



Licenciatura em Engenharia Informática

Projeto de Engenharia Informática em Contexto Empresarial

2023/2024

Ricardo Fernando Valadas Alves

## **Aplicação Web com Chatbot para o ISPGaya**

Relatório de projeto orientado pelo Engenheiro Justino Marco Ronda Lourenço e apresentado  
à Escola Superior de Ciência e Tecnologia

Junho de 2024

## **Agradecimentos**

Gostaria de expressar a minha gratidão a todos que contribuíram para a realização deste projeto. Agradeço à instituição pela oportunidade e pelo suporte prestado ao longo da licenciatura. Agradeço também ao meu orientador, Engenheiro Justino Lourenço pelo valioso apoio e orientação durante todo o processo.

## **Sumário executivo**

O presente projeto, desenvolvido no âmbito da Licenciatura em Engenharia Informática, visa a criação de uma aplicação web com um chatbot que responde a perguntas dos utilizadores com base nas informações recolhidas do site do ISPGaya. Utilizando técnicas de web scraping, foram extraídos dados relevantes das páginas do site, que foram então armazenados num ficheiro JSON. A aplicação foi desenvolvida utilizando tecnologias como HTML, CSS e JavaScript, e o chatbot foi programado para identificar palavras-chave nas perguntas dos utilizadores e fornecer respostas apropriadas, baseando-se nos dados armazenados.

O projeto foi motivado pela necessidade de melhorar a acessibilidade e a experiência do utilizador ao procurar informações sobre o ISPGaya. Com o aumento da procura por soluções rápidas e eficientes para obter informações, a implementação de um chatbot oferece uma forma interativa e imediata de responder às dúvidas dos utilizadores, contribuindo para uma melhor comunicação e satisfação.

A metodologia do projeto incluiu a definição dos requisitos funcionais e não-funcionais, a realização de web scraping para recolher dados específicos das páginas do site do ISPGaya, o armazenamento desses dados num ficheiro estruturado (JSON), e a implementação de um chatbot que utiliza esses dados para responder a perguntas. Foram realizados testes para validar a funcionalidade e a precisão das respostas do chatbot.

O projeto alcançou os seus objetivos principais, demonstrando a viabilidade de integrar um chatbot baseado em web scraping como uma ferramenta útil para melhorar o acesso à informação e a experiência do utilizador em ambientes educacionais.

## **Abstract**

*This project, developed as part of the Bachelor's degree in Computer Engineering, aims to create a web application with a chatbot that responds to user questions based on information collected from the ISPGaya website. Using web scraping techniques, relevant data was extracted from the website's pages and stored in a JSON file. The application was developed using web technologies such as HTML, CSS and JavaScript, and the chatbot was programmed to identify keywords in user questions and provide appropriate responses based on the stored data.*

*The project was motivated by the need to improve accessibility and user experience when searching for information about ISPGaya. With the increasing demand for quick and efficient information retrieval solutions, implementing a chatbot provides an interactive and immediate way to answer user queries, contributing to better communication and satisfaction.*

*The project's methodology included defining functional and non-functional requirements, performing web scraping to collect specific data from ISPGaya's website pages, storing this data in a structured format (JSON), and implementing a chatbot that uses this data to respond to questions. Tests were conducted to validate the chatbot's functionality and response accuracy.*

*The project achieved its main objectives, demonstrating the feasibility of integrating a web scraping-based chatbot as a useful tool to improve information access and user experience in educational environments.*

## **Abreviaturas**

HTML – HyperText Markup Language

CSS – Cascading Style Sheets

JS – JavaScript

JSON – JavaScript Object Notation

API – Application Programming Interface

URL – Uniform Resource Locator

HTTP – HyperText Transfer Protocol

HTTPS – HyperText Transfer Protocol Secure

DOM – Document Object Model

AI – Artificial Intelligence

NLP – Natural Language Processing

## Glossário

**HTML (HyperText Markup Language):** Linguagem padrão para criar páginas e aplicações web.

**CSS (Cascading Style Sheets):** Linguagem de estilo usada para descrever a apresentação de documentos escritos em HTML ou XML.

