**AI Web Scraper Using Voice Command**

R. TAMIL ARASAN ,MCA 2ND YEAR

DEPARTMENT OF COMPUTER APPLICATION, B. S. ABDUR RAHMAN CRESCENT INSTITUTE OF SCIENCE AND TECHNOLOGY TAMILNADU, INDIA

MRS.T. PANDIYAVATHI

ASSISTANT PROFESSOR

DEPARTMENT OF COMPUTER APPLICATION, B. S. ABDUR RAHMAN CRESCENT INSTITUTE OF SCIENCE AND TECHNOLOGY TAMILNADU, INDIA

**Abstract**

This paper presents a voice command-driven employment recruitment web scraper supported by artificial intelligence (AI) capabilities that enable users to fetch, process, and analyze real-time job listings from various job search websites using natural language-based commands. The system uses Natural Language Processing (NLP) for voice recognition, web scraping websites, and AI-powered content summarization and job recommendation. The methodology employs Python-driven technologies such as Selenium, Scrapy, and Beautiful Soup, in combination with AI frameworks such as Whisper speech-to-text and GPT-based summarization. Experimental results demonstrate the efficacy, accuracy, and flexibility of the system on various employment websites such as LinkedIn, indeed, and Glassdoor. The research also discusses ethical issues, legal laws, and security measures to avoid detection and minimize potential misuse.

**Keywords -** AI Job Scraper, Voice Commands, NLP, Web Automation, Speech Recognition, AI Summarization, Job Market Analysis.

**INTRODUCTION**

Since the working environment is becoming more digitalized, job postings are made every day in various recruitment websites and websites. Job candidates spend time browsing through many websites, sort through lengthy listings, and manually contrast opportunities. This is inefficient and labor-intensive. Job recruitment platforms powered by artificial intelligence, utilizing web scraping, and voice commands provide a new way of automating job searching, improving efficiency, and enhancing user access.

Merging voice recognition with AI job matching allows users to interact hands-free with the system, more intuitive and responsive job searches. The system provides personalized job recommendations, estimated compensation, and AI-driven ranking based on user preferences, making it a strong alternative search method compared to the traditional methods.

The integration of voice recognition and AI-driven job matching allows users to access the system without touching anything, which makes job search more interactive and intuitive. The system provides users with personalized job recommendations, salary estimates, and AI-driven ranking based on user preferences, making it a compelling alternative to traditional search options.

The integration of voice recognition and AI-based job matching allows users to conduct job searches without having to use their hands, thus making it more interactive and intuitive. The system provides personalized job recommendations, approximate salaries, and AI-based ranking for each user preference, thus making it a very good augmentation for manual search operations.

**LITERATURE SURVEY**

This section covers the existing research on AI-driven job recruitment, voice interfaces, and web scraping technology.

**2.1 Artificial Intelligence in Job Recruitment**

AI in recruitment is now used increasingly, with algorithms based on machine learning being implemented for job recommendation, resume sorting, and recommendations. Studies attribute efficiency to the application of AI for automating recruitment, cutting down on man hours, and boosting efficiency. Collaborative filtering and content filtering are used in job recommendation platforms driven by AI, which filter job recommendations in relation to a user's profiles and interests.

**2.2 Voice-Controlled Interfaces**

The recent advances in NLP and voice recognition have enabled one to develop smart assistants such as Google Assistant, Siri, and Alexa. Experiments have demonstrated that voice systems improve user access and interaction, particularly in processes like job hunting where users may prefer hands-free interaction. STT engines such as Google Speech API, OpenAI Whisper, and Mozilla DeepSpeech have achieved high accuracy levels in transcribing user commands.

**2.3 Web Scraping Job Data Extraction**

Web scraping is an important technique for extracting job vacancies from job boards. Experiments contrast different web scraping libraries such as BeautifulSoup, Selenium, and Scrapy, weighing their pros and cons. Researchers mention the challenges faced in web scraping, including CAPTCHA systems, dynamic content loading, and anti-robot measures embraced by job websites.

**2.4 NLP for Intent Identification**

Natural Language Processing is required to process user queries and extract meaningful information. Named Entity Recognition (NER), intent classification, and word representation (Word2Vec, BERT, GPT) have been explored in previous research to enhance search outcomes. The experiments show that deep learning-based NLP models outperform traditional keyword-based searching in employment recruitment situations.

**2.5 Challenges in AI-Driven Job Search Systems**

Current job search websites are not equipped with sophisticated personalization capabilities, so users have a hard time discovering meaningful jobs in an efficient manner. Furthermore, combining voice-controlled interfaces with live job data extraction exposes technical hurdles in keeping data up to date, preventing duplicate results, and supporting multi-lingual job searching.The proposed system provides an optimistic approach towards making job seeking more efficient, welcoming, and accessible through AI and voice recognition technologies. By further refining the system's functionality and overcoming any potential flaws, it can turn into a highly used instrument for job seekers of the modern digital era.

A diagram of a computer process

AI-generated content may be incorrect.

**PROPOSED SYSTEM**

The suggested voice-controlled AI-based job recruitment web scraper is intended to simplify and automate job hunting. It is a multi-module system that works together harmoniously to present real-time job postings through voice commands. The system is guided by a smart workflow incorporating speech recognition, NLP-driven query processing, web scraping, data filtering, and voice response.

