A Bilingual LLM-Based Platform for Job Search and Career Guidance: NaukariCraft

Shruti Sharma[1,*], Satvik Bhardwaj[1], Sonakshi Vij[1], Gopal Chaudhary[1]

[1]School of Engineering & Technology, Vivekananda Institute of Professional Studies - Technical Campus, AU-Block (Outer Ring Road) Pitampura, Delhi - 110034, India

Emails: 05217711621_ml@vips.edu; 01417711621_ml@vips.edu; sonakshi.vij92@gmail.com; gopal@vips.edu

Abstract

Finding jobs in today's world is similar to finding a needle in the haystack. The modern job-search platforms present a language barrier for native-speakers and inexperienced candidates, making it difficult for them to compete in the job search race. NaukariCraft, a bilingual (Hindi & English) job search platform makes it easy for users to look for jobs, gain industrial insights, save time by finding relevant jobs tailored to skills and resume, building ATS friendly resume, and ATS score analyzer. NaukariCraft provides full guidance to novel applicants helping them find direction towards jobs tailored to their resume. Using advanced technology like Large Language Model (LLMs) and agents, NaukariCraft enhances user experience, improves employability through resume analysis, and reduces application fatigue. This paper outlines the methodology, proposed work, result, conclusion and future development avenues for NaukariCraft.

Keywords: Bilingual; LLM; Agents; Job search; Employment; Resume

1. Introduction

Job searching is a lengthy and tedious task for inexperienced and new individuals with differences between skills they possess and the needs of the jobs they seek leading to inefficient and inappropriate job search. Many traditional job-search platforms favour English, making it difficult for native speakers to navigate through the app, leading to missed opportunities. Due to this biased favouring of the English language, many native speakers often miss the opportunity to apply to jobs and upskill themselves. Moreover, simple keyword searches that these traditional job-search platforms often fail to capture the skills and qualification of the applicants. Most of the existing systems provide generic, non-personalized employment suggestions instead of skills-focused, fact-based matches. The lack of real-time insights shifts within an industry and the state of the labour market prevents applicants from making fully informed career decisions and staying ahead in the career race by upskilling according to their domain.

To move past these problems, we introduce NaukariCraft, an ultimate guide that uses LLMs to provide employment suggestions, find jobs according to user's qualification and skill sets provide recent trends and insights specifically to their domain, and provide bilingual support (English and Hindi). Most of the existing employment search websites and engines fail to fully capture the nuances of resumes and skills and the changing world of employment. Our vision is to build an AI system that not only extracts resume information and find jobs, but a complete solution to find relevant jobs based on real-time employment intelligence, give guidance on how to improve your resume based on the ATS score, provide a list of skill you lack and what are the recent trends which can be used to upskill. Using LLMs in the analysis of resume and employment matching, the platform attempts to widen the employment opportunity. By incorporating Hindi language, this platform caters to a larger audience,

DOI: https://doi.org/10.54216/FPA.210111
Received: January 27, 2025 Revised: May 16, 2025 Accepted: July 07, 2025

hence involvement of larger unemployed applicants. This platform aims to fill the gaps of the existing traditional technology with the use of LLMs based technology and real-time employment market analysis, providing results in both Hindi and English languages.

NaukariCraft has an effortless and integrated system of employment suggestions with a variety of critical capabilities. The users send in resumes and the platform extracts the critical skills and experience and the preferred kinds of employment using a large language model. Our custom artificial intelligence based web scraper scrapes relevant job postings over the internet based on the parsed resume and provides a matching score of required skills and user's skills. The platform also uses a multi-agent system to provide a deep trend and insights analysis report according to the user's domain of interest in both Hindi and English languages. To avoid unnecessary fatigue while applying to jobs, NaukariCraft provides relevant jobs specific to the user's skills and resume. Moreover, it analyzes the user's resume and provides ATS score, and necessary improvements required according to the user's desired profile using a large language model. Along with these features, it lets you build an ATS friendly resume from scratch.

