

CONTEXTO

“Tempo é dinheiro” (FRANKLIN, 1748) famosa frase dita por Benjamin Franklin na metade do século 18 ainda ecoa na cabeça de muitos seres humanos.

Em busca de mais tempo as pessoas procuram então otimizar suas tarefas. Uma forma de otimizar as tarefas é a criação de automações. As automações buscam por uma melhor produtividade, redução de custos e maior tempo livre para se concentrar em outras tarefas que não podem ser automatizadas. Silva (2019) define que “... automação é um dos processos mais utilizados para a facilitação de inserção dos recursos tecnológicos. Através dessa tecnologia, são utilizadas ferramentas para soluções tecnológicas com o objetivo de otimizar e tornar simples os processos internos, além de diminuir custos operacionais.”

Um grande exemplo disso são as assistentes virtuais, como por exemplo a Alexa da Amazon, a Siri da Apple e o Google Home do Google. Esses assistentes virtuais são capazes de realizar diversas tarefas através de um simples comando de voz. Essa interação entre seres humanos e máquinas está cada vez mais presente nos sistemas, mas para que isso seja possível, é utilizado uma ferramenta chamada PLN (Processamento de Linguagem Natural). (RACKSPACE TECHNOLOGY, 2020)

Johnson (2021) define o PLN sendo um ramo dentro da Inteligência Artificial responsável em fazer com que as máquinas possam compreender a linguagem dos seres humanos, ou seja, o PLN funciona como um tradutor, permitindo assim que as tecnologias possam entender seus usuários, mesmo eles utilizando a linguagem natural.

O PLN também está presente em outras plataformas além das assistentes virtuais. Por exemplo, ele auxilia em sites de busca realizando interpretações entre o que o usuário digita com conteúdos de sites que poderão ser exibidos. Também está presente no auto-completar em plataformas de busca, onde sugestões automáticas são exibidas na tela no momento em que o usuário está digitando. Chatbots, que são utilizados por empresas para se comunicar com seus clientes, também fazem uso do PLN realizando a “tradução” do que o cliente deseja com possíveis soluções das quais as empresas podem oferecer. (TAKE BLIP, 2019).

Para que seja possível criar sistemas voltados a automações, é necessário levantar os requisitos que esse sistema irá possuir. Em engenharia de requisitos, a etapa responsável para o levantamento dessas informações é a elicitacão. Para

Thayer (1997), a elicitación de requisitos é o processo em que os clientes e usuários são questionados pela equipe de desenvolvimento a falarem o quê espera como funcionalidades no sistema que será desenvolvido. Nessa etapa de elicitación serão definidas as exigências, os recursos, os objetivos e as utilidades que o sistema deve cumprir.

Segundo Sommerville(2011, pág. 57):

Os requisitos de um sistema são as descrições do que o sistema deve fazer, os serviços que oferece e as restrições de seu funcionamento. Esses requisitos refletem as necessidades dos clientes para um sistema que serve a uma funcionalidade determinada, como controlar um dispositivo, colocar um pedido ou encontrar informações. O processo de descobrir, analisar, documentar e verificar esses serviços e restrições é chamado engenharia de requisitos.

A especificação de requisitos no desenvolvimento ágil pode ser feito por meio de histórias de usuário (User Stories). Através delas, o usuário utiliza de uma abordagem de escrever sobre os requisitos, tudo isso por meio de uma ou duas frases escritas na perspectiva de quem deseja o recurso/funcionalidade.

Para Cohn (2009, pág. 4), “uma história de usuário descreve a funcionalidade que será valiosa para um usuário ou comprador de um sistema ou software.”. Já Rehkopf (2020) define histórias de usuário como “uma explicação informal e geral sobre um recurso de software escrita a partir da perspectiva do usuário final. Seu objetivo é articular como um recurso de software pode gerar valor para o cliente.”

As técnicas de PLN também podem oferecer diversas vantagens para melhorar a qualidade das histórias de usuário. Segundo Raharjana, Siahaan e Fatichah (2021):

As técnicas de processamento de linguagem natural (PNL) oferecem vantagens potenciais para melhorar a qualidade das histórias de usuários. O PNL pode ser usado para analisar, extrair ou analisar os dados da história do usuário. Tem sido amplamente utilizado para ajudar no domínio da engenharia de software (por exemplo, gerenciamento de requisitos de software, extração de atores e ações no documento de requisitos, recurso de software extração, teste de software, etc.).

PROBLEMA

Cohn (2009) comenta que ao definir os requisitos de software a comunicação pode ser uma adversidade, pois aqueles que desejam um novo software devem se comunicar com quem irá desenvolvê-lo.