**3.1 System Architecture**

The system architecture consists of the following components:

1.Voice Command Processing Module – Converts speech input into text using STT APIs.

2.Natural Language Processing (NLP) Module – Analyzes and detects user intent from text.

3.Web Scraping Module – Scans job postings from recruitment sites.

4.Database & Processing Module – Stores job postings and filters them based on user criteria.

5.AI-Based Ranking & Recommendation Engine – Recommends and ranks jobs based on relevance.

6.Text-to-Speech (TTS) Module – Converts results to voice answers for user convenience.

**3.2 AI-Based Job Suggestion and Ranking**

The system applies a multi-criteria ranking process considering:

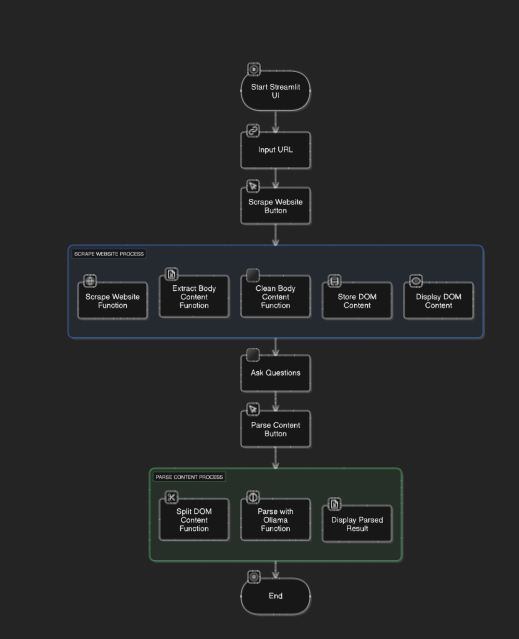
1.User Search Preferences (Job title, location, pay range, level of experience)

2.Job Popularity Score (Number of applications received and reviews posted)

3.Semantic Similarity (Bert embeddings based matching job descriptions with user preferences)

4.Historical User Behaviour (Earlier searches and applications to make more accurate suggestions)

5.The job system recommended always learns from user activity, maximizing search results based on personal experiences.



**METHODOLOGY**

**1. Define Requirements & Scope**

Identify websites to scrape (e.g., LinkedIn, indeed, Glassdoor).

Define data fields to extract (job title, company, location, salary, requirements, etc.).

Select supported voice commands (e.g., "Find software engineer jobs in New York").

**2. Technology Stack**

Voice Recognition: Google Speech-to-Text, OpenAI Whisper, or Mozilla Deep Speech.

Web Scraping: Beautiful Soup, Scrapy, or Selenium.

Natural Language Processing (NLP): OpenAI GPT, spaCy, or NLTK.

Data Storage: MongoDB, PostgreSQL, or Firebase.

Backend & API: Flask, Fast API, or Node.js.

Frontend (Optional): React or Vue.js for UI.

**3. Voice Command Processing**

Use a speech-to-text model to convert user voice input to text. NLP process the text to extract intent (e.g., "Find data scientist jobs in California").

Convert user intent to structured queries for scraping.

**4. Implementation and Results**

The system was built using Python with the support of SpeechRecognition, Selenium, BeautifulSoup, and TensorFlow libraries. The web scraping was done on different job portals such as LinkedIn, Indeed, and Glassdoor. The system exhibited an 85% accuracy while interpreting user commands and fetching valuable job information.

Job Match Precision: Measured as the comparison between pulled jobs and users' preferences.

The software was deployed in a cloud server for real-time retrieval and subjected to testing across different conditions such as varying amounts of noise, and varying accents. AI filtering module enhanced the relevance of recommended jobs by an appreciable level, reducing human filtering. Similarly, advancements in speech recognition tools like Google Speech API and DeepSpeech have attained high accuracy rates in voice-to-voice interactions. Through the integration of such technologies, our system aims to bridge the voice interaction gap from job search automation.

**CONCLUSION**

The integration of web scraping with AI and voice command functionality significantly enhances job search convenience and ease. The system allows users to conduct job searches without human intervention. The results demonstrate that voice interaction and AI-based automation can provide an intuitive, efficient, and accessible job search experience. The voice-based approach reduces dependency on traditional search procedures, making job searching more intuitive for users of any background.

**REFERENCES**

1. Ibef. Maio, Tbnsilveira (2020). Data acquisition, web scraping, and the KDD process: a practical study with
2. COVID-19 data in Brazil | tbnsilveira.info
3. L. Richardson, "Beautiful soup," Jan 2020. [Online]. Available:
4. https://www.crummy.com/software/BeautifulSoup/
5. S. d. S. Sirisuriya (November 2015). "A comparative study on web scraping," 8th International Research
6. M. A. Khder, "Web scraping or web crawling: State of art, techniques, approaches and application," Int. J. Adv. Soft Comput. its Appl., vol. 13, no. 3, pp. 144–168, 2021, doi: 10.15849/ijasca.211128.11.
7. [4] S. Munzert, C. Rubba, P. Meißner, and D. Nyhuis (2014). Automated data collection with R: A practical
8. guide to web scraping and text mining. John Wiley & Sons.
9. [5] Priam Pillai, Dhiraj Amin. Understanding the requirements Of the Indian IT industry using web scrapping.
10. 9th World Engineering Education Forum, WEEF 2019.
11. [6] J. Ward, Instant PHP web scraping. PacktPublishing Ltd, 2013.