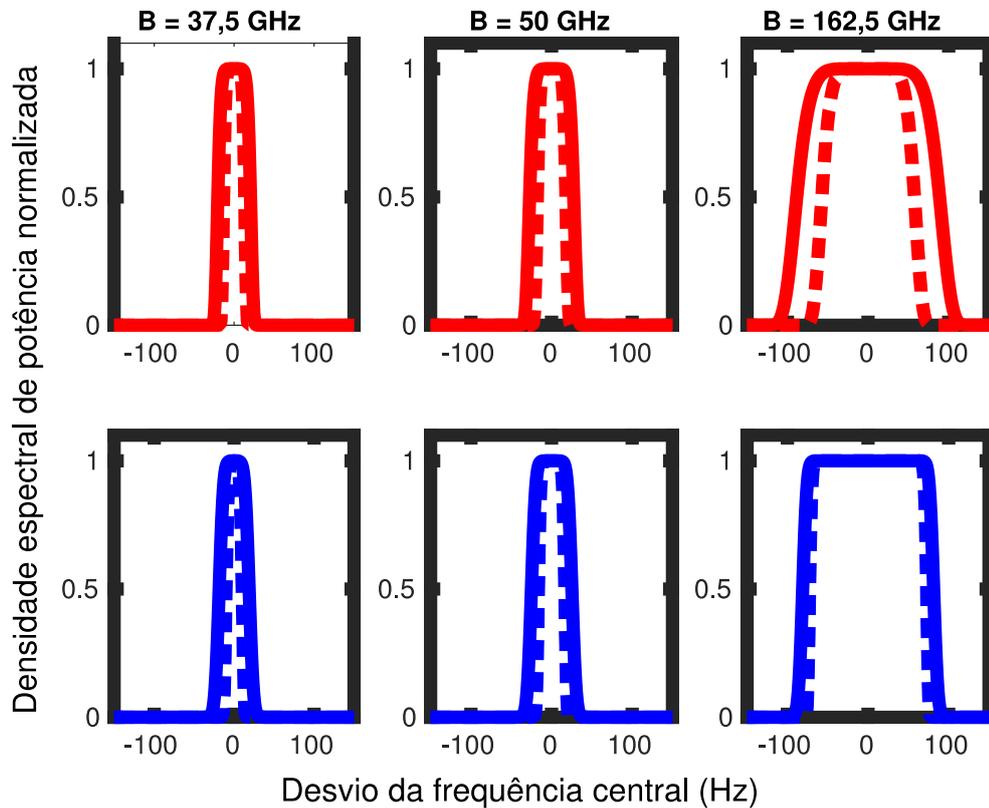


Figura 1: Estreitamento da largura de banda resultante da concatenação de 32 filtros ópticos, considerando o modelo gaussiano (curva em vermelho) e o modelo utilizado por Pulikkaseril (2012) (curva em azul).

Na Tabela I, estão os resultados para diferentes valores de largura de banda do sinal óptico transmitido (37,5; 50 e 167,5 GHz). Dessa forma, utilizando a modelagem gaussiana, a razão entre as potências ópticas final e inicial é fixa, proporcionando uma redução de, aproximadamente, 22% na potência óptica do sinal, após 32 filtros ópticos concatenados. A modelagem gaussiana, para este cenário de simulação, implica em menor atenuação para canais mais estreitos, enquanto que a modelagem proposta por Pulikkaseril (2012) apresenta uma menor perda de potência óptica do sinal ao se considerar sinais ópticos com maior largura de banda.

A análise da diminuição da largura de banda efetiva do filtro óptico em um *lightpath* também pode ser utilizada para se quantificar o estreitamento da largura de banda do sinal óptico causado pelo cascadeamento de filtros ópticos.

Modelo	37,5 GHz	50 GHz	167,5 GHz
Gaussiano	0,78	0,78	0,78
Pulikkaseril [2012]	0,64	0,76	0,91

Tabela I: Razão entre as potências ópticas final e inicial de um sinal óptico gaussiano após ser transmitido por 32 filtros ópticos concatenados e modelados como uma função gaussiana de ordem 4 e com a função de transferência proposta por Pulikkaseril (2012), considerando diferentes valores da largura de banda do sinal óptico transmitido.

Na Fig. 2 estão ilustradas as curvas da largura de banda remanescente do

sinal óptico em função do número de filtros ópticos concatenados, considerando a modelagem gaussiana e a proposta por Pulikkaseril (2012), utilizando diferentes valores de largura de banda do sinal óptico (37,5; 50 e 167,5 GHz). A análise das curvas vermelha e azul reforça que o modelo proposto por Pulikkaseril (2012) tem desempenho superior ao modelo gaussiano em sinais ópticos com maior largura de banda.

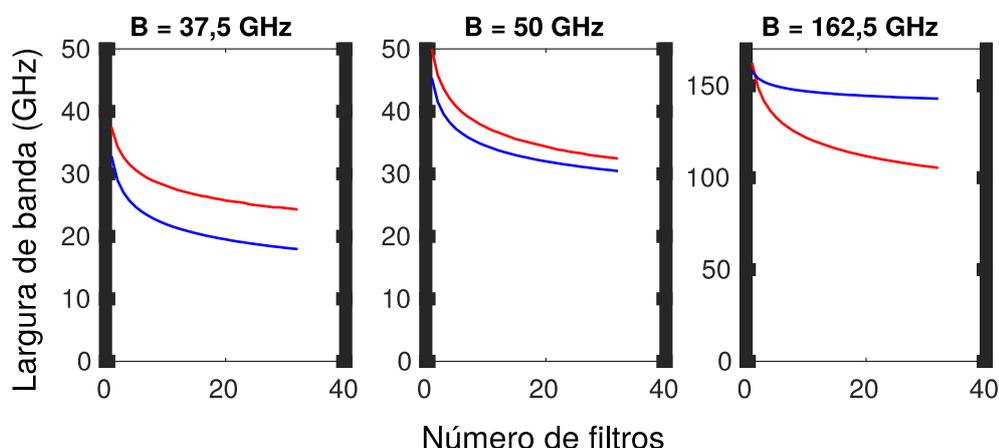


Figura 2: Largura de banda remanescente do sinal óptico em função do número de filtros ópticos concatenados, considerando a modelagem gaussiana (curva vermelha) e a proposta por Pulikkaseril (2012) (curva azul).

4 | CONCLUSÕES E TRABALHOS FUTUROS

Neste artigo foi realizada uma comparação entre a modelagem gaussiana e a modelagem proposta por Pulikkaseril (2012) com o objetivo de quantificar o impacto do cascadeamento de filtros ópticos ao longo de um *lightpath*. Os resultados foram obtidos para filtros ópticos com diferentes larguras de banda, múltiplas de 12,5 GHz. Por meio de simulações em um *lightpath* com 32 filtros ópticos, verificou-se o comportamento da relação entre as potências ópticas dos sinais recebido e transmitido para ambos os modelos, considerando que o estreitamento da largura de banda causada pelo cascadeamento de filtros ópticos era a única limitação da transmissão. A análise dos resultados obtidos reforçou a conclusão de que o modelo proposto por Pulikkaseril (2012) tem desempenho superior ao do modelo gaussiano em sinais ópticos com maior largura de banda. Como trabalhos futuros, espera-se investigar o impacto gerado pela interferência (*crosstalk*) gerada pelos comutadores ópticos em redes ópticas elásticas.

REFERÊNCIAS

CARVALHO, Cleuves C.; ABRANTES, Camila O.; ALMEIDA-JÚNIOR, Raul C. A.; PEREIRA, Helder A. Análise da imperfeição do filtro Óptico em um cenário de redes Ópticas elásticas transparentes. In: **Encontro Anual do Iecom em Comunicações, Redes e Criptografia (ENCOM)**, vol. 1, 2015, pp. 1–2.

MAROM, Dan M.; SINEFELD, David. Beyond wavelength-selective channel switches: Trends in support of flexible/elastic optical networks. In: **2012 14th International Conference on Transparent**

Optical Networks (ICTON). IEEE, 2012. p. 1-4.

PULIKKASERIL, C. Filter Bandwidth Definition of the WaveShaper S-series Programmable Processor. **Finisar product whitepaper**.

STRASSER, Thomas A.; WAGENER, Jefferson L. Wavelength-selective switches for ROADM applications. **IEEE Journal of selected topics in Quantum Electronics**, v. 16, n. 5, p. 1150-1157, 2010.

ZHANG, Guoying et al. A survey on OFDM-based elastic core optical networking. **IEEE Communications Surveys & Tutorials**, v. 15, n. 1, p. 65-87, 2013.

