

**CENTRO UNIVERSITÁRIO DE JOÃO PESSOA - UNIPÊ  
PRÓ-REITORIA ACADÊMICA - PROAC  
CURSO DE BACHARELADO EM CIÊNCIA DA  
COMPUTAÇÃO**

**NITAI CHARAN ÁLVARES PEREIRA**

**CONSULTAS UTILIZANDO LINGUAGEM NATURAL NO  
SERVIDOR DE BUSCA DISTRIBUÍDO ELASTICSEARCH**

**JOÃO PESSOA - PB  
2019**

# **CONSULTAS UTILIZANDO LINGUAGEM NATURAL NO SERVIDOR DE BUSCA DISTRIBUÍDO ELASTICSEARCH**

Monografia apresentada ao Curso de Bacharelado em Ciência da Computação do Centro Universitário de João Pessoa - UNIPÊ, como pré-requisito para a obtenção do grau de Bacharel em Ciência da Computação, sob orientação do Prof. MSc. Fábio Falcão da França

**JOÃO PESSOA - PB  
2019**

**NITAI CHARAN ÁLVARES PEREIRA**

# CONSULTAS UTILIZANDO LINGUAGEM NATURAL NO SERVIDOR DE BUSCA DISTRIBUÍDO ELASTICSEARCH

Monografia apresentada ao curso de Bacharelado em Ciência da Computação do Centro Universitário de João Pessoa - UNIPÊ como pré-requisito para obtenção do grau em Bacharelado em Ciência da Computação, apreciada pela Banca examinadora composta pelos seguintes membros:

Aprovada em                    /                    / 2019.

## BANCA EXAMINADORA

---

Prof. Dr. Nome do Professor A  
UNIPÊ

---

Prof. Dr. Nome do Professor B  
UNIPÊ

---

Prof. Dr. Nome do Professor C  
UNIPÊ

*À minha família, por sua capacidade de acreditar e investir em mim. Mãe, seu cuidado e dedicação foi o que me deram, em alguns momentos, a esperança para seguir. Vovô e Vovó, a presença de vocês significou segurança e certeza de que não estou sozinha nessa caminhada. A minha tia Rosemares (in memoria) por toda a oportunidade e ensinamentos que me deu para seguir com meus sonhos.*

## AGRADECIMENTOS

Gostaria de agradecer minha avó Carminha (in memorian) e primo Luiz Felipe Abreu (in memorian).

Gostaria de agradecer aos meu pais Mathura Pati e Subala Das por sempre prezar e batalhar pela educação dos seu filhos, sem medir esforços para que todos pudesse se graduar e obter a formação acadêmica necessária.

Meus agradecimentos também a minhas irmã Jahnavi Caran por me apoiar, dar suporte e aconselhar na minha trajetória acadêmica. Meus sinceros agradecimentos e minha eterna gratidão, irmã.

Dedico este trabalho ao meu professor orientador MSc. Fábio Falcão de França por me acolher e aconselhar passando seus conhecimentos sem pestanejos e me aceitando como último orientando antes de sua jornada em busca do título de doutorado. Dedico também ao professor MSc. Hugo Vieira Lucena de Souza por todos seus conselhos, ensinamentos e correções que foram possível para conclusão deste título. Também dedico este trabalho e ao professor DSc. Erick France Meira de Souza por sua amizade, suporte e preciosos conselhos dados.

A todos os professores e pessoas não citadas diretamente neste texto, mas que participaram e contribuíram de alguma forma com a minha formação, meu muito obrigado.

## RESUMO

O transtorno do espectro autista consiste em uma condição neurológica permanente que se manifesta na infância comprometendo a capacidade de comunicação, interação social, aprendizagem e adaptação ao meio do indivíduo portador.

O uso de tecnologias assistivas na área educacional tem potencializado o ensino e o desenvolvimento de portadores do espectro, auxiliando no processo de inclusão escolar e na adaptação dos modos de aquisição de conhecimentos,

porém essas ferramentas apesar de apresentarem soluções eficazes muitas vezes não possuem um processo de inclusão realmente eficiente e condizente com a realidade do usuário final.

Muitas das aplicações existentes direcionadas ao público de portadores do espectro autista possuem como propósito auxiliar na comunicação e na aprendizagem da língua portuguesa ou de atividades do cotidiano, satisfazendo apenas campos específicos e havendo a real necessidade de mecanismos que auxiliem em outras áreas.