2. Related Work

Employers also find it hard to sift through the numerous application forms and it becomes hard to spot the suitable candidates the old-fashioned way. One of the techniques involves analysing resumes and shortlisting the fit with the aid of BERT sentence pair classification and it has a 73% fit [1]. Another study highlights the limitations of Applicant Tracking Systems (ATS), with the primary problem of parsing accuracy at 60-70% at times. Through the use of Large Language Models (LLMs), Natural Language Processing (NLP), and better Optical Character Recognition (OCR), the parsing will be incredibly high [2]. The notion of agentic AI systems that allow AI systems to autonomously follow complex objectives with minimal oversight has also been experimented with. The systems can accelerate processes such as finding employment but require cautious integration in a manner that will limit the associated dangers [3]. Traditional wrapper-based approaches are rigid, while LLM-powered language agents can be hard to reuse. Autoscraper, a two-stage framework powered by LLMs, enhances adaptability and efficiency in various web environments by using HTML structures and cross-page similarities [4].

An approach was proposed that coupled unsupervised feature learning, base classifiers, and ensemble techniques with the intention of reducing prediction errors. Through evaluation with a large database of resumes, it was established that AI-augmented models of rankings effectively reduced mismatches and placed great importance on transformer-type architectures in the optimization of employment search experiences [5]. A Conditional Random Fields based hierarchical extraction framework was proposed for annotating resume blocks that had been segmented. The approach dramatically enhanced the extraction accuracy through the combination of content- and layout-based features and the addition of the latter that lifted the F1-score by more than 20% [6]. A study came up with a deep learning algorithm that involves CNN, CRF, and Bi-LSTM with the aim of extracting entities. The technique has a higher accuracy rate compared to the use of conventional keyword- and rule-based methods and fits better in real-world hiring environments [7].

The UTAUT2 model was applied in examining the use of AI-supported tools in finding a job and the factors that drive them. The findings inform the optimization of AI-supported employment processes that benefit the recruiter and the applicant alike in an evolving employment environment [8]. A study applied generative AI in extracting structured labour market intelligence from web-based job adverts and completing real-time workforce informational needs. The results demonstrate the promise of generative AI in the changing analysis of labour market trends and the development of public policy [9]. A study showed how agentic artificial intelligence improves the reasoning abilities of language models by framing them into iterative workflows. The study compares different multiple-agent collaboration (MAC) frameworks and how they adopt the pillars and govern the flow of information. The paper also touches upon the practical use of MAC frameworks and how they can enhance the problem-solving abilities of AI in a variety of fields [10]. The EXPERT was proposed which is an intelligent e-recruitment tool that enhances candidate screening using ontology mapping. The study highlights how ontology-based recruitment can enhance the efficiency of hiring processes in large-scale recruitment scenarios [11].

The study introduced an unsupervised technique of extracting and merging data from partially overlapped web sources with a structured format. They use WEIR, an algorithm that employs redundancy among sources in the fine-tuning of extracting and merging the data. They enhance the automation of web data harvest and merging [12]. LASER, a solution that improves the performance of large language models at web navigation tasks based on state-space problem formulation is also introduced. The authors benchmark LASER with web navigation tasks

DOI: https://doi.org/10.54216/FPA.210111
Received: January 27, 2025 Revised: May 16, 2025 Accepted: July 07, 2025

and demonstrate that it attains higher performance compared with existing methods and converges toward human-level accuracy [13]. Another study introduced the system of World of Bits (WoB), an effort that attempts to advance the frontier of reinforcement learning (RL) research in real web environments that are similar to the real world. Experimental results demonstrate that agents that train with behaviour cloning and RL perform a variety of web tasks effectively and demonstrate WoB's promise of enabling RL study in real web environments [14].