ANÁLISE DE REDE ÓPTICA ELÁSTICA TRANSLÚCIDA CONSIDERANDO DIFERENTES ALGORITMOS DE ROTEAMENTO

Arthur Hendricks Mendes de Oliveira

Universidade Federal de Campina Grande, Centro
de Engenharia Elétrica e Informática, Unidade
Acadêmica de Engenharia Elétrica
Campina Grande - Paraíba

William Silva dos Santos

Faculdade Pitágoras
Imperatriz - Maranhão

Helder Alves Pereira

Universidade Federal de Campina Grande, Centro
de Engenharia Elétrica e Informática, Unidade
Acadêmica de Engenharia Elétrica
Campina Grande – Paraíba

Raul Camelo de Andrade Almeida Júnior

Universidade Federal de Pernambuco, Centro
de Tecnologia e Geociências, Departamento de
Eletrônica e Sistemas
Recife - Pernambuco

RESUMO: Este artigo apresenta uma avaliação do impacto de diferentes algoritmos de roteamento no desempenho de uma rede óptica elástica translúcida. Para isso, considerou-se o ruído de emissão espontânea amplificada gerado nos amplificadores ópticos (de potência, de linha e pré-amplificador), dispostos ao longo dos nós e enlaces da rede, formatos de modulação e taxas de transmissão de bit diferentes para as solicitações de chamada. O desempenho foi avaliado em termos da

probabilidade de bloqueio de chamadas em função do número de regeneradores instalados na rede. Os resultados mostraram que, além da característica do tipo de roteamento (fixo ou dinâmico), deve-se considerar também a política de utilização de regeneradores para otimizar o número de chamadas estabelecidas, considerando tráfego dinâmico, no cenário analisado.

PALAVRAS-CHAVE: Algoritmo de Roteamento, Algoritmo de Utilização de Regeneradores, Rede Óptica Elástica Translúcida.

TRANSLUCENT ELASTIC OPTICAL NETWORK ANALYSIS CONSIDERING DIFFERENT ROUTING ALGORITHMS

ABSTRACT: This paper presents an evaluation of the impact of different routing algorithms on the performance of a translucent elastic optical network. We considered in our simulations the amplified spontaneous emission noise generated in the optical amplifiers (power amplifier, in-line and preamplifier), presented along the nodes and network links, modulation formats and different transmission bit rates for the call requests. The performance was evaluated in terms of the blocking probability of call requests as a function of the number of regenerators installed in the network. The results showed that, in addition to the characteristics of

the routing type (fixed or dynamic), one should also consider the policy of assigning regenerators to optimize the number of established call requests, considering dynamic traffic, in the scenario analyzed.

KEYWORDS: Regenerator Assignment Algorithm, Routing Algorithm, Translucent Elastic Optical Network.

1 | INTRODUÇÃO

As redes ópticas elásticas possuem largura de banda que depende da demanda requerida pelo usuário da rede (CAVALCANTE et al., 2017a). Dessa forma, torna-se mais eficiente na utilização do espectro de frequência em comparação com as redes ópticas que utilizam multiplexação por comprimento de onda (WDM – *Wavelength Division Multiplexing*) (CAVALCANTE et al., 2017b). Com o aumento da necessidade por taxas de transmissão de bit elevadas e o atendimento de chamadas cada vez mais distantes, do ponto de vista geográfico, os sinais ópticos propagantes podem ter sua qualidade de transmissão degradada devido aos efeitos lineares e não-lineares gerados nas fibras ópticas e nos dispositivos presentes ao longo dos nós e enlaces da rede (CAVALCANTE et al., 2015). Dessa forma, pode ser necessário utilizar recursos de regeneração para melhorar a relação sinal-ruído óptica (OSNR – *Optical Signal-to-Noise Ratio*) do sinal óptico recebido (CAVALCANTE et al., 2017b). Com isso, verifica-se na literatura duas linhas de pesquisa no contexto de redes ópticas elásticas translúcidas (CAVALCANTE et al., 2017b): (1) colocação de nós translúcidos e (2) utilização de regeneradores nos nós translúcidos.

Este artigo tem como contribuição avaliar o impacto de diferentes algoritmos de roteamento no desempenho de uma rede óptica elástica translúcida. Para isso, considerou-se o ruído de emissão espontânea amplificada (ASE – *Amplified Spontaneous Emission*) gerado nos amplificadores ópticos (de potência, de linha e pré-amplificador), dispostos nos nós e enlaces da rede, formatos de modulação e taxas de transmissão de bit diferentes para as solicitações de chamada. O desempenho foi avaliado em termos da probabilidade de bloqueio de chamadas em função do número de regeneradores instalados na rede. Está organizado da seguinte forma: na Seção II, apresenta-se os algoritmos de roteamento analisados neste trabalho, bem como os parâmetros e considerações utilizadas nas simulações. Na Seção III, os resultados são descritos e, por fim, na Seção IV, são apresentadas as conclusões e sugestões para trabalhos futuros.

2 | ESTRATÉGIAS DE ROTEAMENTO

Neste trabalho, foram considerados os seguintes algoritmos de roteamento, bem conhecidos na literatura e com adaptações para uso em redes ópticas elásticas: (1) menor número de enlaces (MH – *Minimum Hops*) (ZANG et al., 2000); (2) menor

distância física (SP – *Shortest Path*) (ZANG et al., 2000) e (3) distância e disponibilidade, adaptado para disponibilidade de intervalos de frequência (*slots*) (LORa – *Lenght and Availability Routing*) (CAVALCANTE et al., 2015). Os dois primeiros (MH e SP) são algoritmos de roteamento fixo e o último (LORa) de roteamento dinâmico (ZANG et al., 2000).

2.1 Cenário de Simulação

A rede NSFNet (CAVALCANTE et al., 2017b) foi a topologia utilizada nas simulações. Os ganhos dos amplificadores ópticos foram dimensionados de modo a compensar as perdas do nó e do enlace. Cada nó possui arquitetura denominada de comutação de espectro (*spectrum switching*) (CAVALCANTE et al., 2017b), adaptada para utilização de regeneradores. A relação sinal-ruído óptica obtida no nó destino ($OSNR_{RX}$) pode ser calculada conhecendo-se os valores das perdas nos dispositivos e nas fibras ópticas, além dos ganhos e do ruído ASE gerado nos amplificadores ópticos, levando-se em conta a taxa de transmissão de bit e o formato de modulação da respectiva requisição de chamada (CAVALCANTE, 2015). Assumiu-se que: (1) as chamadas solicitadas seguem uma distribuição poissoniana; (2) a duração de cada chamada estabelecida segue uma distribuição exponencial e (3) a seleção dos pares fonte- destino e da taxa de transmissão de bit, para um dado pedido de conexão, seguem uma distribuição uniforme. O número de requisições simuladas foi de um milhão de chamadas. Para um determinado pedido de requisição de chamada, o algoritmo de roteamento em análise (MH, SP ou LORa) e o de atribuição espectral de primeiro preenchimento (FF – *First Fit*) são utilizados para verificar se a chamada pode ser estabelecida. Primeiro, tenta-se estabelecer a chamada no formato de modulação com maior eficiência espectral (neste trabalho, 64-QAM). Caso não se consiga estabelecer a chamada nesse formato de modulação, tenta-se o próximo mais eficiente, até que não se consiga estabelecer a chamada de forma transparente (sem utilização de regeneradores). Foram considerados dois algoritmos de utilização de regeneradores propostos para redes ópticas elásticas translúcidas (CAVALCANTE et al., 2017b): 1) o de maior alcance transparente (FLR – *First Longest Reach*) e 2) o de melhor utilização espectral (FNS – *First Narrowest Spectrum*). Os valores dos parâmetros utilizados nas simulações foram os seguintes: (1) espaçamento de 80 km entre amplificadores de linha (*span length*); (2) 100% dos nós com capacidade de regeneração; (3) carga da rede igual a 150 Erlangs; (4) 64 *slots* por enlace; (5) formatos de modulação utilizados: 4, 8, 16, 32 e 64-QAM; (6) taxas de transmissão de bit de 100 a 500 Gbps; (7) 12,5 GHz de largura de banda de um *slot*; (8) 12,5 GHz de largura de banda de referência; (9) 5 dB de perda nos elementos de comutação; (10) 5 dB de figura de ruído nos amplificadores ópticos; (11) 0 dBm de potência de entrada do sinal óptico e (12) 30 dB de OSNR na transmissão ($OSNR_{TX}$). Assumiu-se também que os regeneradores podem, além de regenerar o sinal óptico, realizar conversão espectral. As simulações foram realizadas utilizando-se o *software* de código aberto

3 | RESULTADOS

Fig. 1 ilustra a probabilidade de bloqueio de chamadas em função do número de regeneradores instalados na rede, considerando diferentes algoritmos de roteamento (MH, SP e LORa) e políticas de utilização de regeneradores (FLR e FNS). Percebe-se que, independente do algoritmo de roteamento utilizado (MH, SP ou LORa), existe um ponto de cruzamento (R_c) em que, dependendo da política de utilização de regeneradores (FLR ou FNS), os valores da probabilidade de bloqueio (PB) são minimizados. Para um número de regeneradores abaixo do ponto de cruzamento ($R < R_c$), usar a política FLR otimiza o estabelecimento de chamadas na rede visto que essa política prioriza o uso de regeneradores apenas quando necessário (maior alcance transparente), favorecendo um cenário de escassez de recursos de regeneração. Para um número de regeneradores acima do ponto de cruzamento ($R > R_c$), a política FNS fornece valores de PB menores, visto que essa técnica utiliza os regeneradores sempre que possível, favorecendo um cenário onde os recursos de regeneração são abundantes, resultando em compactação do espectro de frequência. Pode-se notar ainda que, o algoritmo de roteamento com característica dinâmica (LORa), forneceu valores menores de PB devido à sua característica de determinar rotas dependendo do estado da rede (roteamento dinâmico) em comparação com o MH e o SP (roteamento fixo). Ao se utilizar a política FLR, obteve-se um ganho (g_{PB}), em termos de PB, de aproximadamente 3,33, utilizando o LORa-FLR em comparação com o MH-FLR e o SP-FLR. Enquanto que, utilizando a política FNS, obteve-se $g_{PB} \approx 5$ para o LORa-FNS, comparando com o SP-FNS, e de $g_{PB} \approx 10$, comparando com o MH-FNS. Comparando o LORa com relação às políticas de utilização de regeneradores (LORa-FNS e LORa-FLR), obteve-se $g_{PB} \approx 3,33$. Com relação ao pior caso, MH-FLR ou SP-FLR, o LORa-FNS obteve $g_{PB} \approx 16,67$.