A respeito de histórias de usuário, Heath (2020) afirma que muitas vezes são genéricas, vagas na coleta e análise de requisitos. Segundo o autor, ele afirma que as maiores dificuldades encontradas são referentes a escopo, priorização e classificação das histórias no backlog, imprecisão sobre o que é problema e o que é solução, e por fim, histórias que escondem seu real valor.

Toda essa falta de clareza nas histórias de usuário podem influenciar caso seja utilizado PLN para a análise dessas histórias. Segundo artigo publicado no blog Take Blip em 2019, o autor comenta que um dos desafios do PLN é a adaptação da linguagem coloquial, que por muitas vezes possuem erros gramaticais ou até mesmo erros de digitação. Outro grande desafio citado pelo autor é fazer com que a solução possa compreender além das palavras, ou seja, quando duas pessoas conversam através da fala, muitas informações passam despercebidas, como tom de voz, respiração e olhares. Esses detalhes podem fazer toda a diferença de compreensão quando se compara apenas com uma conversa através de uma conversa por SMS ou chat.

Atualmente existem inúmeras soluções utilizadas para o PLN. Parker(2019) em seu artigo cita 12 ferramentas open source em diversas linguagens de programação, como por exemplo Python, Node e Java. Dentre as soluções citadas por Parker (2019), destaca-se a Natural Language Toolkit (NLTK) em Python, por ser a solução com mais recursos disponíveis, capaz de implementar todos os componentes de PLN e oferece suporte a vários idiomas.

Outra solução que se destaca é a OpenNLP em Java. É hospedada pela Apache Foundation, ou seja, é fácil integrá-la com outros serviços da Apache. Assim como a NLTK, oferece suporte a vários idiomas e cobre todos os componentes de PLN.

No entanto, ainda é necessário compará-las para que seja possível aferir qual tecnologia possui uma maior exatidão em seus processamentos e qual possui a melhor performance. Sendo assim, qual a solução mais adequada para auxiliar o PLN dentro do contexto de histórias de usuário?

Neste sentido, o objetivo central de desse trabalho é realizar um estudo sistemático de comparação entre pequenas soluções utilizando PLN para criação de histórias de usuário para os idiomas português e inglês, e por fim, serão avaliadas a qualidade das histórias de usuário criadas.

SOLUÇÃO PROPOSTA

Com base no cenário atual onde existem diversas tecnologias voltadas para PLN, se faz necessário uma análise comparativa entre essas tecnologias afim de definir qual ou quais tecnologias são mais adequadas para a validação de histórias de usuário.

Para que isso seja possível, serão selecionadas algumas tecnologias para que sejam previamente avaliadas em: documentação, linguagem de programação, conteúdo disponível na internet a respeito da tecnologia (sites, fóruns e *threads* em redes sociais) e o uso atual no mercado. Após feito esse levantamento de dados, as tecnologias que mais se destacarem serão selecionadas como objetos de estudo, onde será implementado protótipos voltados a validação de histórias de usuário. Esses protótipos serão avaliados levando em consideração critérios de qualidade em requisitos de software.

Wake (2003) cita em seu artigo que as características de uma boa história de usuário devem possuir 6 características: deve ser independente, negociável, valiosa, estimável, pequena e testável. Sendo assim, será desenvolvido uma análise através de uma tabela comparativa sob as potenciais tecnologias onde será aferido se as histórias avaliadas possuem as seis características citadas acima.

Também será avaliado nos protótipos a eficiência no processamento para os idiomas inglês e português e também a produtividade da tecnologia.

OBJETIVO GERAL

Desenvolver um estudo comparativo entre soluções de PLN com o propósito de avaliar qual ou quais tecnologias são mais adequadas para analisar critérios de qualidade em requisitos de software descritos como história de usuário.

OBJETIVOS ESPECÍFICOS

- Analisar e avaliar soluções atuais no mercado, comparando-as dentro dos critérios prestabelecidos (documentação, linguagem, conteúdo disponível na internet e uso no mercado);

- Implementar um protótipo para as soluções mais bem avaliadas onde serão voltadas a validação de histórias de usuário;
- Avaliar os protótipos desenvolvidos a nível de eficiência no processamento em inglês e português e também da produtividade ao utilizar a tecnologia.

METODOLOGIA

Segundo Brown (2006), define que a metodologia é a estrutura filosófica dentro da qual a pesquisa é conduzida ou a base sobre a qual a pesquisa se baseia. Já O'Leary (2004) descreve a metodologia como a estrutura que está associada a um conjunto particular de suposições paradigmáticas que usadas para conduzir nossa pesquisa.

Sendo assim, dado o contexto de metodologia, o estudo seguirá o modelo científico em camadas (Research Onion) de Saunders (2007), seguindo a forma transversal, indutiva e interpretativa. Seguirá um modelo multimétodo, com procedimento de pesquisa bibliográfica (GIL, 2010), estudo comparativo das tecnologias (FACHIN, 2001), design e prototipação (SOMMERVILLE, 2011) e Goal Question Metric (GQM) (BASILI, CALDIERA, ROMBACH, 1994).

