



**UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
UNIDADE ACADÊMICA DE SISTEMAS E COMPUTAÇÃO
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

GILVAN RAMOS DA CRUZ NETO

COMO PESSOAS DESCREVEM LUGARES?

CAMPINA GRANDE - PB

2019

GILVAN RAMOS DA CRUZ NETO

COMO PESSOAS DESCREVEM LUGARES?

Trabalho de Conclusão Curso apresentado ao Curso Bacharelado em Ciência da Computação do Centro de Engenharia Elétrica e Informática da Universidade Federal de Campina Grande, como requisito parcial para obtenção do título de Bacharel em Ciência da Computação.

Orientador: Professor Dr. Claudio Elízio Calazans Campelo.

CAMPINA GRANDE - PB

2019



C957c Cruz Neto, Gilvan Ramos da.
Como pessoas descrevem lugares? / Gilvan Ramos da Cruz Neto. - 2019.

12 f.

Orientador: Prof. Dr. Claudio Elízio Calazans Campelo.

Trabalho de Conclusão de Curso - Artigo (Curso de Bacharelado em Ciência da Computação) - Universidade Federal de Campina Grande; Centro de Engenharia Elétrica e Informática.

1. Descrição de lugares. 2. Inteligência artificial. 3. Localização - informações sobre. 4. Descrições espaciais. 5. Posicionamento - descrições. 6. Rota - descrições. 7. Dados geoespaciais. 8. Linguagem espacial verbal. 9. Processamento de linguagem natural. I. Campelo, Claudio Elízio Calazans. II. Título.

CDU:004(045)

Elaboração da Ficha Catalográfica:

Johnny Rodrigues Barbosa
Bibliotecário-Documentalista
CRB-15/626

GILVAN RAMOS DA CRUZ NETO

COMO PESSOAS DESCREVEM LUGARES?

Trabalho de Conclusão Curso apresentado ao Curso Bacharelado em Ciência da Computação do Centro de Engenharia Elétrica e Informática da Universidade Federal de Campina Grande, como requisito parcial para obtenção do título de Bacharel em Ciência da Computação.

BANCA EXAMINADORA:

**Professor Dr. Claudio Elízio Calazans Campelo
Orientador – UASC/CEEI/UFCG**

**Professor Dr. Herman Martins Gomes
Examinador – UASC/CEEI/UFCG**

**Professor Dr. Tiago Lima Massoni
Disciplina TCC – UASC/CEEI/UFCG**

Trabalho aprovado em: 25 de novembro 2019.

CAMPINA GRANDE - PB

Como pessoas descrevem lugares?

Trabalho de Conclusão de Curso

Gilvan Ramos da Cruz Neto (Aluno), Cláudio Campelo (Orientador)

Departamento de Sistemas e Computação
Universidade Federal de Campina Grande
Campina Grande, Paraíba - Brasil

ABSTRACT

The way we describe locations can be important in several situations, not only in warnings about occurrences of greater severity such as traffic accidents or natural disasters, but in other events in which communication is done verbally. Despite the considerable advancements in the fields of Artificial Intelligence, current Natural Language Processing models are not yet able to maintain a dialogue that involves non-trivial geographic awareness and spatial reasoning. In this work, an experiment was carried out to collect and analyze data referring to 570 place descriptions made by volunteers, in order to understand human behavior under different conditions. The results provided meaningful insights on the way people describe places in different situations they are exposed, such as known and unknown places. It also revealed some language resources that are employed under different circumstances and indicated that there may be correlations between features like gender/age and the language resources used. We believe this work is an important step towards a deeper understanding of spatial language and that the results obtained in this research are useful resources for the development of spatial-aware intelligent systems.

KEYWORDS

Place Descriptions, Dialogue, Artificial Intelligence, Response, Systems

RESUMO

A forma como descrevemos locais pode ser importante em várias situações, não só em avisos sobre ocorrências de maior gravidade, como acidentes de trânsito ou desastres naturais, mas em outros eventos em que a comunicação é feita verbalmente. Apesar dos avanços consideráveis nas áreas da Inteligência Artificial, os atuais modelos de Processamento de Linguagem Natural ainda não são capazes de manter um diálogo que envolva consciência geográfica não trivial e raciocínio espacial. Neste trabalho, foi realizado um experimento para coletar e analisar dados referentes

a 570 descrições de lugares feitas por voluntários, a fim de compreender o comportamento humano em diferentes condições. Os resultados forneceram insights significativos sobre a forma como as pessoas descrevem lugares em diferentes situações em que estão expostas, como lugares conhecidos e desconhecidos. Também revelou alguns recursos linguísticos que são empregados em diferentes circunstâncias e indicou que pode haver correlações entre características como sexo/idade e os recursos linguísticos utilizados. Acreditamos que este trabalho é um passo importante para uma compreensão mais profunda da linguagem espacial e que os resultados obtidos nesta pesquisa são recursos úteis para o desenvolvimento de sistemas inteligentes de reconhecimento espacial.

PALAVRAS-CHAVE

Descrições de Lugares, Diálogo, Inteligência Artificial, Resposta, Sistemas

1. INTRODUÇÃO

Em diversas circunstâncias do dia a dia informações precisas sobre localizações podem facilitar nossa vida bem como auxiliar profissionais em eventos que envolvam uma maior gravidade. Por exemplo, um turista que visita determinada cidade e deseja encontrar um local que lhe foi indicado previamente de forma vaga. Outro exemplo seria um serviço de emergência que recebe um pedido de socorro relacionado a um acidente onde o solicitante descreve o local do fato, ou ainda uma denúncia numa rede social sobre ocorrências de assaltos numa determinada localização.