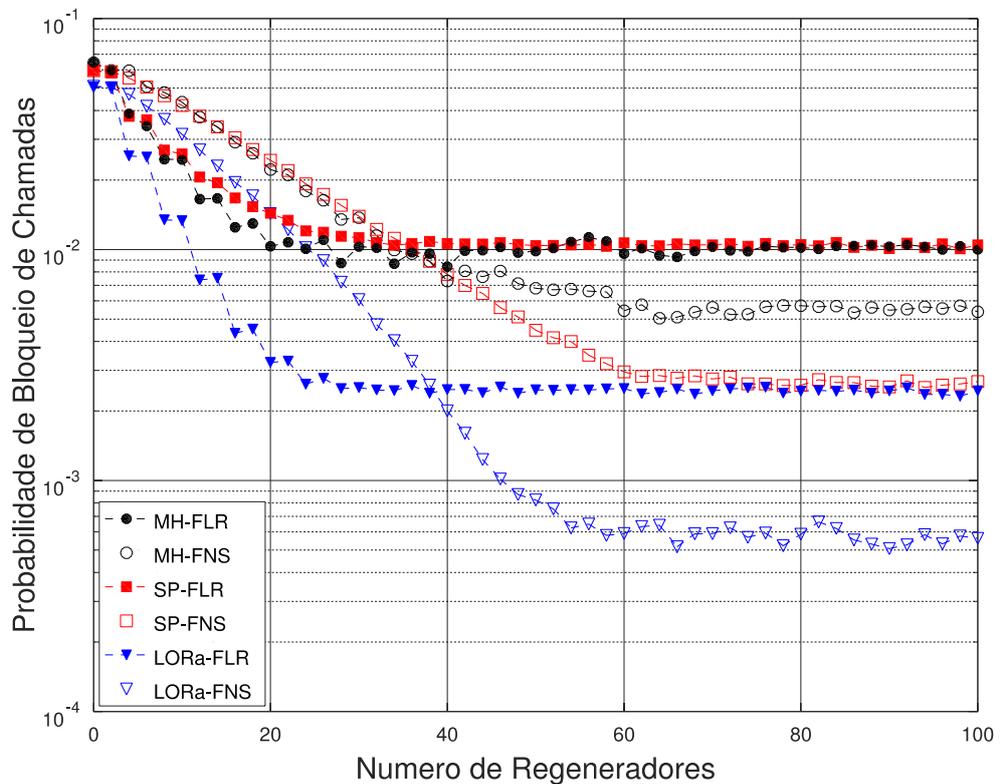


Figura 1: Probabilidade de bloqueio de chamadas em função do número de regeneradores instalados na rede, considerando diferentes algoritmos de roteamento (MH, SP e LORa) e políticas de utilização de regeneradores (FLR e FNS).

4 | CONCLUSÕES E TRABALHOS FUTUROS

Neste artigo avaliou-se o impacto de algoritmos de roteamento no desempenho de uma rede óptica elástica translúcida. Para isso, considerou-se o ruído de emissão espontânea amplificada gerado nos amplificadores ópticos (de potência, de linha e pré-amplificador), dispostos nos nós e enlaces da rede, formatos de modulação e taxas de transmissão de bit diferentes para as solicitações de chamada. O desempenho foi avaliado em termos da probabilidade de bloqueio de chamadas em função do número de regeneradores instalados na rede, considerando a topologia NSFNet. Os resultados mostraram que o LORa (roteamento dinâmico) obteve melhor desempenho, utilizando FLR ou FNS, comparado aos algoritmos MH e SP (roteamento fixo). Considerando a política de utilização de regeneradores, o LORa-FNS obteve um ganho de aproximadamente 16,67 quando comparado aos algoritmos MH-FLR e SP-FLR no cenário analisado neste artigo. Como trabalhos futuros, espera-se considerar outras topologias presentes na literatura e investigar o impacto das políticas de utilização de regeneradores, bem como de outros algoritmos de roteamento, em diversos cenários de redes ópticas elásticas translúcidas.

REFERÊNCIAS

CAVALCANTE, Matheus A. et al. Applying power series routing algorithm in transparent elastic optical networks. In: **2015 SBMO/IEEE MTT-S International Microwave and Optoelectronics Conference (IMOC)**. IEEE, 2015. p. 1-5.

CAVALCANTE, Matheus A.; PEREIRA, Helder A.; ALMEIDA, Raul C. SimEON: an open-source elastic optical network simulator for academic and industrial purposes. **Photonic Network Communications**, v. 34, n. 2, p. 193-201, 2017.

CAVALCANTE, Matheus A. et al. Estado da Arte de Redes Ópticas Elásticas Translúcidas para Cenários de Tráfego Dinâmico em Redes Metropolitanas e de Longas Distâncias. **Revista de Tecnologia da Informação e Comunicação**, v. 7, n. 2, p. 24-32, 2017.

ZANG, Hui et al. A review of routing and wavelength assignment approaches for wavelength-routed optical WDM networks. **Optical networks magazine**, v. 1, n. 1, p. 47-60, 2000.

ANÁLISE DE REDE ÓPTICA ELÁSTICA TRANSLÚCIDA CONSIDERANDO ALGORITMOS DE ATRIBUIÇÃO ESPECTRAL

Arthur Hendricks Mendes de Oliveira

Universidade Federal de Campina Grande, Centro
de Engenharia Elétrica e Informática, Unidade
Acadêmica de Engenharia Elétrica
Campina Grande - Paraíba

William Silva dos Santos

Faculdade Pitágoras
Imperatriz - Maranhão

Helder Alves Pereira

Universidade Federal de Campina Grande, Centro
de Engenharia Elétrica e Informática, Unidade
Acadêmica de Engenharia Elétrica
Campina Grande – Paraíba

Raul Camelo de Andrade Almeida Júnior

Universidade Federal de Pernambuco, Centro
de Tecnologia e Geociências, Departamento de
Eletrônica e Sistemas
Recife - Pernambuco

RESUMO: Neste artigo, avalia-se o impacto na probabilidade de bloqueio de chamadas em função do número de regeneradores instalados na rede para diversos algoritmos de atribuição espectral, considerando um cenário de rede óptica elástica translúcida. Para isso, considerou-se o ruído de emissão espontânea amplificada, gerado nos amplificadores ópticos dispostos nos nós e enlaces da rede, como penalidade física, além de diferentes taxas de transmissão de bit e formatos de modulação. Observou-se

que, além da característica do algoritmo de atribuição espectral de compactar o espectro de frequência, é importante considerar a política de utilização de regeneradores, dependendo do cenário de escassez ou abundância de recursos disponíveis, para regeneração do sinal óptico na rede.

PALAVRAS-CHAVE: Algoritmo de Atribuição Espectral, Algoritmo de Utilização de Regeneradores, Rede Óptica Elástica Translúcida.

TRANSLUCENT ELASTIC OPTICAL NETWORK ANALYSIS CONSIDERING DIFFERENT SPECTRUM ASSIGNMENT ALGORITHMS

ABSTRACT: In this article, we evaluate the impact on the blocking probability of call requests as a function of the number of regenerators installed in the network for several spectral assignment algorithms, considering a translucent elastic optical network scenario. We considered in our simulations the amplified spontaneous emission noise, generated in the optical amplifiers presented along the nodes and network links, as physical penalty, in addition to different transmission bit rates and modulation formats. It was observed that in addition to the characteristic of the spectral assignment algorithm to compact the frequency

spectrum, it is important to consider the policy of regenerator assignment, depending on the scenario of scarcity or abundance of available resources, for regeneration of the optical signal in the network.

KEYWORDS: Regenerator Assignment Algorithm, Spectrum Assignment Algorithm, Translucent Elastic Optical Network.

1 | INTRODUÇÃO

O sinal óptico ao se propagar ao longo dos nós e enlaces da rede pode sofrer degradação na sua qualidade de transmissão (QoT – *Quality of Transmission*) (CAVALCANTE et al., 2015). Em redes ópticas elásticas (EON – *Elastic Optical Network*), a QoT depende da taxa de transmissão de bit e do formato de modulação, além da largura de banda utilizada para transmissão da informação, como nas redes ópticas que utilizam multiplexação por comprimento de onda (WDM – *Wavelength Division Multiplexing*) (CAVALCANTE et al., 2017a). Dessa forma, em um cenário que considera o impacto das penalidades físicas no estabelecimento de chamadas, é importante considerar a utilização de recursos que podem regenerar o sinal óptico e aumentar sua respectiva QoT. Em redes ópticas translúcidas existem duas estratégias para melhorar o desempenho dessas redes, são elas (CAVALCANTE et al., 2017a): (1) alocação de nós translúcidos (RP – *Regenerator Placement*) e (2) utilização de regeneradores (RA – *Regenerator Assignment*). A primeira estabelece os nós que possuem a capacidade de melhorar a QoT dos sinais ópticos por meio da regeneração eletrônica. Enquanto que a segunda estabelece em qual nó translúcido e de que forma o sinal óptico será regenerado.