Portanto o presente trabalho tem o intuito de gerar uma aplicação de apoio ao ensino dos numerais de 1 a 9 para crianças portadoras do transtorno do espectro autista, utilizando um processo criativo baseado no usuário

, o Design Thinking, cujo objetivo principal é a resolução de problemas com foco primordial no ser humano, levando em consideração a realidade do público alvo, satisfazendo as suas necessidades e respeitando as suas limitações.

Através de ciclo de etapas baseados nesta metodologia foi possível realizar o desenvolvimento da solução proposta, cuja última etapa consiste na validação da solução, através da aplicação da mesma em campo por crianças portadoras do espectro.

**Palavras-chave:** <Reconhecimento Facial. Envelhecimento Facial. Celebidades. Redes Neurais>.

1. O que é o transtorno do espectro autista
2. Tecnologia ajuda espectro auxiliando na inclusão nas escolas
3. apresentar problemas das ferramentas
4. ferramentas apenas auxiliam em outras áreas
5. Falar sobre a aplicação de apoio
6. Falar sobre Design Thinking
7. Falar sobre o desenvolvimento através das etapas propostas pela metodologia

## Construção do Resumo

- Indicativo sobre o trabalho e não sobre o conteúdo do trabalho
- Entre 150 e 500 palavras
- Não usar abreviações, símbolos, formulas, equações ou diagramas a não ser que seja estritamente indispensáveis
- Não usar citações
- Não faça mais de um paragrafo
- Não crie tópicos
- Sem frases negativas
- Após o título do trabalho e nome dos autores
- Primeira frase indica o tema do trabalho
- Informar os objetivos geral e específicos, metodologia, resultados
- 3 a 5 palavras chaves que represente a essência do trabalho separadas por pontos

## Partes do Resumo

- Introdução  
Objetivos e relevância do trabalho do trabalho.
- Metodologia  
Elementos essenciais do método de pesquisa
- Resultados  
Principais resultados ou de maior destaque que pode ser representativo da pesquisa
- Discussão
- Conclusão  
Contribuição e limites do trabalho

## Normas ABNT resumo

- Frases afirmativas, curtas, escritas em voz ativa e na terceira pessoa do singular
- Texto claro, conciso, seguir uma ordem lógica e sem completamente fiel ao trabalho
- Deve mencionar o tema, objetivo, métodos, resultados e conclusões
- Deve indicar, no fim, palavras chaves

## ABSTRACT

Facial appearance may change as age progresses, aging the human face for facial recognition is still a challenging problem, and one of the challenges faced in it is to achieve temporal invariance. This work proposes a technological solution that is believed to contribute to the solution of this problem, will be used for the proposal the celebrity data set called CACD (Cross-Age Celebrity Dataset), the dataset contains 163,446 images of 2,000 celebrities collected from the Internet, where it is possible to visualize images of celebrities along their careers, being they male or female, in the age range between 16 and 62 years. Images are collected from search engines using celebrity name and year (2004-2013) as keywords. In this dataset were manually removed noisy images of celebrities with a rating of less than or equal to five. This image bank was designed primarily for cross-age face recognition and recovery. The choice for the dataset was based on it being by far the largest dataset of age-matched faces publicly available. This work will present the application of Siamese Neural Networks to treat the comparative characteristics extracted from these images, giving the result of the greater proximity between them. A binary tree structure is used for dimensionality reduction. A significant improvement in accuracy can be observed in this method.

**Key-words:** <Facial recognition. Facial Aging. Celebrities. Neural networks. >



## LISTA DE FIGURAS

## LISTA DE TABELAS

## LISTA DE ABREVIATURAS

## Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>14</b>
1.1	JUSTIFICATIVA . . . . .	14
1.2	OBJETIVOS . . . . .	15
1.2.1	Objetivo Geral . . . . .	15
1.2.2	Objetivos Específicos . . . . .	15
1.3	METODOLOGIA . . . . .	15
1.4	ORGANIZAÇÃO DO TRABALHO . . . . .	16
<b>2</b>	<b>CONCEITOS GERAIS E REVISÃO DA LITERATURA</b>	<b>17</b>
<b>3</b>	<b>METODOLOGIA</b>	<b>18</b>
<b>4</b>	<b>APRESENTAÇÃO E ANÁLISE DOS RESULTADOS</b>	<b>19</b>
<b>5</b>	<b>CONCLUSÕES E TRABALHOS FUTUROS</b>	<b>20</b>
	<b>REFERÊNCIAS</b>	<b>20</b>
	<b>ANEXO A - Códigos utilizados no desenvolvimento do trabalho</b>	<b>22</b>

# 1 INTRODUÇÃO

A barreira do aprendizado para a interação entre homem e sistemas computacionais até os dias atuais ainda é um obstáculo a ser conquistado devido a necessidade de tempo e interesse por parte dos usuário. (BARBOSA; SILVA, 2010) afirmam que usuários que se dispõem a aprender novos sistemas interativos com características únicas e distintas precisam dispor de tempo e interesse para posteriormente ser capazes de usufruir e usar as funcionalidades deste sistema.