Observa-se que as respostas geradas pela maioria dos sistemas que se comunicam de forma verbal com humanos não entregam um resultado que se aproxima ao de um diálogo entre seres humanos, retornando em grande parte dos casos respostas pouco coerentes ou até em algumas vezes totalmente incorretas e, em ferramentas que necessitam identificar locais, respostas corretas se tornam ainda mais necessárias para que o uso desses sistemas se tornem viáveis. Baseado nisso, entender como

humanos descrevem as localizações de lugares se torna um importante artefato para auxiliar na implementação de sistemas mais precisos que se comuniquem com humanos e que tenham por objetivo identificar locais descritos por pessoas.

Definimos as descrições de lugares em descrições de posicionamento e descrições de rota. As descrições baseadas em posicionamento definem onde está inserido um determinado local a partir do uso de termos como "*próximo de*", "*ao lado de*", entre outros. Por outro lado, as descrições baseadas em rota representam uma tentativa de indicar uma localização com base em algum trajeto que pode ser percorrido com o objetivo de se alcançar o local alvo da descrição. Um exemplo de descrição desta natureza seria "*...chegando pela rua do sol, vira-se à direita na farmácia da saúde e...*". As descrições podem ainda combinar estas duas categorias, como em "*...é na frente da Praça da República, logo após passar pelo viaduto José da Silva, do lado direito*".

Analisou-se neste trabalho como seres humanos descreveram localizações de lugares, estabelecendo-se como estudo de caso a cidade de Campina Grande-PB. Foram visualizados diversos aspectos relacionados às descrições, tais como a utilização de termos que descrevessem rotas ou posicionamentos, o uso de termos vagos, dentre outros pontos considerados importantes para o trabalho em questão.

Com base nesses aspectos, buscou-se responder algumas indagações relacionadas ao comportamento humano na descrição dos locais, expostas a seguir:

- Como nível de conhecimento sobre a cidade afeta nas descrições?
- Qual a influência da idade nas descrições?
- O sexo e o conhecimento sobre a cidade influenciam no tempo gasto nas descrições?
- Que tipo de descrição as pessoas mais utilizam? Descrições de posicionamento ou de rota?
- Quando descrevem lugares, as pessoas usam nomes alternativos para fazer referência a pontos específicos?
- Com que frequência pessoas usam pontos de referência nas descrições?

A pesquisa contou com a participação de 57 voluntários, que proveram descrições de 10 lugares diferentes onde estavam supostamente localizados (através de um ambiente simulado). Destas pessoas, cerca de 84% residiam em Campina Grande-PB no momento da aplicação da pesquisa e, em uma escala de 0-10, informaram que possuíam conhecimento mediano sobre locais da cidade. Dentre os resultados, foi verificado que a idade do indivíduo tem influência na quantidade de caracteres utilizados nas descrições. Também observou-se que os indivíduos descrevem, em grande parte das vezes, posicionamentos e não rotas que permitam chegar a um determinado local.

O restante do artigo está estruturado da seguinte maneira. A próxima seção introduz alguns conceitos relevantes para o entendimento do trabalho e discute os trabalhos relacionados. Em seguida, a Seção 3 descreve a metodologia utilizada para condução desta pesquisa. A Seção 4 apresenta os resultados obtidos e, por fim, a Seção 5 conclui o artigo e aponta para trabalhos futuros.

2. FUNDAMENTAÇÃO TEÓRICA

Nos últimos anos, o estudo de descrições espaciais através da linguagem natural ganhou grande relevância, principalmente para a inteligência artificial. Bons exemplos disso são assistentes virtuais como Siri, Cortana ou Alexa, atualmente muito conhecidas no mercado. Uma das razões subjacentes a essa relevância - embora não seja o único - é a massificação de dados georreferenciados, implícita ou explicitamente.

Em ambientes altamente interativos ou ainda em plataformas informais de mídia social, as pessoas fazem uso pesado de linguagem informal, inclusive quando se trata de locativo expressões. Esse uso inibe a capacidade de uso das tradicionais abordagens de processamento de linguagem natural para recuperar dados geoespaciais e informações essenciais de textos (LIU et. al, 2014).

VILLEGAS, et al. (2019) observaram que a extração automática de informações geoespaciais em tempo real terá um enorme impacto em tecnologias disruptivas, como assistentes inteligentes e mecanismos de pesquisa. Gerar modelos capazes de decodificar uma descrição e depois transformá-lo em endereços de dados estruturados pode se tornar a solução de problemas como: a identificação automática de frases que se referem a alguma entidade geoespacial; a manipulação de sinônimos referentes à mesma localização (ambiguidade); a caracterização taxonômica de expressões locativas; assim como a automação de processos complexos de interpretação de linguagem para a determinação de coordenadas geográficas em mapas e bancos de dados.