Este artigo apresenta uma avaliação de uma rede óptica elástica translúcida, em termos da probabilidade de bloqueio de chamadas em função do número de regeneradores instalados na rede, considerando diversos algoritmos de atribuição espectral. Para isso, considerou-se o ruído de emissão espontânea amplificada (ASE – *Amplified Spontaneous Emission*), gerado nos amplificadores ópticos dispostos nos nós e enlaces da rede, como penalidade física, além de diferentes taxas de transmissão de bit e formatos de modulação. Este artigo está organizado da seguinte forma: na Seção II, as estratégias de atribuição espectral são descritas, bem como o cenário, os parâmetros e as considerações utilizadas nas simulações. Na Seção III, são discutidos os resultados e, por fim, na Seção IV, as conclusões deste artigo e os trabalhos futuros são apresentados.

2 | ESTRATÉGIAS DE ATRIBUIÇÃO ESPECTRAL

Neste trabalho, foram considerados os seguintes algoritmos de atribuição espectral, com adaptações para uso em redes ópticas elásticas (ZANG et al, 2000): (1) melhor preenchimento (BF – *Best Fit*); (2) preenchimento exato (EF – *Exact Fit*); (3)

primeiro preenchimento (FF – *First Fit*) e (4) preenchimento aleatório (RF – *Random Fit*).

2.1 Cenário de Simulação

A NSFNet (CAVALCANTE et al., 2017a) foi a topologia utilizada nas simulações. Os ganhos dos amplificadores (de potência, de linha e pré-amplificadores) foram dimensionados de modo a compensar as perdas do nó e do enlace. Cada nó possui arquitetura denominada de comutação de espectro (*spectrum switching*) (CAVALCANTE et al., 2017a), adaptada para utilização de regeneradores. A relação sinal-ruído óptica obtida no nó destino ($OSNR_{RX}$) pode ser calculada conhecendo-se os valores das perdas nos dispositivos e na fibra óptica, além dos ganhos e da potência de ruído ASE gerada nos amplificadores ópticos, levando-se em conta a taxa de transmissão de bit e o formato de modulação da chamada (CAVALCANTE et al., 2015). Assumiu-se que: (1) as chamadas solicitadas seguem uma distribuição poissoniana; (2) a duração de cada chamada segue uma distribuição exponencial; (3) a seleção dos pares fonte-destino, bem como a seleção da taxa de transmissão de bit para uma dada requisição de chamada, seguem uma distribuição uniforme. O número de requisições simuladas foi de 10^6 chamadas. Para um determinado pedido de requisição de chamada, o algoritmo de roteamento de menor distância física (SP – *Shortest Path*) e o de atribuição espectral em análise (BF, EF, FF ou RF) verificam se a chamada pode ser estabelecida. Primeiro, tenta-se estabelecer a chamada no formato de modulação com maior eficiência espectral (neste trabalho, 64-QAM). Caso não se consiga estabelecê-la nesse formato de modulação, tenta-se o próximo mais eficiente, até que não se consiga estabelecê-la de forma transparente (sem utilização de regeneradores). Para isso, foram considerados dos algoritmos de utilização de regeneradores (CAVALCANTE et al., 2017a): (1) o de maior alcance transparente (FLR – *First Longest Reach*), que privilegia o menor uso de regeneradores, e (2) o de melhor utilização espectral (FNS – *First Narrowest Spectrum*), que privilegia o uso do formato de modulação mais eficiente do ponto de vista espectral. Assumiu-se também que os regeneradores podem, além de regenerar o sinal óptico, realizar conversão espectral. Os valores dos parâmetros utilizados nas simulações foram os seguintes: (1) espaçamento de 80 km entre amplificadores de linha (*span length*); (2) 100% dos nós com capacidade de regeneração (nós translúcidos); (3) carga da rede igual a 150 Erlangs; (4) 64 *slots* por enlace; (5) formatos de modulação utilizados: 4, 8, 16, 32 e 64-QAM; (6) taxas de transmissão de bit de 100 a 500 Gbps; (7) 12,5 GHz de largura de banda de um *slot*; (8) 12,5 GHz de largura de banda de referência; (9) 5 dB de perda nos elementos de comutação; (10) 5 dB de figura de ruído nos amplificadores ópticos; (11) 0 dBm de potência de entrada do sinal óptico e (12) 30 dB de OSNR na transmissão ($OSNR_{TX}$). As simulações foram realizadas utilizando-se o *software* de código aberto SimEON (CAVALCANTE et al., 2017b).

3 | RESULTADOS

Fig. 1 ilustra a probabilidade de bloqueio de chamadas em função do número de regeneradores instalados na rede, considerando algoritmos de atribuição espectral (BF, EF, FF e RF) e políticas de utilização de regeneradores (FLR e FNS). Percebe-se que, para os algoritmos de atribuição espectral simulados, obtém-se um valor crítico no número de regeneradores (R_c) de modo que, acima dele ($R > R_c$), a política FNS fornece valores menores de probabilidade de bloqueio de chamadas (PB). Isso ocorre porque a política FNS privilegia utilizar os recursos de regeneração disponíveis nos nós translúcidos sempre que possível, realizando, além da regeneração do sinal óptico, a conversão espectral. Isso permite que os formatos de modulação mais eficientes, do ponto de vista espectral, sejam utilizados, tornando possível um maior estabelecimento de chamadas no cenário de abundância de recursos de regeneração. Para o cenário de escassez de recursos de regeneração ($R < R_c$), a política FLR se destaca porque utiliza os recursos de regeneração só quando necessário, privilegiando o estabelecimento do caminho óptico (*lightpath*) com o maior alcance transparente, disponibilizando recursos para que outras chamadas possam utilizá-los. Do ponto de vista do algoritmo de atribuição espectral utilizado, tanto para o cenário de escassez quanto para o de abundância de recursos de regeneração, os que obtiveram valores menores de PB foram os algoritmos BF, EF e FF, respectivamente, associados com as devidas políticas de utilização de regeneradores para o cenário avaliado (FLR - escassez e FNS - abundância). Isso ocorre devido à característica de compactação espectral desses algoritmos, de modo que utilizam o espectro de frequência de forma mais eficiente, diminuindo o impacto da fragmentação espectral (CAVALCANTE et al., 2017a), tornando possível o estabelecimento de um maior número de chamadas. Interessante perceber também que, dependendo do algoritmo de atribuição espectral, pode não fazer diferença utilizar qualquer uma das políticas de utilização de regeneradores (por exemplo, RF), resultando em valores de PB bem similares.

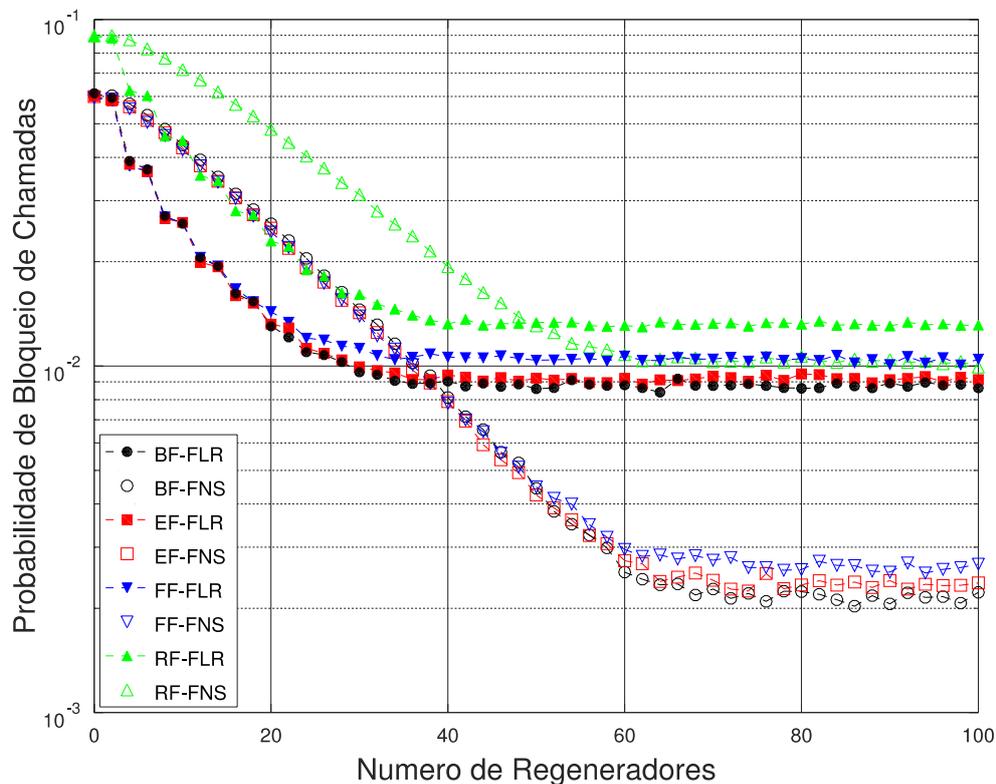


Figura 1: Probabilidade de bloqueio de chamadas em função do número de regeneradores instalados na rede, considerando algoritmos de atribuição espectral (BF, EF, FF e RF) e políticas de utilização de regeneradores (FLR e FNS).