A detailed review on why and for what purpose recruiters use 'Apply' and 'Easy Apply' feature for recruitment was presented. The study indicated that most recruiters prefer 'Easy Apply' for all recruitment on LinkedIn because of its simplicity. The "Easy Apply" feature was discovered to be the only feature used by all of the participants for their LinkedIn job recruitment [15]. A way to simplify the application process, using python programming with the power of LinkedIn, was proposed, by creating an engine that downloads jobs from LinkedIn and submits applications on behalf of users. Their system automatically fills out an application form using the information provided by the user. Their system provides an application tracking system, tracks employee responses and sets follow-up activities [16]. An innovative approach that leverages artificial intelligence for career growth was introduced. Results show the system improves efficiency and accuracy, enabling individuals to make informed career decisions and accelerate professional growth [17].

3. Proposed Work and Methodology

NaukariCraft proposed system is a bilingual LLM-based employment seek platform that will improve employment suggestions with the help of resume analysis, employment categorization, business intelligence, and skills matching. The system has a workflow as shown in Fig. 1 that involves a set of integrated modules that enable smooth, individualized, and evidence-supported employment suggestions.

The methodology is made up of data collection, pre-processing all data obtained, as well as models that are implemented in order to categorize or summarize all data. This project heavily relies on web scraping tools, LLMs and agents.

A. Resume parser and processing

Users upload resumes, with text extracted via PdfReader. Gemini model (gemini-2.5-flash-preview-04-17) parses a resume and extracts structured data, which includes name, contact, skills, educational background, work experience, qualifications, languages, suggested job roles, and suggested category of a resume. Parsed data is stored in a Flask session for further processing.

B. Job Scraper and Resume Processing

LinkedIn job listings are scraped using httpx and BeautifulSoup, with user-agent rotation and delays to prevent detection. Jobs are accessed based on proposed job roles that are categorized by the Gemini model (gemini-2.5-flash-preview-04-17). Posting dates are standardized with the dateparser library. Resumes and job skills are collected, standardized, and made comparable. Jobs are displayed in order from highest match score to lowest. Percentage match is worked out as:

$$\left(\frac{Matched\ Skills}{Total\ Job\ Skills}\right) X\ 100 \qquad (1)$$

Built with Flask, users are allowed to upload resumes, display parsed data, and browse job matches dynamically. Job postings are displayed with match scores, skills, and job data. Session management makes a transition between pages seamless. The pre-trained spaCy model (en_core_web_sm) is used in text processing. The job description is pre-processed and tokenized in preparation for downstream work. The technical skills are also enriched with a specialized regular expression although spaCy's NER is also utilized. A specialized regex pattern extracts technical skills. This is utilized in detecting skills from job postings. Comparison between a job posting and a CV is made by converting the text into a numerical form with the Term Frequency-Inverse Document Frequency (TF-IDF) algorithm. Term frequency (TF) indicates how many times a term is repeated in a document. Inverse Document Frequency (IDF) normalizes terms that are highly frequent in the entire collection. Formula for the same is given by:

DOI: https://doi.org/10.54216/FPA.210111
Received: January 27, 2025 Revised: May 16, 2025 Accepted: July 07, 2025

Figure 1. Flow-chart of proposed work

$$TF - IDF(t,d) = TF(t,d) \times log(\frac{N}{DF(t)}) \quad (2)$$

where TF(t,d) is the frequency in a single document d of a term t, N is the total number of documents, and DF(t) is the frequency of a document with a term t. Once both job descriptions and the resume are in terms of TF-IDF representations, cosine similarity is calculated between these two texts. Formula for the same is given by:

$$Cosine\ Similarity = \frac{A.B}{\|A\| \times \|B\|} \quad (3)$$

where A is a CV's TF-IDF vector, and B is a job posting's TF-IDF vector. Jobs are categorized in order of similarity scores. Duplicated job postings (same web address) are removed, making postings unique. Final outputs (JSON form) are job data, skills, and sources.