HAMZEI, et al. (2019) porém, destacam que as respostas geradas pelo homem em um diálogo humano-humano são diferentes das respostas recuperadas na interação humano-computador. Enquanto na pergunta humana-humana, a pessoa recebe respostas relevantes com informação contextual suficiente, as ferramentas atuais baseadas em computador não são capazes de fornecer respostas de qualidades similares. Segundo BENNET (2006), a ambiguidade e a imprecisão da linguagem natural são alguns dos maiores obstáculos na construção de ontologias robustas de uso geral. Por causa desses fenômenos é extremamente difícil fornecer informações precisas dos significados dos conceitos aceitáveis para uma ampla comunidade de usuários. O estudo do uso de termos vagos - propriedade inerente às linguagens naturais, que se manifesta na síntese de

sentenças imprecisas, que falham em transmitir as ideias originais pois os termos utilizados podem possuir um espaço contínuo de interpretações - em descrições se torna nesse cenário um importante auxiliar para a evolução dessas ferramentas.

De acordo com RICHTER, et al. (2012), a criação de sistemas automatizados para a interpretação de descrições de lugares continua sendo um desafio. Atualmente, não existem ferramentas de software capazes de compreender de forma coerente linguagem espacial verbal complexa ou responder outras questões. HAMZEI, et al. (2019) também corroboram com esse pensamento e explanam que ferramentas que fornecem respostas semelhantes às respostas geradas por humanos serão essenciais para sua utilização nos mais diversos tipos de sistema. Para isso, o comportamento de resposta humana deve ser investigado como um pré-requisito importante.

2.1 TRABALHOS RELACIONADOS

Apesar de termos na literatura uma considerável quantidade de trabalhos relacionados a identificação automática de informações geoespaciais a partir de dados textuais, o entendimento de como o ser humano descreve localizações de lugares foi explorado em menor escala, conquanto estudos relacionados a esse tema podem ser um importante facilitador na melhoria de ferramentas que identifiquem localizações.

VASARDANI, et al. (2016) estudaram a viabilidade de construir um gerador de representações computacionais de instâncias de lugares analisando um conjunto de propriedades com suporte cognitivo para a aplicabilidade destas propriedades estruturais na detecção de locais a partir de descrições constituídas de linguagem natural. Foi explorado um subconjunto de propriedades cognitivas para a análise da viabilidade do uso deste subconjunto ser usado por um gerador de representações computacionais de locais. Quatorze estudantes de pós-graduação da Universidade de Melbourne descreveram três locais cada - um local na área urbana de Melbourne, um local na zona rural e um local interno. O processo de coleta de dados resultou na descrição de 32 áreas. Os resultados indicaram que um possível construtor de local deveria tratar de propriedades que se destacavam não apenas no experimento cognitivo em questão, mas também relacionadas a conceitos básicos de lugar presentes em outras teorias cognitivas.

A análise e o uso de informações transmitidas por uma descrição de locais, tais como referências a lugares, relações espaciais e atividades humanas, que representa o locativo e semântico conhecimento do descritor sobre os locais referidos foi proposto por CHEN, et al. (2018) postulando-se que tais informações poderiam ser aproveitadas para criar um sistema de informações para modelar e utilizar recursos do conhecimento humano sobre determinado local. Os autores verificaram que as descrições de lugares ocorrem na comunicação cotidiana como

forma de transmitir informações espaciais sobre o lugar e que o conhecimento modelado em gráficos de locais estendidos podem ajudar a entender melhor as descrições humanas como entrada para serviços espaciais e, portanto, auxiliar a melhoria na interação humano-computador.

RICHTER, et al. (2012) exploraram um conjunto de descrições de lugares coletadas através de mecanismos de *crowdsourcing* dentro de um aplicativo para celular e introduziram um esquema geral para classificar descrições de locais de acordo com diferentes parâmetros característicos usando esse esquema para demonstrar existência de certos grupos de tipos predominantes de lugares presentes em descrições feitas através da comunicação humana. Observou-se que a maioria das pessoas pretendia descrever suas reais localizações, como solicitado. Muitas descrições continham informações de endereço (nível da rua ou do edifício), que geralmente são claramente identificáveis e úteis em um contexto urbano. No entanto, várias descrições pareciam ser destinadas a propósitos diferentes e eram direcionadas a outros destinatários, como amigos, pois compartilhavam referências vernaculares e certas abreviações.

Utilizando o MS-MACRO, um conjunto de dados de compreensão de máquina de propósito geral fornecido pela Microsoft, HAMZEI, et al. (2019) analisaram a relação entre as perguntas e respostas em termos de tipo, escala e proeminência de lugares referidos na base de dados. Os resultados mostram que as respostas geradas por humanos às perguntas seguem um padrão específico, partindo de valores menos conhecidos de tipo, escala e proeminência, e continuam até lugares bem conhecidos. Este estudo revelou ainda que o tipo, a escala e a proeminência dos lugares mencionados nas questões têm relação direta com a formação das respostas analisadas.

A exploração do nível de granularidade para a localização de um local descrito foi estudado por RICHTER, et al. (2013). O estudo foi focado em conhecer esse melhor nível de granularidade e seu suporte a resolução de descrições de locais, por exemplo, na recuperação de informações geográficas além da integração das relações espaciais nesse processo. Foi proposto e aplicado um procedimento mecanicista para determinar o nível de granularidade para um conjunto de descrições de lugares. A viabilidade do procedimento foi avaliada em uma comparação de descrições de locais e suas relações com a posição autorreferida das pessoas em um mapa. Os resultados mostraram que o procedimento geralmente oferecia boas respostas de acordo com os locais correspondentes do mapa.