**JavaScript (JS):** Linguagem de programação que permite a criação de conteúdos dinâmicos em páginas web.

**JSON (JavaScript Object Notation):** Formato leve de intercâmbio de dados, fácil para leitura e escrita por humanos e máquinas.

**API (Application Programming Interface):** Conjunto de rotinas e protocolos que permitem a comunicação entre diferentes sistemas de software.

**URL (Uniform Resource Locator):** Endereço que identifica de maneira única um recurso na web.

**HTTP (HyperText Transfer Protocol):** Protocolo de comunicação usado para transferência de dados.

**HTTPS (HyperText Transfer Protocol Secure):** Versão segura do HTTP, que utiliza criptografia para aumentar a segurança das transferências de dados.

**DOM (Document Object Model):** Modelo de objeto do documento que representa a estrutura lógica de documentos HTML e XML.

**AI (Artificial Intelligence):** Inteligência artificial, ramo da ciência da computação que trata da criação de sistemas capazes de realizar tarefas que normalmente requerem inteligência humana.

**NLP (Natural Language Processing):** Processamento de linguagem natural, área da inteligência artificial que se preocupa com as interações entre computadores e linguagens humanas.

## Índice

Agradecimentos .....	2
Sumário executivo .....	3
Abstract .....	4
Abreviaturas .....	5
Glossário .....	6
Introdução .....	8
Enquadramento .....	8
Objetivos .....	8
Estrutura do relatório .....	9
Projetos similares e associados .....	10
Soluções existentes .....	10
Análise comparativa.....	11
Bases teóricas.....	12
Requisitos.....	13
Requisitos Funcionais .....	13
Requisitos não-funcionais.....	13
Descrição do trabalho .....	14
Testes de validação .....	16
Cronograma.....	18
Meios previstos e meios necessários.....	19
Meios Humanos .....	19
Meios materiais .....	20
Problemas e decisões .....	21
Problemas Detectados .....	21
Decisões Tomadas .....	21
Conclusões .....	22
Apêndices.....	<b>Erro! Indicador não definido.</b>

## **Introdução**

### **Enquadramento**

Este projeto visa a criação de uma aplicação web que integra um chatbot capaz de responder a perguntas feitas por utilizadores com base nas informações retiradas do site do ISPGaya. A necessidade de melhorar a experiência de acesso à informação levou ao desenvolvimento desta solução, utilizando técnicas de web scraping para recolher dados e tecnologias web para disponibilizá-los aos utilizadores.

### **Objetivos**

O principal objetivo deste projeto é criar uma ferramenta de chatbot que possa responder eficazmente a perguntas sobre a instituição do ISPGaya, utilizando dados extraídos automaticamente do seu website. Os objetivos específicos incluem:

- Desenvolver um sistema de web scraping para extrair dados relevantes do site do ISPGaya;
- Armazenar os dados extraídos num ficheiro JSON;
- Implementar um chatbot capaz de ler os dados JSON e fornecer respostas baseadas em palavras-chave identificadas nas perguntas dos utilizadores;
- Criar uma interface web amigável utilizando HTML, CSS e JavaScript.



## Estrutura do relatório

Este relatório está organizado da seguinte forma:

- **Introdução:** Apresenta o enquadramento do projeto, objetivos e organização do relatório;
- **Projetos similares e associados:** Analisa soluções tecnológicas semelhantes disponíveis no mercado;
- **Bases teóricas:** Revisão da literatura que fundamenta o projeto;
- **Requisitos:** Definição dos requisitos técnicos e funcionais do sistema;
- **Descrição do trabalho:** Detalha o desenvolvimento e implementação da aplicação;
- **Testes de validação:** Descreve os casos de teste utilizados para validar o sistema;
- **Cronograma:** Compara o cronograma previsto com o cumprido, justificando desfasamentos;
- **Meios previstos e meios necessários:** Descreve os recursos humanos e materiais utilizados;
- **Problemas e decisões:** Identifica problemas enfrentados e decisões tomadas;
- **Conclusão:** Apresenta a conclusão do projeto e sugestões para trabalhos futuros.