Consequentemente, pessoas que não possuem hábito de utilização da internet e sem muito conhecimento técnico em manusear softwares de acesso podem deixar de participar de eventos, oportunidades e movimentos que utilizam como meio a internet. Pessoas com estas características podem se ausentar da participação no controle sobre a Administração Pública atuando na verificação, acompanhamento e fiscalização da regularidade de gastos públicos devido a se depararem com barreiras até mesmo no momento do recolhimento dos dados públicos que utilizam como canal mediador a internet.

No que se refere a fiscalização, com a promulgação da Constituição de 1988, o Brasil se classificou se como um Estado Democrático de Direito onde uma de suas características, para este tipo de estado, é a participação popular no controle sobre a Administração Pública. Neste sentido, o cidadão tem o poder de acompanhar e fiscalizar a regularidade dos atos governamentais juntamente com os órgãos institucionais legalmente criados para esta finalidade (ARRUDA; TALES, 2010).

Tendo em vista que em 2000 foi promulgada a Lei de Responsabilidade Fiscal, que entre muito outros pontos, no artigo 48, definiu que as prestações de contas e outros instrumentos de transparência da gestão fiscal pública devem ter ampla divulgação, inclusive em meios eletrônicos de acesso público. Porém, todo benefício gerado a partir das regras impostas por esta lei, podem não ser devidamente aproveitadas por cidadãos que não possuem conhecimento prévio na utilização da internet.

Surgindo assim a pergunta central que este trabalho se propõe a intervir: como facilitar o acesso às informações públicas com o propósito que pessoas, sem conhecimento técnico em sistemas computacionais, possam auxiliar no processo de fiscalização de gastos governamentais, ajudando assim a diminuir os índices de corrupção no país?

## 1.1 JUSTIFICATIVA

A partir da promulgação da Lei de Acesso à Informação, Lei nº 12.527/2011, foi garantido o direito constitucional de acesso às informações públicas, possibilitando que qualquer pessoa física ou jurídica, sem a necessidade de apresentar motivos, receba informações públicas de órgãos e entidades. Desta forma, surgem novos mecanismos os quais possibilitam a participação do cidadão na fiscalização de gastos públicos e combate à corrupção.

A melhoria do acesso à informação pública e a criação de regras que permitem a disseminação de informações produzidas pelo governo reduzem os abusos que podem ser cometidos (STIGLITZ, 2002). Porém, tais dados estão mais disponíveis em sistemas computacionais mediante a internet, utilizando-se muitas vezes de ferramentas carentes

de parametrização ou disponibilização dos dados em grandes planilhas, dificultando assim o entendimento por parte de não-especialistas em informática.

Na tentativa de melhorar a divulgação dos dados públicos, o Ministério da Transparência em conjunto com a Controladoria-Geral da União, em 2004, criou o site Portal da Transparência do Governo Federal que possibilita o acesso livre, no qual o cidadão pode encontrar informações sobre como o dinheiro público é utilizado, além de se informar sobre assuntos relacionados à gestão pública do Brasil. Ainda assim, o problema de acesso por pessoas sem muitos conhecimentos técnicos sobre internet e como utilizá-la continua a existir.

Como forma de intervenção, este trabalho sugere a utilização de linguagem natural para busca e manipulação dos dados em bases de dados disponibilizadas pelo Governo Federal. “O processamento de linguagem natural permite ao computador compreender e reagir a declarações e comandos de voz realizados em uma linguagem natural” (STAIRS; REYNOLDS, 2006, p. 508). Para tal, este trabalho propõe a utilização do serviço de API de reconhecimento vocal Language Understanding Intelligent Service (LUIS) que aplica inteligência de aprendizado personalizado de máquina a um texto de linguagem natural e extrai informações relevantes a futuras aplicações.

Para isso, este trabalho também propõe a utilização da ferramenta Elasticsearch que, entre muitas funcionalidades, possui recursos e tecnologias que permitem realizar consultas através de índices em grandes volumes de dados em tempo real (GORMLEY; TONG, 2015).