Os trabalhos descritos nesta seção tiveram sua importância em contextos locais, onde a língua utilizada é de primordial importância para os resultados obtidos. Em nosso estudo, o enfoque na língua portuguesa foi um importante diferencial, além da análise de outros aspectos relacionados à descrição de localização de lugares que não foram explorados em

trabalhos disponíveis na literatura e que podem ser úteis na implantação e/ou evolução de sistemas que tratem de informações geoespaciais.

3. METODOLOGIA

Neste estudo contamos com a participação de 57 pessoas, que de forma anônima e voluntária, descreveram as localizações de 10 lugares previamente repassados e que estão situados em Campina Grande-PB. Foi definido um tempo limite de 2 minutos para a descrição de cada um dos locais. Durante a entrevista, foram coletados os seguintes dados do participante: sexo, faixa de idade, informação se o participante reside ou já residiu na cidade de Campina Grande, além de uma autoavaliação em que se pedia numa escala entre 0-10 o quanto o indivíduo conhecia locais em Campina Grande.

De acordo com WILSON, et al. (2002), na comunicação verbal os indivíduos conseguem transmitir uma gama muito ampla de significados, apesar do fato de não haver uma camada básica de informações identificável de forma independente para o ouvinte captar. Como resultado, a comunicação verbal pode atingir um grau de explicitação não disponível na comunicação não-verbal. Baseado nisso, foi recomendado aos entrevistados que descrevessem os locais de maneira informal, como se essa descrição estivesse sendo feita verbalmente a um amigo que lhe buscaria no lugar mostrado.

Os locais escolhidos para as descrições possuem distintas características. Buscou-se disponibilizar lugares com diferentes níveis de dificuldade para que os participantes da pesquisa descrevessem sua localização. No Quadro 1 estão descritas as métricas utilizadas para a escolha dos locais utilizados na coleta de dados.

Quadro 1: Tipos dos locais utilizados na pesquisa, de acordo com suas características

| | |
|--------|--|
| Tipo 1 | Lugares localizados em Ruas/Avenidas bem conhecidas da cidade. |
| Tipo 2 | Locais centrais e com muitos pontos de referência ao seu redor. |
| Tipo 3 | Pontos mais afastados de locais conhecidos da cidade. |
| Tipo 4 | Lugares situados em ruas pouco conhecidas e com poucos pontos de referência. |

Para a coleta dos dados, foi desenvolvida uma ferramenta web com uma interface que disponibilizou aos entrevistados os

lugares a serem descritos a partir do Street View do Google¹, além do controle do tempo gasto em cada descrição, conforme o que é apresentado na Figura 1. A interatividade com o Street View possibilitou aos entrevistados visualizarem os locais de diversas formas, bem como circularem por ruas próximas às localidades escolhidas para as descrições.

A partir da integração da aplicação com a API do Street View, foi possível remover a sobreposição textual do Street View que indica o endereço do local associado, bem como o link para abrir o local no Google Maps². Isto proporcionou uma maior aproximação da realidade pois fez com que os participantes da pesquisa só mencionassem os nomes do logradouro (e.g., rua, avenida, travessa) que realmente conheciam, ou em casos onde tiveram capacidade de identificar essa informação no próprio local, através de placas.

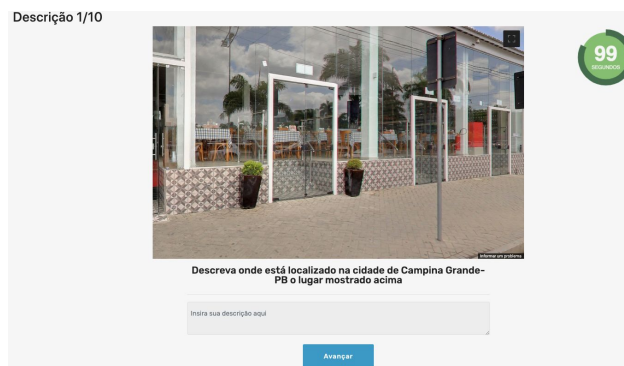


Figura 1: Interface da ferramenta desenvolvida para coleta das descrições dos locais

O trabalho em questão propôs analisar dados observando aspectos considerados importantes e que podem interferir nas descrições como a idade, a autodeclaração de conhecimento sobre o locais da cidade, o comportamento dos entrevistados no que se diz respeito às características dos elementos utilizados nas descrições como por exemplo o uso dos pontos de referência, o uso de termos alternativos e de termos que descrevem posicionamentos e rotas, visando buscar respostas que auxiliem na melhoria de sistemas que tratem da identificação automática de locais.

Para a identificação dos termos que descrevem rotas ou posicionamentos, termos vagos, termos vernaculares e termos históricos foi adotado o uso de vocabulários, estes definidos de forma manual, a partir de uma análise prévia das descrições coletadas na pesquisa.

O Processamento de Linguagem Natural (do inglês, Natural Language Processing - NLP) é de grande valia para

¹<https://developers.google.com/maps/documentation/javascript/streetview>

²<https://maps.google.com/maps>

verificar em que medida o texto está escrito de forma apropriada. De posse de uma base de descrições textuais, as mesmas precisaram ser devidamente tratadas e, para isto, foram executados procedimentos relacionados à normalização, comuns em pré-processamentos de linguagens naturais. Durante a normalização, foram abrangidas tratativas como *tokenização*, modificação de letras maiúsculas para minúsculas, remoção de caracteres especiais, dentre outras.