DOI: https://doi.org/10.54216/FPA.210111
Received: January 27, 2025 Revised: May 16, 2025 Accepted: July 07, 2025

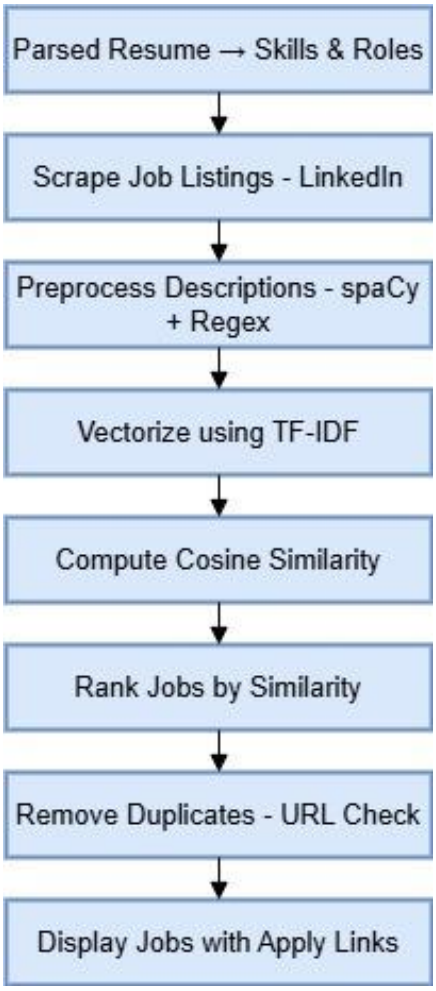

Figure 2. Job Matching & Ranking Pipeline

C. News Agent Workflow

The platform orchestrates a stepwise AI workflow in order to produce a guide on upskilling a given professional domain through three AI-based agents that are LLM-driven. First, it looks for articles, which is a domain-specific career insight generator. Then it scrapes articles and extracts in-depth information from a given article URL. It gives insights that are more technical, new trends, suggested skills, as well as requirements in terms of recruiting. All three agents are sequentially invoked by the central workflow. Web searcher agent processes the initial search results and refines them into a structured format. Article scrape agent processes the article contents in order to pull out in-depth career-based information. Career advisory agent processes the gathered data and present a markdown-based upskilling guide that consists of new technologies, breakdown of core skills, learning roadmaps, interview preparation, and career strategy advice. All agents in the workflow are equipped with a Groq model instance (specifically with ID "llama-3.3-70b-specdec").

D. ATS score calculator and analyser

The user uploads a resume and text is extracted and stored in the user's flask session to maintain user-specific data during analysis workflow temporarily. The user selects a desired job category and specific role, if no input is provided it use the category and role based on the parsed resume. A custom prompt is generated in the backend that includes the extracted resume text, parsed resume data and the selected or inferred job category or role. The generated prompt is passed to the LLM ('gemini-2.5-flash-preview-04-17'). The model analyses the resume content in context of the given role and responds back with a JSON-formatted evaluation which includes, a total ATS score, section-wise analysis (skills, education, experience, certification, formatting), specific missing information, suggestions for improvement.

DOI: https://doi.org/10.54216/FPA.210111
Received: January 27, 2025 Revised: May 16, 2025 Accepted: July 07, 2025

Figure 3. Newsagent workflow

The JSON response is cleaned using python string operations. The cleaned string is parsed into structured python objects using the JSON library. Validation checks are added to ensure that even if some sections are missing in the LLM response, default placeholder values are used, preventing breakdowns in further steps. The total ATS score provided by the model is normalized to ensure it falls between 0 and 100. Based on the normalized score, the resume is categorized into four categories- excellent, good, and average and needs improvement. The backend cross-reference a predefined dataset of recommended courses and career development videos based on the user's job category and role. Matching is implemented through simple dictionary lookups and filtering mechanisms in python. Finally, all the collected outputs are compiled into a structured JSON response and is sent back to the frontend. Flask's flash messaging system is used to alert users about upload issues, invalid formats, or analysis errors.