## **1.2 OBJETIVOS**

### **1.2.1 Objetivo Geral**

Desenvolver uma solução de software que utiliza Graphical User Interface (GUI) para aumentar a usabilidade a pessoas leigas. Recebe como entrada uma solicitação de consulta a dados públicos, em linguagem natural, e é respondida com ou sem os dados solicitados.

### **1.2.2 Objetivos Específicos**

Realizar um estudo acerca das tecnologias utilizadas para a criação de um sistema de interação descrevendo suas características e possibilidades possíveis de utilização para se atingir o objetivo geral.

Gerar protótipos de baixa fidelidade para os testes de usabilidade com os requisitos impostas na concepção do sistema.

Avaliar a efetividade da aplicação com a realização de um estudo de caso uso solicitando a atores escolhidos de forma aleatória tentando buscar dados sobre gastos públicos sem e com a solução de software posposto.

## **1.3 METODOLOGIA**

“Método é o conjunto das atividades sistemáticas e racionais que, com maior segurança e economia, permite alcançar o objetivo de produzir conhecimentos válidos e

verdadeiros, traçando o caminho a ser seguido, detectando erros e auxiliando as decisões” (LAKATOS; MARCONE, p. 79, 2019). As autoras descrevem também que a ciência caracteriza-se pela utilização de métodos científicos e que estes não são de uso exclusivo pela ciência. Porém, não é possível a ciência estar apartado do emprego de métodos científicos (LAKATOS; MARCONE, 2019).

Com o objetivo de facilitar análise e avaliações futuras, este trabalho é dividido sobre a características dos tipos de pesquisas que foram aplicada do processo de construção deste trabalho.

Este trabalho se classifica como pesquisa exploratória pois entre outros fatores é tentado proporcionar maior familiaridade com as ferramentas utilizadas pela solução de software proposta. Realizando para isso, levantamento bibliográficos com o propósito de explicar e expor possibilidades proporcionadas por estas ferramentas. Como (GIL, p. 2018) explana que o levantamento bibliográfico utilizada pelas pesquisas exploratórias é uma das maneiras utilizadas para a coleta de dados relevantes.

Este trabalho também se classifica como uma pesquisa aplicada devido ao propósito ser uma possível abordagem ao problema identificado da dificuldade de acesso às informações públicas em que pessoas, sem conhecimento técnico em sistemas computacionais, possam posteriormente auxiliar no processo de fiscalização de gastos governamentais. (GIL, p. 25 2018) “pesquisa aplicada, abrange estudos elaborados com a finalidade de resolver problemas identificados no âmbito das sociedades em que os pesquisadores vivem”.

Com o intuito a avaliações futuras sobre a qualidade dos resultados mostrados neste trabalho, foi adotado a análise e interpretação dos dados expostos de forma qualitativa. Devido a descrição dos resultados serem em formas verbais e não em termos numéricos como proposto pelas pesquisas quantitativa. (GIL, p. 39 2018) “Nas pesquisas quantitativas os resultados são apresentadas em termos numéricos e, nas qualitativas, mediante descrições verbais”.

Os procedimentos adotados na análise, interpretação e coleta dos dados exposto neste trabalho são realizados de forma bibliográfica e experimental. A forma bibliográfica se dá pela referenciação de dados através de citações de materiais já publicados. Enquanto que a experimental se dá pela adoção de testes de caso de uso utilizado para demonstrar a eficácia do software proposto para intervenção ao problema.

## 1.4 ORGANIZAÇÃO DO TRABALHO

Após esse capítulo introdutório, o conteúdo deste trabalho organiza-se da seguinte forma:

- No capítulo 2 irá conter um estudo a respeito da Application Programming Interface (API) da ferramenta utilizada Language Understanding Intelligent Service (LUIS).
- O capítulo 3 tratará sobre o estudo da ferramenta Elasticsearch.

- O capítulo 4 apresentará a metodologia baseada no design thinking, utilizada para criação do projeto, as fases e artefatos gerados que apoiam no desenvolvimento da aplicação produzida, assim como os resultados obtidos na validação.

## **2 CONCEITOS GERAIS E REVISÃO DA LITERATURA**



### **3 METODOLOGIA**

## **4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS**

## **5 CONCLUSÕES E TRABALHOS FUTUROS**

## REFERÊNCIAS

## **ANEXO A – Códigos utilizados no desenvolvimento do trabalho**