Para realização das análises, foi necessário marcar os pontos de referência utilizados pelos voluntários nas descrições textuais. Se executada manualmente, esta seria uma tarefa com um custo muito elevado, inviabilizando a pesquisa e comprometendo sua reprodutibilidade. Portanto, foram utilizadas técnicas de aprendizagem de máquina para realizar a anotação automática. Para este fim, foi utilizada uma biblioteca *Open Source* de processamento de linguagem natural, o RASA NLU³. Esta biblioteca pode ser utilizada para realizar classificação de intenção e extração de entidades⁴.

Para processar a extração dos pontos de referência, um modelo de extração foi treinado utilizando exemplos de descrições gerados manualmente, coletadas durante a pesquisa. A partir de uma descrição como "Ao lado do açude velho" era esperado que o modelo retornasse o ponto de referência "açude velho". O RASA, vale salientar, pode ser treinado para executar esta tarefa a partir de uma lista resumida de exemplos.

Após algumas iterações de treinamentos e testes, foi gerado um modelo provido de *precision* e *recall* 0.9, resultados estes obtidos através da comparação com os valores analisados manualmente de uma amostra de cerca de 10% dos dados. A *precision* indica a relação entre as previsões positivas realizadas corretamente e todas as previsões positivas, enquanto o *recall* demonstra a relação entre as previsões positivas realizadas corretamente e todas as previsões que realmente são positivas. A partir destes resultados o modelo pôde ser considerado satisfatório e os pontos de referência foram então coletados e anotados.

4. RESULTADOS E DISCUSSÕES

Neste trabalho, a caracterização dos participantes da pesquisa está apresentada na Tabela 1. Pode-se observar que a maioria dos entrevistados é do sexo masculino, na faixa de idade entre 18-24 anos, residentes em Campina Grande e declarando um moderado conhecimento de locais da cidade.

Tabela 1: Caracterização dos participantes da pesquisa considerando sexo, idade, residência em

| Campina Grande e autoavaliação do conhecimento da cidade. | | | |
|--|----------|-----------------------|---------------------------------|
| Variáveis | n | Frequência (%) | Frequência Acumulada (%) |
| <i>Sexo</i> | | | |
| Masculino | 49 | 85,96 | 85,96 |
| Feminino | 8 | 14,04 | 100,00 |
| <i>Idade (anos)</i> | | | |
| 18 a 24 | 41 | 71,93 | 71,93 |
| 24 a 30 | 10 | 17,55 | 89,48 |
| 30 a 40 | 3 | 5,26 | 94,74 |
| Acima de 40 | 3 | 5,26 | 100,00 |
| <i>Residência em Campina Grande</i> | | | |
| Sim | 48 | 84,21 | 84,21 |
| Não | 9 | 15,79 | 100,00 |
| <i>Auto Avaliação (Conhecimento da Cidade)</i> | | | |
| 0 a 3 | 9 | 15,79 | 15,79 |
| 4 a 7 | 38 | 66,67 | 82,46 |
| 8 a 10 | 10 | 17,54 | 100,00 |

4.1 O conhecimento da cidade e sua interferência nas descrições

Durante a pesquisa foi disponibilizado aos participantes a indicação de uma autoavaliação quanto ao conhecimento da cidade numa escala de valores compreendida entre 0-10, sendo 0 desconhecimento total de lugares na cidade e 10 conhecimento amplo de locais na cidade. A partir da análise dos dados, observou-se que há semelhança na quantidade de caracteres utilizados nas descrições considerando o atributo autoavaliação de conhecimento de locais da cidade. Em média, os entrevistados que declararam conhecer poucos locais da cidade utilizaram 65 caracteres. Os que declararam conhecer de forma mediana utilizaram 59 caracteres, enquanto os que autoavaliaram conhecer muitos locais fizeram o uso de 62 caracteres.

Observando ainda os valores apresentados na Figura 2, pode-se verificar uma variação de elevada magnitude nos dados referentes às pessoas que declararam amplo conhecimento de locais da cidade. Tendência diferente foi observada para pessoas

³ <https://rasa.com/docs/rasa/nlu/about/>

⁴ <https://blog.rasa.com/rasa-nlu-in-depth-part-2-entity-recognition/>

que declararam pouco conhecimento, pois verificou-se baixa variação nas respostas e médias semelhantes às demais classes das autoavaliações. Este fato indica que autoavaliação de conhecimento do local é uma variável que deve ser melhor estudada e compreendida em trabalhos posteriores.

Não foram verificadas correlações significativas entre a autoavaliação e outros critérios utilizados para a análise das descrições tais como o tempo gasto, o uso de pontos de referência, uso termos que descrevem rotas ou posicionamentos ou a utilização dos termos vagos.

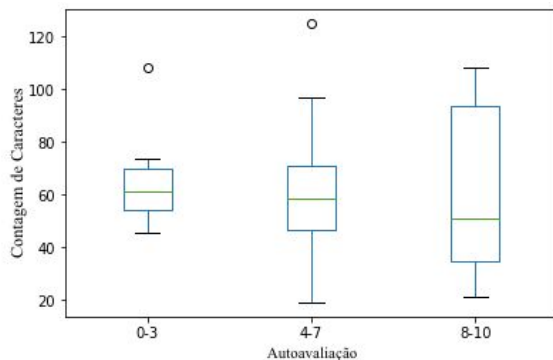


Figura 2: Medidas de posição da quantidade de caracteres utilizados nas descrições de lugares de acordo com a autoavaliação de conhecimento da cidade de Campina Grande-PB.