## Projetos similares e associados

Esta secção analisa soluções tecnológicas similares disponíveis no mercado e projetos desenvolvidos por outras instituições que têm como objetivo melhorar a interação com utilizadores através de chatbots e sistemas de resposta automática.

### Soluções existentes

Existem várias soluções no mercado que utilizam chatbots para melhorar a interação com os utilizadores em websites. Por exemplo:

- **Chatbot da Universidade de Stanford:** A universidade de stanford implementou um chatbot para auxiliar estudantes e visitantes com informações sobre cursos, horários de aulas, eventos e outros serviços da universidade. Utiliza processamento de linguagem natural avançado para entender e responder a perguntas dos utilizadores de forma contextualizada.
- **Chatbot da Uber:** A Uber utiliza um chatbot nos seus canais de suporte ao cliente para responder a perguntas comuns sobre viagens, pagamentos e promoções. Integrado com APIs avançadas de processamento de linguagem natural para oferecer suporte automatizado aos utilizadores.
- **Chatbot da Sephora:** A Sephora, uma empresa de cosméticos e produtos de beleza, implementou um chatbot no seu website e aplicações móveis para oferecer recomendações personalizadas de produtos, dicas de maquilhagem e responder a perguntas frequentes sobre os seus produtos e serviços. Utiliza processamento de linguagem natural para entender as preferências dos clientes e recomendar produtos de acordo com as suas necessidades e preferências.

## Análise comparativa

Comparando estas soluções com o projeto desenvolvido:

- **Funcionalidade:** O chatbot desenvolvido para o ISPGaya é específico para a instituição, utilizando dados extraídos diretamente do seu website, enquanto outros chatbots podem ter um escopo mais geral ou adaptável a diferentes contextos.
- **Tecnologia Utilizada:** Muitas soluções no mercado utilizam APIs de processamento de linguagem natural (NPL) avançadas, como Dialogflow do Google, enquanto o projeto desenvolvido utiliza um arquivo JSON e uma abordagem baseada em palavras-chave.

## **Bases teóricas**

A fundamentação teórica deste projeto envolve conceitos de web scraping, armazenamento de dados em formato JSON, desenvolvimento web e implementação de chatbots.

### **Web Scraping**

Web scraping é uma técnica utilizada para extrair informações de websites. Utiliza-se bibliotecas como BeautifulSoup em Python para navegar e extrair dados de páginas HTML.

### **Desenvolvimento Web**

O desenvolvimento da aplicação web envolve HTML, CSS e JavaScript. HTML estrutura a página, CSS estiliza e JavaScript adiciona interatividade.

### **Implementação de Chatbots**

Os chatbots são programas que simulam conversas humanas. Podem ser baseados em regras, utilizando padrões de palavras-chave, ou em aprendizagem máquina, utilizando processamento de linguagem natural.