4 | CONCLUSÕES E TRABALHOS FUTUROS

Neste artigo, avaliou-se o impacto na probabilidade de bloqueio de chamadas em função do número de regeneradores instalados na rede para diversos algoritmos de atribuição espectral, considerando um cenário de rede óptica elástica translúcida. Para isso, considerou-se o ruído de emissão espontânea amplificada, gerado nos amplificadores ópticos dispostos nos nós e enlaces da rede, como penalidade física, além de diferentes taxas de transmissão de bit e formatos de modulação. Os algoritmos de atribuição espectral que realizam compactação espectral obtiveram os melhores resultados em termos de probabilidade de bloqueio de chamadas (BF, EF e FF), enquanto que os que não tem essa característica (RF-FLR e RF-FNS) não apresentaram bons resultados, independente da política de utilização de regeneradores utilizada. Como trabalhos futuros, espera-se utilizar outras topologias e considerar outros cenários de simulação para investigar o impacto dos algoritmos de atribuição espectral em redes ópticas elásticas translúcidas.

REFERÊNCIAS

CAVALCANTE, Matheus A. et al. Applying power series routing algorithm in transparent elastic optical networks. In: **2015 SBMO/IEEE MTT-S International Microwave and Optoelectronics Conference (IMOC)**. IEEE, 2015. p. 1-5.

CAVALCANTE, Matheus A. et al. Estado da Arte de Redes Ópticas Elásticas Translúcidas para

Cenários de Tráfego Dinâmico em Redes Metropolitanas e de Longas Distâncias. **Revista de Tecnologia da Informação e Comunicação**, v. 7, n. 2, p. 24-32, 2017.

CAVALCANTE, Matheus A.; PEREIRA, Helder A.; ALMEIDA, Raul C. SimEON: an open-source elastic optical network simulator for academic and industrial purposes. **Photonic Network Communications**, v. 34, n. 2, p. 193-201, 2017.

ZANG, Hui et al. A review of routing and wavelength assignment approaches for wavelength-routed optical WDM networks. **Optical networks magazine**, v. 1, n. 1, p. 47-60, 2000.

A NEW MULTI OBJECTIVE APPROACH FOR OPTIMIZING P-MEDIAN MODELING IN SCHOOL ALLOCATION USING GENETIC ALGORITHM

Clahildek Matos Xavier

Operations and Management Center of the Amazonian Protection System (CENSIPAM) – Manaus, AM – Brazil
clahildek@gmail.com

Marly Guimarães Fernandes Costa

Federal University of Amazonas (UFAM) – Manaus, AM – Brazil
marlygfcosta@gmail.com

Cícero Ferreira Fernandes Costa Filho

Federal University of Amazonas (UFAM) – Manaus, AM – Brazil
cfcfilho@gmail.com

ABSTRACT: This paper proposes a new methodology for a multi parameter approach using the p-median. The optimized solution must fulfill the following criteria: It must minimize distances between cities without a university to a city that has a university; It must prioritize cities with a higher population; It must prioritize cities with a lower United Nation Human Development Index. The use of a multiparametric approach was only possible by introducing the concept of a generalized distance. The results compare the existing distribution of campuses of the Federal system with the best location resulting from the multi parameter method proposed here. Locations for expansion of the current established university in Amazonas State,

Brazil, are proposed.

KEYWORDS: p-Median modelling. Generalized Distance. Locating Higher Education Schools. Genetic Algorithm.

1 | INTRODUCTION

The percentage of the population in Brazil with access to college education is only 14%, according to an OECD report (OECD, 2015), so there is much room for growth of higher education in Brazil. There are national policies in place intended to increase the population of college graduates, by means of grants that allow student admission to private Higher Education Institutions (HEI) through programs such as REUNI – Restructuring and Expansion of Federal Universities (Ministry of Education, 2010). The REUNI program provide slots for 30% of the population within the age group of 18 to 24, based on expanding current degree programs or creating new ones.

Provisions for expanding Higher Education Establishments (HEE) under REUNI towards the interior of Brazil are meant to establish, between years 2013 and 2018, an HEE in every city with either over 200,000 inhabitants or over 200km away from state HEE campuses. Both criteria determined for the country, in a broad sense, require further analysis as far as the Brazilian

Amazon region is concerned due to its distinctive features, including long distances between cities and low population density overall.

In order to provide an overview of this context, the state of Amazonas-Brazil comprises 62 cities and accounts for more than 180,000 Secondary School Pupils (SSP) living in the interior, totaling some 50% of the entire SSP population in the state (Brazilian Institute of Geography and Statistics, 2012). The other 50% lives in the capital, Manaus. The SSP population is the one that will enter in an HEE. Access to such cities is typically managed by inland waterway navigation and the distances between locations can reach 1,500km (National Waterway Transportation Agency, 2013). Only the State capital city, Manaus, is home to a population greater than 200,000 dwellers (Brazilian Institute of Geography and Statistics, 2010).

Transportation of students from the interior to the nearest higher education establishment is time consuming and usually done by river. An alternative way to diminish such transportation difficulties, as per REUNI proposals, is to establish new higher education units in the interior. In Amazonas State, the federal system of higher education comprises 6 units. The aims of this paper concerns the higher education federal system of Amazonas State and are as follows: show that a high yet impractical number of higher education units are necessary to satisfy the REUNI criteria; evaluate the present distribution of higher education units of federal system; and propose a new methodology for selecting the best sites for constructing new college campuses.

The optimal location of schools has been addressed in the literature. Some studies solve a single objective optimization problem as follows. Pizzolato et al. (2004) proposed the use of uncapacitated p -median models (PMP) as well as capacitated (CPMP) to determine the location for p schools in Vitoria, Espirito Santo state, Brazil. The optimization criterion is to minimize students total travel distance to the nearest school. A Lagrangian heuristic was used for the optimization problem. Currently, the total distance that students travel to schools they are assigned to was 99,220m. After determining location, by means of the uncapacitated p -median method, this distance decreased to 70,252m and to 85,080m with capacitated modeling.

In Teixeira et al. (2007), a variant of the p -median model was employed in order to determine p schools, in the city of Coimbra, Portugal. The optimization criterion is maximizing the accessibility of students to schools, with constraints on maximum and minimum capacity occupation. The optimization problem considered 11 existing secondary schools, 43 population centers and the total student-to-school distance of 21,097.8km. An XPRESS-MP mixed-integer programming optimizer was used. They achieved an optimized value of 10,861.2km for the sum of distances, for 14 schools: 10 existing schools are maintained, 4 new schools are created and only one existing school is closed.

Menezes and Pizzolato (2014) proposed location of new schools, during the years 2015 – 2020, in the Guaratiba area, assessment region 5, Rio de Janeiro, Brazil, relying on capacitated p -median and maximum coverage models. They sought

to minimize total student travel distance to the nearest school. Each school had a 1,300-student maximum capacity and the total number of new schools was 15. For the second model they sought to maximize population coverage within a pre-set 1,500m maximum distance. Both approaches were solved using an AIMMS (Advanced Interactive Multidimensional Modeling System) tool based on a CPLEX solver.

Some studies solve multi objective problems. In Gac et al. (2009) the authors aim at maximizing individuals' utility and school profits. The objective function to be maximized defines the social benefits, which are modeled as aggregated quantities and measure the economic profits gained by schools and aggregate individuals' surplus. The authors employed a linear deterministic model. Amaya et al. (2015) proposed a model for determining the location of schools aiming at minimizing school fixed and overhead costs such as per-hour rates. Such work took into consideration the cost for opening and closing existing school facilities. The problem was optimized using CPLEX 9.1. They managed to reduce over 30% of total costs for a 5-year period.

The aforementioned papers that use the p-median model to determine school location have only a single objective: minimize student travel distance, with the center population as weight. This paper aims to evaluate application of a p-median model, using a new multi objective approach. The multi objectives considered are the following: minimize student travel distance, maximize population on centers with allocated schools and minimize the United Nation's Human Development Index (UN HDI) of centers with allocated schools. This new multi objective proposal for p-median is only possible through the use of a generalized distance that encompasses three parameters: student travel distance, center population and UN HDI. In this paper, these parameters are normalized using median approach. Kariv and Hakimi (1979) have shown that the p-median problem is NP-hard. With the aim of obtaining the best location with the 62 cities, we employed the Genetic Algorithm - GA.

This paper is organized as follows. Section 2 presents how the Data used in the study were obtained. Section 3 shows the generalized distance, the normalization parameters and the evaluation index used to compare two location solutions. Section 4 presents the location problem modeling with the p-median model. Section 5 shows how GA was modeled to solve the p-median problem. Section 6 shows the results of locating a university or college for the entire set of cities using GA. Section 7 presents the conclusions and outlines the main contributions of this study.

2 | DATA

The Data used in this work consist of distances between cities, their populations and UN HDI. Population and UN HDI were acquired from the National Geography and Statistics Institute – IBGE (2010, 2012). The distances considered between cities were those corresponding to fluvial distances, which refer to the main means of transportation used in that region.

Some fluvial distances were established through data from the National Waterway Transportation Agency – ANTAQ (2013). Other distances missing from this database were acquired by means of one of the following methods: 1) subtraction between known routes along the same river; 2) using Google Maps API to measure the route between two cities set in different rivers or tributaries.

The Federal University of Amazonas (UFAM) has campuses in six cities: Manaus (capital), Parintins, Itacoatiara, Benjamin Constant, Coari, and Humaitá (Federal University of Amazonas, 2006).