4.2 Influência da idade nas descrições

Considerando o tamanho da resposta das descrições dos lugares em função da idade do entrevistado, pode-se observar que pessoas com idades mais elevadas tendem a descrever os locais se utilizando de mais caracteres, conforme pode ser observado através das medidas de posição apresentadas na Figura 3. Este fato também é corroborado através da correlação de Pearson entre a idade e a quantidade de caracteres utilizados que foi de magnitude moderada, com valor de 0,50.

Pessoas na faixa de idade compreendida entre 18-24 anos utilizaram em média 54 caracteres em suas descrições; para a faixa entre 24-30 anos essa média foi de 74 caracteres; os entrevistados com idades de 30-40 anos fizeram o uso de 92 caracteres, em média; por fim, os entrevistados com idades acima de 40 anos apresentaram em média 76 caracteres para a descrição dos lugares propostos nesta pesquisa. Este fato indica que a idade do usuário se torna uma métrica importante a ser considerada no uso de sistemas que necessitam de informações repassadas por seres humanos.

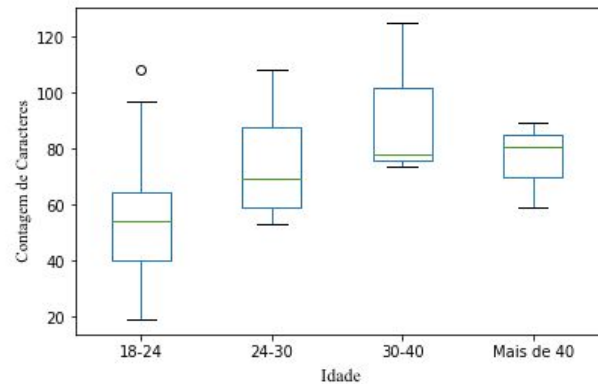


Figura 3: Medidas de posição da quantidade de caracteres utilizados nas descrições de lugares por faixa de idade.

4.3 Relação entre o gênero, a autoavaliação de conhecimento da cidade e o tempo gasto nas descrições

Ao se avaliar o tempo gasto pelos participantes nas descrições dos locais, observou-se que pessoas do sexo feminino gastaram, em média, 12 minutos para descreverem os 10 lugares propostos na pesquisa, enquanto participantes do sexo masculino fizeram o uso de 9 minutos e 45 segundos, em média, para a descrição dessa mesma quantidade de locais. Este fato deve estar relacionado à quantidade média de caracteres utilizados por homens e mulheres nas descrições dos locais. Enquanto os homens utilizaram em média 58 caracteres, as mulheres fizeram o uso de 76 caracteres, isto indica que a diferença de tempo está associada à intenção de prover mais detalhes, e não à dificuldade em prover descrições.

Levando em conta a variável de autoavaliação de conhecimento dos locais na cidade de Campina Grande, o tempo gasto nas descrições foi, em média, de 10 minutos e 20 segundos para pessoas que declararam conhecer a cidade pouco (faixa de 0-3) ou moderadamente (faixa de 4-7). Para pessoas que informaram um elevado conhecimento (faixa de 8-10), o tempo gasto foi de 8 minutos e 34 segundos, em média.

Em 6% das descrições os participantes utilizaram integralmente o tempo limite de 2 minutos definido na pesquisa. Destes participantes, 73% residiam em Campina Grande-PB e 84% avaliaram conhecer locais na cidade de forma moderada.

4.4 Descrições de posicionamento e de rota

A compreensão de descrições de rotas e posicionamento geradas por humanos têm sido fundamentais para pesquisas que visam o entendimento da cognição humana de informações geoespaciais (ZHANG et al., 2012). Baseado nisto, foi analisado neste estudo o comportamento humano relacionado ao tipo de descrição, estas divididas em descrições de rota e descrições de posicionamento.

As descrições de rota geradas por humanos são informações textuais que especificam caminhos possíveis entre

dois pontos. Elas contêm sequências de nomes de estradas, pontos de referência, pontos de decisão e ações a serem tomadas durante o trajeto de uma origem a um destino. Descrições de posicionamento especificam onde estão inseridos determinados locais através da utilização de termos que informem sua localização como "ao lado de", "próximo de", "em frente ao", entre outros.

A partir da análise da Figura 4 pode-se constatar a utilização em uma maior escala de termos que descrevem posicionamento e o uso em menor frequência de termos que descrevem rotas, e isso é recorrente em casos que os participantes avaliaram conhecer pouco, moderadamente ou muitos locais na cidade de Campina Grande. Nas Tabelas 2 e 3, estão exibidos alguns dos termos que foram utilizados pelos participantes da pesquisa para a descrição de posicionamento e rotas, respectivamente .