Etapa	Atividades	Métodos	Resultados
Etapa 1 - Síntese da fundamentação teórica	- Sintetizar contexto histórico de processamento de linguagem natural, de histórias de usuário e de critérios de qualidade	Pesquisa bibliográfica (GIL, 2010)	Fundamentação teórica
Etapa 2 - Estudo comparativo	- Pesquisar as tecnologias mais utilizadas - Análise das tecnologias conforme os requisitos preestabelecidos - Definir as duas tecnologias mais promissoras	Estudo comparativo (FACHIN, 2001)	Análise comparativa de potenciais soluções
Etapa 3 - Prototipação	- Implementar protótipos das soluções A e B em português e inglês	- Design e prototipação (SOMMERVILLE, 2011)	Protótipos das tecnologias selecionadas
Etapa 4 - Avaliação comparativa	- Avaliar e comparar os resultados obtidos das tecnologias A e B	- GQM (BASILI et al., 1994)	Avaliação das tecnologias e tabela comparativa

REFERÊNCIAS BIBLIOGRÁFICAS

BARKER. D. **“12 open source tools for natural language processing”**, 2019.

Disponível em:

<https://opensource.com/article/19/3/natural-language-processing-tools> Acesso em 11 dez. 2021

BASILI, V. R.; CALDIERA, G.; ROMBACH, H. D. **“Goal Question Metric Paradigm”**.

In: MARCINIAK Encyclopedia of Software Engineering. [S.l.]: John Wiley & Sons, 1994.

BROWN R. B, **“Doing Your Dissertation in Business and Management: The Reality of Research and Writing”**, 2006. Sage Publications

COHN, M. **User stories applied for agile software development**, 13. ed.

Crawfordsville, Indiana. 2009. 263 p.

FRANKLIN. B. **“Advice to a Young Tradesman”**, 1748. Disponível em:

<https://founders.archives.gov/documents/Franklin/01-03-02-0130> Acesso em 10 dez. 2021

GIL, A. C. **“Como elaborar projetos de pesquisa”**. São Paulo: Atlas, 2010. ISBN 5ª edição.

HECK, P. KLABBERS, M. VAN EEKELEN, M. C. J. D. **“A software product certification model,”** Software Quality Journal, vol. 18, no. 1, pp. 37–55, 2010.

HEATH, F. **The trouble with user stories**. 2020, DZone. Disponível em:

<https://dzone.com/articles/the-trouble-with-user-stories-1> . Acesso em 04 dez. 2021

O’LEARY Z. **“The essential guide to doing research”**, 2004. Sage.

RAHARJANA, I. K, SIAHAAN. D e FATICHAH. C, **"User Stories and Natural Language Processing: A Systematic Literature Review,"** in IEEE Access, vol. 9, pp. 53811-53826, 2021, doi: 10.1109/ACCESS.2021.3070606.

RACKSPACE TECHNOLOGY, **"Dos chatbots à Alexa: a evolução do Processamento de Linguagem Natural"**, 2020. Disponível em: <https://www.rackspace.com/pt/solve/evolution-nlp> Acesso em 31 jan 2022.

REHKOPF. M. **"Histórias de usuários com exemplos e um template"**, 2020. Disponível em: <https://www.atlassian.com/br/agile/project-management/user-stories> . Acesso em 09 dez. 2021

SAUNDERS, M., LEWIS, P., & THORNHILL, A. **"Research Methods for Business Students"**, 2007, (6th ed.) London: Pearson.

SILVA, P. **"O que é automação e para que serve? Conversando com o CTO"**, 2019. Disponível em: <https://gobacklog.com/blog/o-que-e-automacao-e-para-que-serve/> . Acesso em 6 jan. 2022.

SOMMERVILLE, I. . **"Engenharia de software"**, 9. ed. Pearson. 2011. 529 p.

TAKE BLIP, **"Tudo sobre NLP: o que é processamento de linguagem natural e seus desafios na Inteligência Artificial"**, 2019. Disponível em: <https://www.take.net/blog/tecnologia/nlp-processamento-linguagem-natural/> . Acesso em 01 dez. 2021.

THAYER, R. H. e DORFMAN, M.; **"Introduction to Tutorial Software Requirements Engineering"** in Software Requirements Engineering, IEEE-CS Press, Second Edition, 1997, p.p. 1-2.

WAKE, B. **"INVEST in Good Stories, and SMART Tasks"**, 2003. Disponível em: <https://xp123.com/articles/invest-in-good-stories-and-smart-tasks/> Acesso em 11 jan 2022.