3 | GENERALIZED DISTANCE, NORMALIZATION OF PARAMETERS AND EVALUATION INDEX

Choosing a particular city for locating a college or university must abide by the following criteria: 1) cities with a larger SSP population must be prioritized; 2) cities with the lowest UN HDI must be prioritized; 3) minimizing the distances between those cities with established campuses and those without must be pursued. In order to insert such criteria into the solution by means of p-median, the concept of generalized distance was created to combine all these criteria into a single equation (Xavier et al., 2016). One generalized distance has been proposed, the generalized Manhattan distance, Eq.(1). This distance was named generalized for they include non-geometric coordinates.

$$d_{ij}^{Man_gen} = d_{ij} + \frac{|p_i - p_j|}{\left(\frac{p_i + p_j}{2}\right) |p_i - p_j| + k_1} + \frac{\left(\frac{HDI_i + HDI_j}{2}\right) |HDI_i - HDI_j|}{|HDI_i - HDI_j| + k_2} \quad (1)$$

In Eq.(1): $d_{ij}^{Man_gen}$ is the generalized Manhattan distance; d_{ij} is the normalized river navigation distance between cities i and j ; p_i and p_j are the normalized student population for cities i and j , respectively; HDI_i and HDI_j are the normalized UN HDI for cities i and j , respectively; k_1 and k_2 are extremely small constants compared to other terms and, in this case, were determined to equal 10^{-4} . These two constants are intended to prevent indetermination in expression (1), when $i=j$, for it satisfies condition of reflexivity.

This distance hereby proposed is proximity measure, for they satisfy conditions of similarity, positivity, and reflexivity (Xu and Wunsch, 2008). However, such distance is not proximity metric, for it fails to satisfy the triangle inequality condition. The parameters used in Eq.(1) were normalized using median values, Eq.(2):

$$x_i = x_i^* / X_i \quad (2)$$

In Eq.(2): x_i^* stands for non-normalized parameters distance, population or UN's

HDI for cities; x_i is the normalized parameter value; X_i is the median (Bussab and Morettin, 2009) of a parameter x_i^* . The index proposed in this study for evaluating a location solution is given by means of Eq.(3):

$$Index = \left(\frac{\sum_{i=1}^n \sum_{j=1}^n d_{ij} x_{ij}}{n - p} \right)^{-1} + \frac{\sum_{i=1}^n p_i x_{ii}}{p} + \left(\frac{\sum_{i=1}^n UN's\ HDI_i x_{ii}}{p} \right)^{-1} \quad (3)$$

In Eq.(3): d_{ij} is the fluvial distance obtained in a location solution, normalized by maximum and minimum values, between city i with no college to city j with a college nearest to it; p_i is the population of city i , normalized by maximum and minimum values, of population of cities with a college; UN's HDI are the UN HDI, normalized by maximum and minimum values, of cities with a college or university; p and n are the total number of colleges and total number of cities, respectively, with $0 < p < n$. Eq.(8) defines the variable x_{ij} .

4 | PROBLEM MODELING WITH P-MEDIAN

This model aims to find an optimal solution for allocating higher education institution using a multiobjective approach. Mathematical modeling for the p-median problem on locating education schools in this study corresponds to:

Minimize:

$$f = \sum_{i=1}^n \sum_{j=1}^n d_{ij}^{Man_gen} x_{ij} \quad (4)$$

Subject to:

$$\sum_{i=1}^n x_{ii} = p \quad (5)$$

$$\sum_{j=1}^n x_{ij} = 1, \quad \forall i \in \{1, \dots, n\} \quad (6)$$

$$\sum_{i=1}^n x_{ij} \leq n x_{jj}, \quad \forall j \in \{1, \dots, n\} \quad (7)$$

$$x_{ij} = \begin{cases} 1, & \text{if } \exists \text{ a school in municipality } j \text{ and if} \\ & \text{municipality } i \text{ is nearest to } j \text{ than any other} \\ & \text{municipality with an allocated school;} \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

The p-median modeling just presented is a classical one. The only difference is that the objective function of Eq.(4) minimizes the generalized Manhattan distance. The constraint of Eq.(5) determines that precisely p facilities must be open. Constraint of Eq.(6) requires that each city be assigned a single campus. Constraint of Eq.(7) forces cities with no allocated schools to be associated only with cities with allocated schools. Eq.(8) indicates the binary nature of the decision variables.

5 | GENETIC ALGORITHM MODELING FOR SOLVE P-MEDIAN PROBLEM

Holland (1975) is credited with originally designing the first genetic algorithm (GA). According to Chong and Zak (2013), the main steps of GA are defined in the flowchart of Figure 1. The genetic algorithm starts with a population $P(0)$ of N chromosomes. The chromosome is an encoding of the problem information and encompasses the following data: alphabet, length and encoding. For the p-median problem of this paper, we use the following data: alphabet – {0, 1}; length – 62; encoding – 0 corresponds to a city without a college and 1 corresponds to a city with a college. The stopping criterion used is the number of iterations of the GA. Originally, the fitness function is the function given by Eq.(4), using the generalized Manhattan distance. Next, we discuss the three operators of the GA, selection, crossover and mutation used for solving the p-median problem. The last two are used in the evolution stage shown in Figure 1.

The selection operator is employed to select the members of the mating pool, $M(k)$. The size of the mating pool, N, is the same size of the population $P(k)$. To select the members for the mating pool, the roulette-wheel scheme was used. The probability of selecting a member of $P(k)$ for the mating pool using the roulette was defined through the linear rank method. This probability is given in Eq.(9).

$$p(s_i) = \text{rank } s_i / \sum_{k=1}^n \text{rank } s_k \quad (9)$$

As the problem is of minimization, the fitness function values were ordered by decreasing values. The value of the fitness function was not used for the probability calculation because high values of the fitness function dominate the population, generating a premature convergence of GA.

To pick parents for crossover, two chromosomes were randomly selected from the mating pool. So, the probability of selecting a chromosome from the mating pool is given by $p_c = 1/N$. The crossover operation was carried out as proposed by Picek et al. (2013). These authors showed that a combination of several operators result in a better performance for the GA. In this study, the following operators were used: 1) one-point crossover; 2) two-point crossover; 3) uniform crossover; and 4) random operator. For each one of these operators, an occurrence probability of 0.02, 0.03, 0.05 and 0.9 was defined. The three first operators can generate a solution with a number of cities

with allocated schools different from p . These solutions are called inviable solutions. Figure 2 illustrates the generation of an inviable solution with the uniform operator. For this example, a chromosome was employed with length equal to 10 and $p=4$. With the crossover random mask, the offspring chromosome has five cities with allocated schools. To solve this problem, the second term of a new fitness function, Eq.(10), penalizes chromosomes with a number of cities with allocated schools different from p . In some results that will be presented in the next section, some cities must be considered fixed in the solution. In these cases, term 3 of Eq.(10) penalizes the solutions that do not have these cities.

$$F(X) = f(X) + |S - p| * \max(d_{ij}^{Man_gen}) + N * \max(d_{ij}^{Man_gen}) \quad (10)$$

In Eq.(10): $F(X)$ is the fitness function with type 1 and type 2 penalties; $f(X)$ is the original fitness function defined by Eq.(4); S is the number of cities with schools in the location solution; N is the number of fixed cities not present in a location solution.

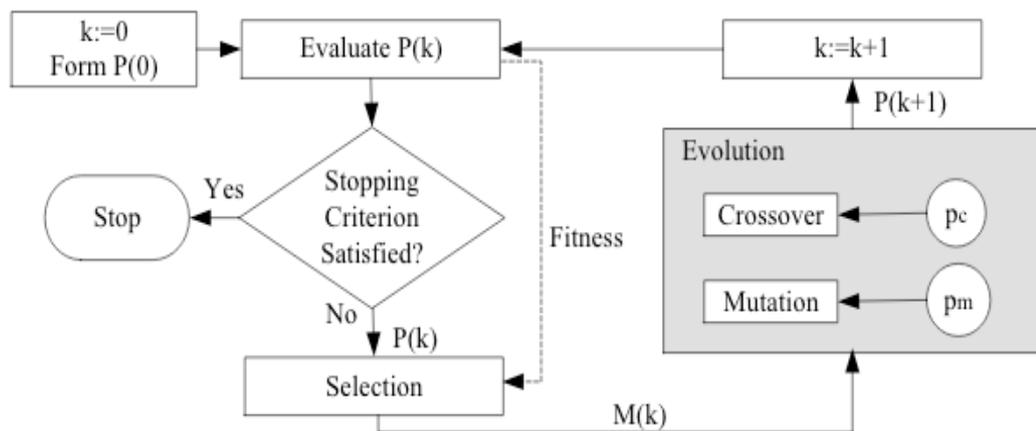


Figure 1 – Steps of Genetic Algorithm (After Chong and Zak, 2013). $P(k)$ – population at instant k ; p_c - crossover probability; p_m - mutation probability.

Parent 1	0	0	1	0	0	1	0	1	1	0
Parent 2	0	1	0	0	1	0	0	1	0	0
Crossover mask	0	1	1	1	0	1	0	0	1	0
Offspring	0	0	1	0	1	1	0	1	1	0

Figure 2 – Generating inviable solutions with the uniform operator. Only for this example, the chromosome length was equal to 10 and $p = 4$.