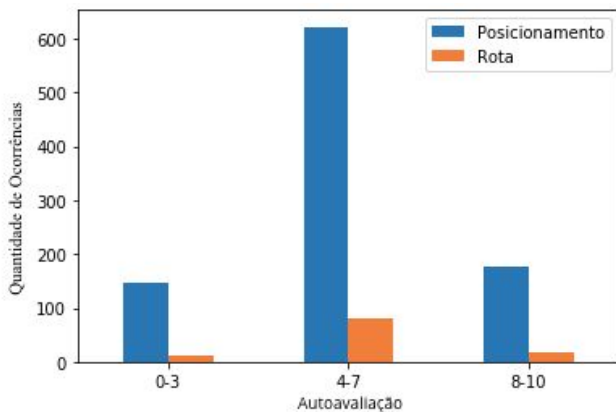


Figura 4: Quantidade de termos relacionados a descrições de posicionamento e rota encontrados na pesquisa de acordo com a autoavaliação de conhecimento da cidade de Campina Grande-PB.

Ainda com relação ao uso de termos que descrevem posicionamentos e rotas, foi verificado que 78% das descrições continham termos que descrevem posicionamentos, enquanto em 10% das descrições os participantes da pesquisa se utilizaram de termos que especificam rotas. Em 12% das descrições não foi feito o uso de termos que descrevessem posicionamentos ou rotas, o que torna a maioria dessas descrições pouco minuciosas.

Tabela 2: Principais termos utilizados pelos participantes da pesquisa que descrevem posicionamento.

| Termo | Quantidade de Ocorrências |
|--|---------------------------|
| <i>Termos que descrevem posicionamento</i> | |
| "Frente" | 273 |

| | |
|-----------|-----|
| "Próximo" | 118 |
| "Lado" | 106 |
| "Perto" | 83 |
| "Vizinho" | 14 |

Tabela 3: Principais termos utilizados pelos participantes da pesquisa que descrevem rota.

| Termo | Quantidade de Ocorrências |
|----------------------------------|---------------------------|
| <i>Termos que descrevem rota</i> | |
| "Indo" | 16 |
| "Sentido" | 16 |
| "Desce" | 9 |
| "Direção" | 8 |
| "Vindo" | 5 |

4.5 Uso de nomes alternativos para referenciar pontos específicos

Uma maneira adicional de se referir a um lugar é usando termos vernaculares, que geralmente acontece quando as pessoas se comunicam se referindo a uma região ou grupo cultural específico. Nesta pesquisa também foi verificado o uso destes termos, e se constatou sua utilização com certa regularidade. Dentre esses termos pode-se destacar "Praça dos Hippies", "Praça da Morgação", "Igreja dos Crentes", "Praça da Ternura", "Praça dos Pombos" e "Praça do Assalto".

Além dos termos vernaculares, a utilização de termos vagos também foi observado na pesquisa e seu uso pode ser visto com mais detalhes na Tabela 4. Os termos históricos, isto é, expressões que descrevem nomes de antigos locais da cidade, também foram utilizados nas descrições, tais como "Hotel Ouro Branco" e "Hotel Onigrat" cujo nome atual é "Nord Hotel".

Estes resultados corroboram com a necessidade de que sistemas que trabalham com mapeamento de locais devam ser capazes de utilizarem técnicas para a identificação correta dos dados relacionados a esses e também a termos associados ao que é chamado de conceitos espaciais vagos.

Tabela 4: Termos vagos utilizados pelos participantes da pesquisa.

| Termo | Quantidade de Ocorrências |
|---------------------|---------------------------|
| <i>Termos Vagos</i> | |

| | |
|-----------|----|
| "Perto" | 83 |
| "Pequeno" | 5 |
| "Grande" | 4 |
| "Baixo" | 1 |

4.6 Uso de pontos de referência nas descrições

Ao se analisar o uso dos pontos de referência, também definidos na literatura como *POI (Point of interest)* foi observado que pessoas se utilizam em grande parte dos casos destes pontos para descreverem locais. Em nossa pesquisa, 92% das descrições continham ao menos um ponto de referência. Os dados mostram que a frequência no uso dos pontos de referência tem relação com a idade dos entrevistados, pois obtivemos uma correlação moderada de 0,50 entre essas duas variáveis. Essa relação estatística nos mostra que pessoas em faixas de idade mais elevadas tendem a utilizar mais pontos de referência em suas descrições. Não foi observado, vale salientar, correlação significativa entre a frequência no uso de pontos de referência e a variável de autoavaliação de conhecimento de locais na cidade de Campina Grande - PB.

A partir dos dados da *Google Places API*⁵, pôde-se obter outras informações referentes aos locais utilizados nas descrições, como dados geométricos e o atributo *Rating*, que define a classificação de um local com base em avaliações agregadas de usuários. A utilização do atributo que retorna a quantidade de usuários que avaliaram certo local também foi avaliada como importante para pesquisa pois pode auxiliar no entendimento de quanto uma região é ou não conhecida.

As análises mostraram uma forte associação entre o número de usuários que avaliaram locais próximos aos pontos descritos e o uso de pontos de referência nas descrições. A correlação de 0,70 mostra que em regiões com locais com uma maior quantidade de avaliação dos usuários da Google, as localizações dos lugares próximos a estes que foram propostos na pesquisa eram descritas se utilizando de uma maior quantidade de pontos de referência, enquanto os locais em regiões onde a quantidade de avaliações era inferior, o uso dos pontos de referência foi feito em menor frequência.