## Requisitos

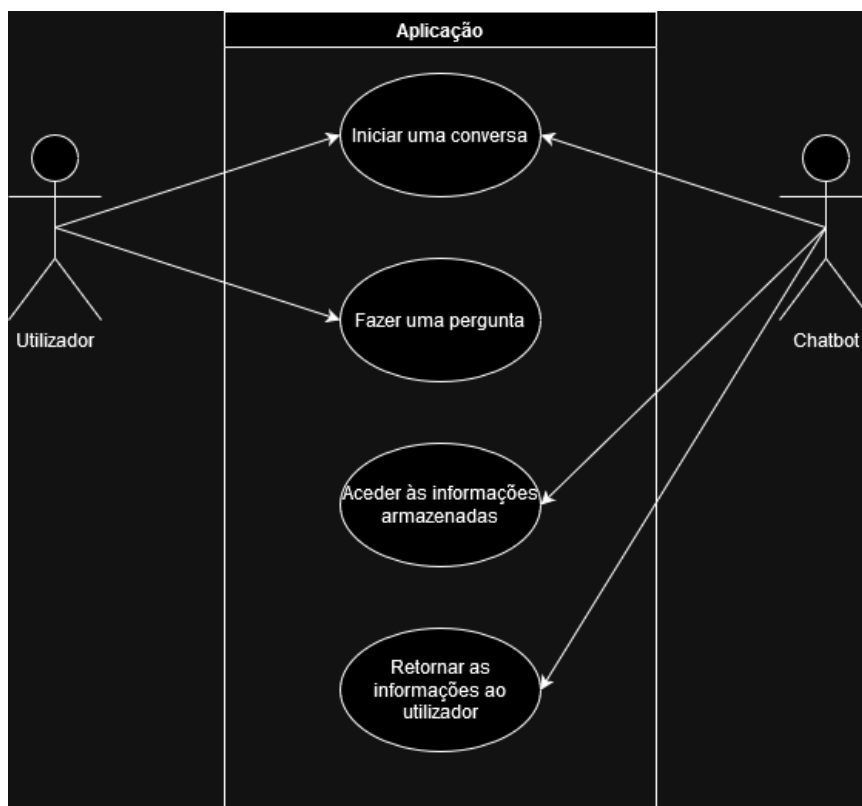
### Requisitos Funcionais

- Integração com o website [www.ispgaya.pt](http://www.ispgaya.pt) para obter informações sobre licenciaturas, mestrados, etc;
- Respostas a perguntas feitas por utilizadores;
- Interface web amigável e responsiva.

### Requisitos não-funcionais

- Suporte para português e inglês;
- Text-to-speech

Figura 1. *Diagrama de casos de uso*



## **Descrição do trabalho**

Nesta secção, descreve-se o trabalho realizado, detalhando cada etapa do desenvolvimento da aplicação web e chatbot.

### **Web Scraping**

Utilizou-se a biblioteca BeautifulSoup em Python para extrair os dados das páginas do site do ISPGaya. O script foi configurado para procurar classes CSS específicas e extrair o texto relevante.

### **Armazenamento em JSON**

Os dados extraídos foram armazenados num ficheiro JSON, estruturado com chaves correspondentes às categorias de informação (ex.: “licenciaturas”, “mestrados”, etc).