To implement the random operator, we adopted the following approach: first, the chromosome length was transformed from 62 to p . Each one of the p genes represents the decimal position of the 1's in the original chromosome. In the sequence, a random mask, of size p , like the one in the uniform crossover, is generated. With this mask, the

first offspring is generated. For generating the second offspring, this mask is inverted. Figure 3(a) illustrates this process. As shown, in the offspring there are genes duplicated. To solve this problem, from the set of decimal numbers of both parents that are not in one offspring, one chooses a random decimal number to change the duplicated value. Figure 3(b) illustrates this process for offspring 1.

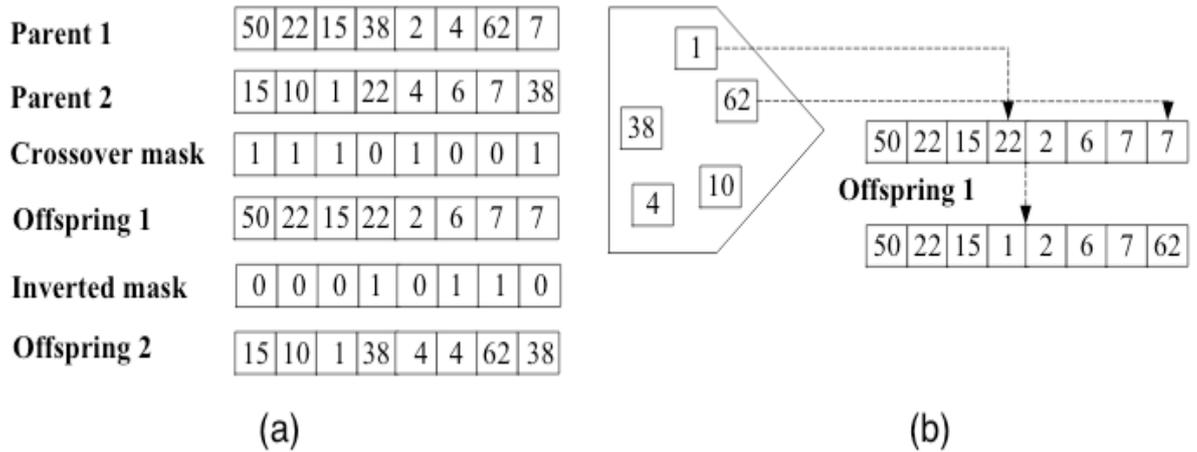


Figure 3 – Random operator used for implementing crossover: (a) generating offspring; (b) eliminating duplicated genes.

For mutation step, the reciprocal exchange mutation called a Twors mutation operator was used (Abdoun et al., 2012). This operator selects two genes randomly and then swaps its values. Evidently, no mutation takes place when the two selected genes have the same value. To prevent this situation, we have adapted the Twors mutation operator as follows. Initially, the genes of a chromosome are divided in two subsets: a) one subset composed of genes with values equal to 1, ; b) another subset composed of position genes with values equal to 0,. Subsequently, one gene is selected randomly from each subset, and with uniform probability and , respectively. Then, these two genes are swapped between and . The probability of a given chromosome suffers a mutation is 0.2.

The parameters used in this study for GA are described in the sequence. Population size = 300. Maximum number of iterations = 200*p. Number of individuals selected for crossover = 80% of population size. Number of individuals selected for mutation = 20% of population size. The simulation for this heuristic was performed at MATLAB R2014a on a ASUS Intel(R) Xeon(R) Processor @1.80-2.40 GHz computer with 8 GB of RAM and Windows 10.

6 | RESULTS OBTAINED WITH THE GA

The results section comprises three subsections. The first one shows results obtained with our GA that meet the REUNI requirements. The second section evaluates the existing campus locations of the federal system, in Amazonas state, comparing them with similar locations, obtained with our GA algorithm. Finally, the third section,

presents a proposal for expanding federal system to 10 and 15 campuses, preserving the current system location.

6.1 Location solutions that meet reuni criteria

In this study, location solutions with number of colleges varying from 10 to 50 were obtained for GA algorithm. Table 1 shows a reduced set of these results (only 13 values), for GA.

Number of colleges	Mean fluvial distance	Mean SSP	Mean UN HDI	Evaluation index	Meets REUNI requirement?	Maximum fluvial distance	Execution time (s)
10	272.90	21,938.40	0.578	21.08	No	1001	275.92
14	223.27	16,657.57	0.579	25.24	No	616	398.93
18	191.23	13,424.67	0.574	29.23	No	505	549.29
22	170.03	11,285.27	0.566	32.79	No	505	691.95
26	148.50	9,785.15	0.559	37.41	No	402	882.24
30	130.72	9,064.67	0.556	42.32	No	324	1,186.82
34	129.96	8,082.76	0.552	42.66	No	285	1,214.22
38	133.92	7,488.74	0.553	41.41	No	285	1,505.66
42	116.65	6,856.14	0.555	47.22	No	285	1,749.85
46	93.19	6,778.52	0.559	58.70	No	226	1,966.97
48	93.93	6,531.63	0.559	58.23	Yes	170	2,398.75
49	92.54	6,535.10	0.560	59.09	Yes	170	2,074.14
50	93.08	6,458.48	0.560	58.74	Yes	170	2,387.90

Table 1 – Results of locations with number of universities or colleges varying from 10 to 50, for GA algorithm.

The simulations use the generalized distance shown in Eq.(1). For these solutions, we calculated: the average fluvial distances between cities without a university or college to the nearest city with a university or college; the average SSP population and UN HDI of cities with an allocated high school; and the evaluation index. As shown in Table 1, for GA algorithm, the locations with 48, 49 and 50 campuses meet the REUNI requirement: the maximum distance of a city without a university or college to a city with a university or college is less than 200km. The conclusion is that to meet the REUNI requirements, a high number of universities or colleges must be constructed. This is impractical, not only due to the high cost of the solution, but also because it implies founding a university or college in small cities, with an insufficient SSP population.

6.2 Evaluation the existing campus locations of the federal system in amazonas state

This section compares the locations obtained with our GA algorithm with the current distribution of Federal University campuses. Section 2 shows that UFAM has 6 campuses in Amazonas State. Figure 4 shows a comparison of GA location with the current location of UFAM. In this figure, the urban and rural areas of the cities are shown. The blue color indicates that a city was chosen by both GA and UFAM. The gray color highlights cities selected only by UFAM. The green color indicates cities

selected only by GA. Table 2 shows the mean fluvial distance, mean SSP, mean UN HDI and evaluation index computed for these two locations.



Figure 4 – Location solutions performed by GA compared to current distribution of UFAM.

As shown in Figure 4, the GA algorithm selected the following cities: Manaus, Parintins, Tefé, Tabatinga, Eirunepé and Boca do Acre. The p-median modeling optimized by GA method prioritizes cities with a high number of inhabitants, such as Manaus, Parintins, Tefé and Tabatinga. Those cities have a population greater than 60,000 dwellers and cities with lower values of UN HDI, such as Eirunepé and Boca do Acre, 0.563 and 0.588, respectively. Differing from our approach, the location produced by UFAM prioritizes, in general, cities with a higher population.

Table 2 shows mean fluvial distance, mean SSP, mean UN HDI and evaluation index computed for these two locations.

According to Table 2, GA algorithm results in better value of the evaluation index than the UFAM location. The best value for the evaluation index corresponds to GA algorithm, 17.28. This behavior of the evaluation index is due to the lower value of mean fluvial distance obtained with our GA algorithm.

Distribution	Mean fluvial distance	Mean SSP	Mean UN HDI	Evaluation index
UFAM	536 64	38131 00	0 634	11 23
GA	328 50	36 538 33	0 634	17 28

Table 2 – Mean values of fluvial distance, SSP, UN HDI and evaluation index for locations of 6 campuses of UFAM and GA algorithm.

6.3 Expansion proposal for Federal campus

It is very expensive to reallocate existing schools already established in cities of the interior of Amazonas State. So, in this section, we propose an expansion of the current campuses of UFAM, preserving the current existing schools. For this system, we propose an expansion to 10 and 15 schools. Figures 5(a) and 5(b) show the predicted expansion of UFAM for 10 and 15 campuses, respectively. In these figures, the current existing schools are highlighted in color, and the new allocation schools are marked

green.

As shown in Figure 5(a), the p-median problem solved by the GA algorithm selected the following cities for developing 4 new schools: Pauini, Eirunepé, Guajará and Tefé. Our proposal of opening 4 new campuses for UFAM using the p-median modeling, presented in this study, prioritizes cities with a high number of inhabitants, such as Tefé that has a population greater than 60,000 inhabitants and cities with lower values of UN HDI, such as Pauini, Eirunepé and Guajará, 0.496, 0.563 and 0.532, respectively.

As shown in Figure 5(b) the p-median problem solved with the GA algorithm method selected the following cities for creating 9 new schools: Apuí, Boca do Acre, Canutama, Eirunepé, Guajará, Ipixuna, Itamarati, Pauini and Tefé. This expansion prioritizes cities with a high number of inhabitants, such as Tefé that has a population greater than 60,000 inhabitants and cities with lower UN HDI values, such as Boca do Acre, Canutama, Eirunepé, Guajará, Ipixuna, Itamarati, Pauini, 0.588, 0.530, 0.563, 0.532, 0.481, 0.477 and 0.496, respectively. The city of the Apuí has a relatively high UN HDI value of 0.637, and a lower population, around 20,648 inhabitants. The question is why was this city selected instead of the city of São Gabriel da Cachoeira? The answer is that the first city is farther away from large urban centers such as Manaus than the second one, 1,365km versus 1,001km of fluvial distance, respectively.