5. CONCLUSÃO E TRABALHOS FUTUROS

Indivíduos tendem a utilizar na maioria dos casos termos que descrevem posicionamento e fazem, com menor frequência, o uso de termos relacionados a descrição de rotas. O uso dos termos que descrevem rotas foi observado em maior escala em locais situados em ruas e avenidas movimentadas, porém dispostas de poucos pontos de referência conhecidos, como, por exemplo, pequenos estabelecimentos comerciais. Por outro lado, a utilização de termos que descrevem posicionamentos foi feita em maior quantidade em lugares localizados em ruas e avenidas movimentadas, mas que possuíam pontos de referência bem conhecidos, como shoppings e pontos turísticos da cidade.

Pessoas do sexo feminino se utilizam de mais caracteres e gastam mais tempo nas descrições que indivíduos do sexo masculino. Além disso as mulheres fazem o uso de mais pontos de referência que os homens nas descrições. Verificando ainda separadamente as descrições de homens e de mulheres, observou-se que em cerca de 4% das descrições feitas por homens, foi informado o nome correto da rua onde o lugar está localizado, enquanto nas descrições feitas por mulheres foi indicado corretamente o nome da rua em 10% dos casos. Esses dados podem ser uma métrica importante para se concluir que mulheres descrevem locais de forma mais detalhada que os homens.

Com relação ao uso dos termos vagos, os resultados médios obtidos mostraram que homens se utilizam mais de termos vagos que mulheres. Avaliando a variável idade, pessoas na faixa de 30-40 anos se utilizam em maior frequência desses termos. Ao se analisar a variável de autoavaliação conhecimento de locais da cidade, foi verificado que pessoas que declararam conhecer muitos locais fizeram o uso em maior frequência dos termos vagos.

Em grande parte dos casos as pessoas fazem o uso de pontos de referência para descreverem os locais. O uso desses pontos de referência está moderadamente relacionado a idade do indivíduo, isto indica que pessoas de idades mais elevadas tendem a utilizar pontos de referência com mais frequência. Também foi verificado que o uso destes pontos de referência é altamente relacionado a região onde o local está inserido, pois em regiões mais conhecidas os locais inseridos nelas eram descritos com o maior uso de pontos de referência.

Como trabalhos futuros, pretende-se replicar esse experimento com uma maior quantidade de participantes, variando o tempo disponibilizado para cada descrição, de forma a avaliar como os indivíduos reagem a diferentes níveis de pressão. Aplicar o experimento para que participantes descrevam locais em diferentes cidades também se torna um elemento importante na obtenção de novos resultados que possam auxiliar na busca de

⁵<https://developers.google.com/places/web-service/search#nearby-search-and-text-search-responses>

soluções para problemas relacionados a sistemas que identifiquem posições geográficas.

6. REFERÊNCIAS

Bennett, B. A theory of vague adjectives grounded in relevant observables. In Patrick Doherty, John Mylopoulos, and Christopher A. Welty, editors, *Proceedings of the Tenth International Conference on Principles of Knowledge Representation and Reasoning*, pages 36–45. AAAI Press, 2006.

Chen, H.; Vasardani, M.; Winter, S.; Tomko, M. A Graph Database Model for Knowledge Extracted from Place Descriptions. *ISPRS Int. J. Geo-Inf.* 2018,7, 221.

Hamzei, E.; Winter, S.; Tomko, M. Initial Analysis of Simple Where-Questions and Human-Generated Answers, University of Melbourne, Parkville, Victoria, Australia, 2019.

Liu, F.; Vasardani, M.; Baldwin, T. Automatic Identification of Locative Expressions from Social Media Text: A Comparative Analysis. In *Proceedings of the 4th International Workshop on Location and the Web*, Shanghai, China, 3 November 2014; Ahlers, D., Wilde, E., Martins, B., Eds.; ACM: New York, NY, USA, 2014; pp. 9–16.

Richter, D.; Winter, S.; Richter, K.F.; Stirling, L. How people describe their place: Identifying predominant types of place descriptions. In *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information*, Redondo Beach, CA, USA, 6 November 2012; pp. 30–37.

Richter, D.; Winter, S.; Richter, K.F.; Stirling, L. Granularity of locations referred to by place descriptions. *Comput. Environ. Urban Syst.* 2013,41, 88–99.

Vasardani, M.; Tomko, M.; & Winter, S. (2016). The cognitive aspect of place properties. *International Conference on GIScience Short Paper Proceedings*,1, 1. doi: 10.21433/b3116928w12f

Villegas, A.M.; Siordia, O.S.; Bobadilla, E.A; Aguilar, C.A; Acosta, A. Extracción automática de referencias geoespaciales en discurso libre usando técnicas de procesamiento de lenguaje natural y teoría de la accesibilidad. *Procesamiento del Lenguaje Natural*, 63:143 – 146, 2019.

Wilson, D.; Sperber, D.; Dan, S. Relevance theory. In *Handbook of Pragmatics*. Blackwell, 2002.

Zhang, X.; Baojun, Q.; Mitra, P.; Xu, S.; Klippel, A.; MacEachren, A. 2012. Disambiguating Road Names in Text Route Descriptions Using Exact-All-Hop Shortest Path Algorithm. In *Proceedings of the 20th European Conference on Artificial Intelligence (ECAI'12)*. IOS Press, Amsterdam, The Netherlands,876–881. <https://doi.org/10.3233/978-1-61499-098-7-876> Session 2A: Collection Access and Indexing JCDL'18, June 3-7, 2018, Fort Worth, TX, USA 88.