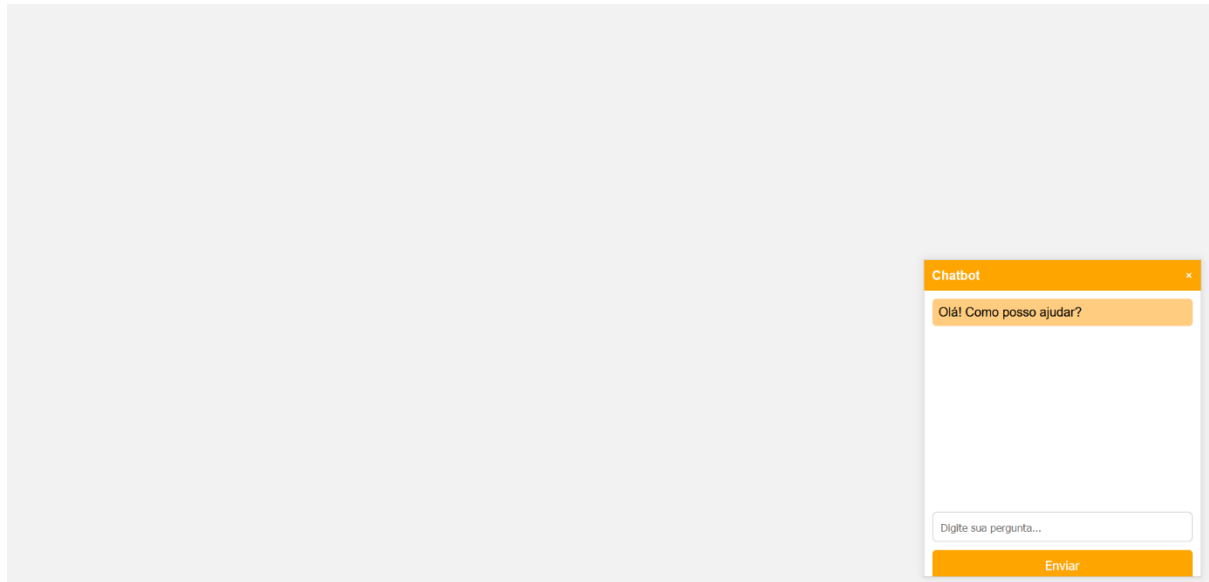
### **Desenvolvimento do Chatbot**

O chatbot foi implementado utilizando JavaScript, com lógica para reconhecer palavras-chave nas perguntas dos utilizadores e procurar as respostas no ficheiro JSON.

### **Interface Web**

A interface web foi desenvolvida com HTML e CSS, garantindo que fosse responsiva e amigável. A interação com o chatbot foi facilitada por JavaScript.

Figura 2. *Resultados finais*



## Testes de validação

ESPECIFICAÇÃO DE CASO DE TESTE	
<b>Identificador:</b>	PROJECT-TCS-001
<b>Objetivo:</b>	Garantir que o sistema é capaz de aceder ao site <a href="http://www.ispgaya.pt">www.ispgaya.pt</a> , realizar web scraping e armazenar as informações relevantes num ficheiro JSON.
<b>Autor(es):</b>	Ricardo Alves
<b>Especificação de Entradas</b> a. URL das páginas do website b.	
<b>Especificação de Saídas</b> a. Ficheiro JSON com as informações extraídas, organizadas de acordo com as categorias b.	
<b>Outros</b>	
<b>Dependências</b>	

ESPECIFICAÇÃO DE CASO DE TESTE	
<b>Identificador:</b>	PROJECT-TCS-002
<b>Objetivo:</b>	Garantir que o chatbot pode responder corretamente às perguntas dos utilizadores, com as informações armazenadas no ficheiro JSON.
<b>Autor(es):</b>	Ricardo Alves
<b>Especificação de Entradas</b> a. Pergunta do utilizador b.	
<b>Especificação de Saídas</b> a. Resposta relevante baseada nas informações armazenadas b.	
<b>Outros</b>	
<b>Dependências</b>	



ESPECIFICAÇÃO DE CASO DE TESTE	
<b>Identificador:</b>	PROJECT-TCS-003
<b>Objetivo:</b>	Garantir que a interface web do chatbot seja amigável e responsiva, proporcionando uma boa experiência de utilização
<b>Autor(es):</b>	Ricardo Alves
<b>Especificação de Entradas</b> a. Interação do utilizador através da interface web (cliques, perguntas digitadas, etc) b.	
<b>Especificação de Saídas</b> a. Respostas e interações adequadas exibidas de forma correta e responsiva b.	
<b>Outros</b>	
<b>Dependências</b>	

## Cronograma

Atividade	Previsto	Cumprido	Desfasamento
Definição de requisitos	1 semana	1 semana	0
Implementação do Chatbot	3 semanas	4 semanas	+ 1 semana
Desenvolvimento Web	2 semanas	3 semanas	+ 1 semana
Testes	1 semana	1 semana	0

## Meios previstos e meios necessários

### Meios Humanos

#### 1. Estudante de Engenharia Informática

##### Responsabilidades:

- Definição dos requisitos do projeto;
- Desenvolvimento e implementação do sistema de web scraping;
- Desenvolvimento do chatbot;
- Criação da interface web;
- Testes e validação do sistema;
- Redação do relatório final.