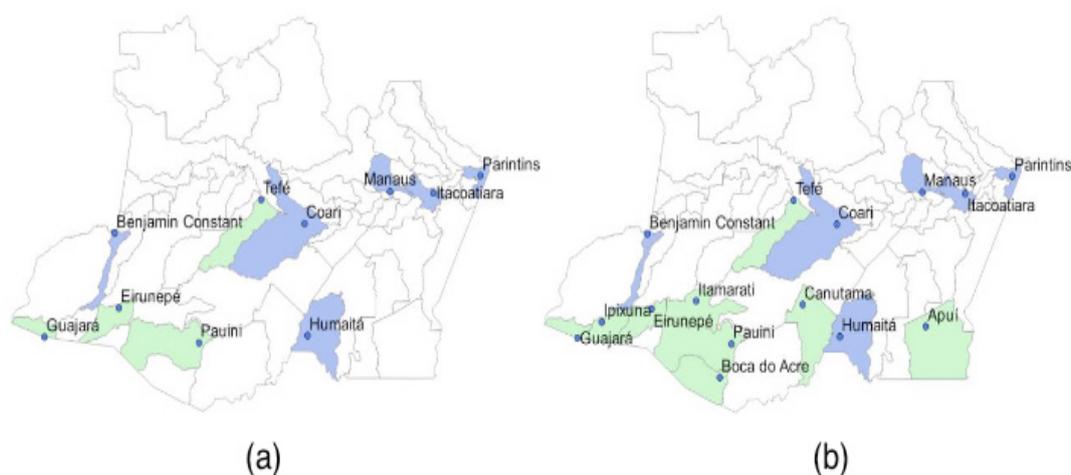


Figure 5 – Proposal for expanding the Federal University of Amazonas: (a) 10 campuses; (b) 15 campuses. Blue color: cities currently with institutions of higher learning. Green color: cities chosen for establishing new schools.

One can observe that the proposal for expanding UFAM, based on a multiparametric approach, selects cities mainly in the Southwest subregion of Amazonas state. In this region there are many cities with lower UN HDI values, with considerable fluvial distances between them and great population centers, such as Manaus, Parintins, etc. Thus, another contribution of this work is to emphasize the necessity of social development of these cities.

7 | CONCLUSION

This study proposed a new methodology for a multiparameter approach using the p-median. This was possible by introducing generalized Manhattan distance, a proximity measure, that, besides the geographic distance, allows for including other parameters relevant to the nature of the problem. In this study, the other parameters considered were the population of secondary pupils and the United Nations' Human Development Index. These parameters are normalized using median approach. The study proposes an optimization expression that allows direct and inverse dependence with the parameter.

This new multiparametric approach for the p-median problem was applied to a case study of university in Amazonas State, situated in the northern region of Brazil, with 62 cities. In this study, genetic algorithm heuristic was employed to solve the p-median location problem in 62 cities. To comply with REUNI requirements in Amazonas State, using the multiparametric approach, a minimum of 48 cities with colleges is required. This is an impractical number of schools, mainly due to the cost and because it does not justify establishing a university in a very small city.

In Amazonas State, the Federal University of Amazonas, an institution of higher learning, operating in 6 cities. To compare with these currently established schools, simulations using GA algorithm with the same number of schools were done. The locations obtained with the our GA algorithm result in lower mean distances between cities with schools and cities with no schools compared with the existing distribution of schools.

Finally, the main contribution of this study is to propose a guide for expanding the current Federal University, based on a multiparametric approach. In a future study, with the objective of finding a set of solutions in a search space, the Pareto-optimal method will be applied to solve the same problem.

REFERENCES

Abdoun, O., Abouchabaka, J., & Tajani, C. (2012). Analyzing the Performance of Mutation Operators to Solve the Travelling Salesman Problem. *International Journal of Emerging Sciences*, 2(1), 61-77.

Amaya, J., Peeters, D., Uribe, P., & Valenzuela, J. P. (2015). Optimization Modeling for Resource Allocation in the Chilean Public Education System. *International Regional Science*, 39(2), 155-176.

Brazilian Institute of Geography and Statistics – IBGE (2012). Union Official Journal, 1, 65 - 246. Secondary School Students Achievements.

Brazilian Institute of Geography and Statistics – IBGE (2010). XII Demographic Census.

Bussab, W. O., & Morettin, P. A. (2009). *Basic Statistics*. (6th ed.). Saraiva Publisher. (Section 3.1: Position Measures).

Chong, E. K. P & Zak, S. H. (2013). *An Introduction to Optimization*. (4th Ed.). Canada. John Wiley & Sons. (Chapter 14: Global Search Algorithms).

Federal University of Amazonas –UFAM (2006). Institutional Development Plan 2006-2015. http://www.proplan.ufam.edu.br/PDI_06a15.pdf. Accessed: 08.08.16.

Gac, I., Martínez, F., & Weintraub, A. (2009). A deterministic linear optimization model for allocating schools to zones. *Journal of the Operational Research Society*, 60, 895-905.

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press.

Kariv, O., & Hakimi, S. L. (1979). An algorithmic approach to network location problems. II: the p-medians. *Journal of Applied Mathematics*, 37, 539–560.

Menezes, R. F., & Pizzolato, N. D. (2014). Locating public schools in fast expanding areas: application of the capacitated p-median and maximal covering location models. *Pesquisa Operacional*, 34(2), 301-317.

Ministry of Education (2010). VIII National REUNI Workshop: Federal Universities - Consolidation and Expansion 2011-2020, Support for the National Education Plan. Brazil.

National Waterway Transportation Agency – ANTAQ (2013). Characterization of supply and demand in the inland waterways of the Amazon region passengers.

OECD report (2015). Education at a Glance 2015: OECD Indicators. *OECD Publishing, Paris*. <http://dx.doi.org/10.1787/eag-2015-en>. Accessed in 29.08.16.

Picek, S., Jakobovic, D., & Golub, M. (2013). On the recombination operator in the real-coded genetic algorithms, *IEEE Congress on Evolutionary Computation*, pp. 3103-3110.

Pizzolato, N. D., Barcelos, F. B., & Lorena, L. A. (2004). School location methodology in urban areas of developing countries. *International Transactions in Operational Research*, 11(6), 667–681.

Teixeira, J., Antunes, A., & Peeters, D. (2007). An Optimization-Based Study on the Redeployment of a Secondary School Network. *Environment and Planning B: Planning and Design*, 34(2), 296-315.

Xavier, C. M., Costa, M. G. F., & Costa Filho, C. F. F. (2016). Optimizing Universities Allocation with Integer Programming Using a New Distance Measure. In: 5th International Conference on Engineering Optimization (EngOpt), Iguassu Falls-Brazil. Proceedings of 5th EngOpt, 2016. v. 1. p. 1-10.

Xu, R., & Wunsch, D. C. (2008). *Clustering*. IEEE Press Series on Computational Intelligence (Section 2.2: Definition of Proximity Measures).

SOBRE O ORGANIZADOR

Ernane Rosa Martins - Doutorado em andamento em Ciência da Informação com ênfase em Sistemas, Tecnologias e Gestão da Informação, na Universidade Fernando Pessoa, em Porto/Portugal. Mestre em Engenharia de Produção e Sistemas, possui Pós-Graduação em Tecnologia em Gestão da Informação, Graduação em Ciência da Computação e Graduação em Sistemas de Informação. Professor de Informática no Instituto Federal de Educação, Ciência e Tecnologia de Goiás - IFG (Câmpus Luziânia) ministrando disciplinas nas áreas de Engenharia de Software, Desenvolvimento de Sistemas, Linguagens de Programação, Banco de Dados e Gestão em Tecnologia da Informação. Pesquisador do Núcleo de Inovação, Tecnologia e Educação (NITE), certificado pelo IFG no CNPq. ORCID: <https://orcid.org/0000-0002-1543-1108>

ÍNDICE REMISSIVO

B

Big data 76, 77

C

Computação 2, 5, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 65, 67, 92, 112, 114, 115, 116, 117, 119, 120, 124, 127, 168, 169

Computação vestível 67

Comunicação 39, 42, 68, 75, 148, 154

Conhecimento 52, 53, 58, 59, 60, 61, 62, 70, 76

D

Dispositivos 35

E

Ensino 34, 35, 37, 40, 41, 42, 43, 45, 46, 107, 125, 127

G

Gestão do conhecimento 63

I

Informação 39, 52, 53, 56, 57, 58, 61, 63, 76, 89, 90, 91, 148, 154, 168

Internet 5, 7, 21, 22, 43, 57, 58, 92, 105, 106, 107, 112, 113, 115, 132

Internet das coisas 5

M

Monitoramento 135

O

Organização do conhecimento 54

P

Programação 43, 168

R

Recuperação de dados 24

Redes 21, 43, 130, 131, 137, 141, 148, 153, 154

S

Sistemas de arquivos 24, 33

T

Tecnologia 57, 60, 75, 112, 143, 148, 149, 154, 168

Testes 26, 27, 28, 29, 30, 31, 32, 122

Agência Brasileira do ISBN
ISBN 978-85-7247-488-7