##### Perfil:

- Conhecimentos em programação (Python, JavaScript);
- Familiaridade com desenvolvimento web (HTML, CSS, JavaScript);
- Experiência básica em web scraping (biblioteca BeautifulSoup);
- Habilidades de gestão de projetos e documentação técnica.

#### 2. Orientador

##### Responsabilidades:

- Fornecer orientação técnica e suporte ao longo do projeto;
- Ajudar na definição dos objetivos e requisitos do projeto;
- Rever e fornecer *feedback* sobre o progresso e a documentação do projeto.

##### Perfil:

- Habilidades de mentoria e orientação académica.

## **Meios materiais**

### **1. Hardware**

Computador Pessoal:

- Sistema operativo: Windows, macOS ou Linux

### **2. Software**

Visual Studio Code (VS Code):

- Editor de código-fonte para desenvolvimento do script de web scraping e da aplicação web.

Navegador web:

- Google Chrome, Mozilla Firefox ou outro navegador moderno para testar a interface web.

Meios previstos de acordo com os meios necessários, sem ocorrência de desfasamentos.

## **Problemas e decisões**

### **Problemas Detectados**

- Inicialmente, dificuldades na extração precisa de dados devido à estrutura HTML variável das páginas web;
- Problemas de performance na resposta do chatbot com grandes volumes de dados.

### **Decisões Tomadas**

- Ajuste do script de web scraping para lidar com diferentes estruturas HTML;
- Implementação de otimizações no código do chatbot para melhorar a performance.

## **Conclusões**

Este projeto de fim de curso resultou na criação de uma aplicação web com um chatbot, que utiliza informações extraídas do site do ISPGaya para responder a perguntas dos utilizadores. O desenvolvimento desta solução envolveu a aplicação de técnicas de web scraping para recolher dados relevantes das páginas web, o armazenamento desses dados em formato JSON, e a implementação de um chatbot em JavaScript capaz de processar perguntas e fornecer respostas baseadas nesses dados.

O projeto atingiu os objetivos propostos, proporcionando uma ferramenta eficaz para melhorar a acessibilidade e a experiência do utilizador no acesso a informações sobre o ISPGaya. Através da utilização de tecnologias web, foi possível criar uma interface amigável e responsiva, permitindo que os utilizadores interajam facilmente com o chatbot e obtenham respostas rápidas e precisas.

Ao longo do desenvolvimento do projeto, foram enfrentados desafios, como a variação na estrutura HTML das páginas web e a necessidade de otimizar a performance do chatbot. Essas dificuldades foram superadas com ajustes no script de web scraping e otimizações no código do chatbot.

Futuramente, o projeto pode ser expandido para incluir técnicas avançadas de processamento de linguagem natural (NLP), melhorando ainda mais a precisão e a relevância das respostas do chatbot. Além disso, a cobertura web scraping pode ser ampliada para incluir mais secções do site do ISPGaya, oferecendo uma gama mais ampla e informações aos utilizadores.

Estas conclusões demonstram o sucesso do projeto e as suas potencialidades para evolução, evidenciando a importância de soluções tecnológicas inovadoras para a melhoria da comunicação e acesso à informação em instituições de ensino superior.

## Referências bibliográficas

- Documentação oficial da biblioteca BeautifulSoup:  
<https://www.crummy.com/software/BeautifulSoup/bs4/doc/>
- W3Schools para tutoriais de HTML, CSS e JavaScript: <https://www.w3schools.com/>

## Apêndices

Código do projeto: <https://gitfront.io/r/ricardo10alves/sskWTMgY7Zbs/chatbot/>

Website com documentos:

[https://ricardo10alves.github.io/website\\_projeto\\_final\\_2019102163/](https://ricardo10alves.github.io/website_projeto_final_2019102163/)

(username == admin / password == password123)