E. Resume Builder

A DOCX document is created using the python-docx library's Document class. The Professional template is applied, with custom styles defined for a formal layout. User data, stored in a dictionary, is populated into the document using python-docx. The _format_list_items function processes input data, handling both string and list formats. Document margins are set to 0.5 inches (top/bottom) and 0.7 inches (left/right) using python-docx's sections API for efficient space utilization. The completed document is saved to a BytesIO buffer and returned for integration with the Flask web interface. Error handling is implemented with try-except blocks, logging exceptions via the traceback module for debugging. The generated DOCX file is made available for download through the Flask-based web interface, complementing the resume parsing and job matching functionalities. The Professional template's formal design ensures compatibility with corporate job applications.

DOI: https://doi.org/10.54216/FPA.210111
Received: January 27, 2025 Revised: May 16, 2025 Accepted: July 07, 2025

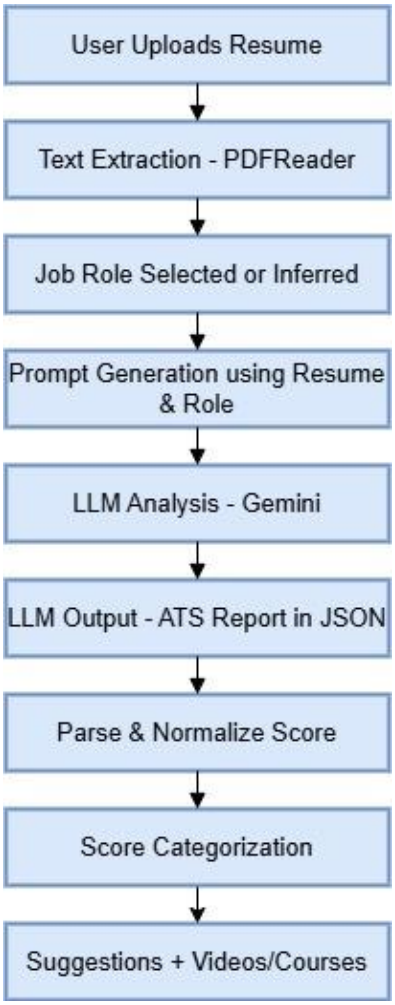

Figure 4. ATS Score Calculation and Analysis

5. Results

Successfully implemented NaukariCraft: LLM-Powered Job Search Platform that leverages Agentic AI for job listing, skill matching and for suggesting latest trends and news to users based on their resume. A registration/login window, which appears first when a user opens the NaukariCraft application, this helps us to keep track of user information by registering users. After successfully registering the user will be directed to the home page and will be asked to upload their resume to parse the content from it, when user uploads the resume then the user is directed to the dashboard on webpage.

From the dashboard, users can select from multiple features. The detailed analysis features what is missing and present in the user's resume, what modifications can be made to make it better, along with some YouTube tutorial on how to make your resume better and some upskilling videos specific to the user's domain. If the user selects ‘view parsed resume’, they will be redirected to the parsed resume page. Another feature is that a user can build their own ATS-friendly resume. When user clicks on ‘find jobs’ options then a list of jobs related to the resume category is displayed along with basic details such as company name, job title, job posted date, number of skills matched, and an apply link button by clicking on which the user is redirected to a page from where they can apply to the particular job.

One other feature of NaukariCraft is showing latest trend and news to users in their work field, it shows latest news and trends based on the resume category, which is determined earlier, it helps user to find latest news for the field in which they are applying for jobs and helps them to prepare well. All the above features that exists in English language are available in Hindi language as well. News and trends insights are specific to a domain in Hindi. These bilingual features helps to serve a larger audience.

DOI: https://doi.org/10.54216/FPA.210111
Received: January 27, 2025 Revised: May 16, 2025 Accepted: July 07, 2025

This section also holds the evaluation of the NaukariCraft platform, focusing on its functional integrity and technical performance. The system was assessed through feature level comparison and internal latency profiling. These evaluation helps in placing NaukariCraft relative to existing job search platforms and show its technical viability for real-time web deployment.

In order to analyze NaukariCraft's strength, a comparative study was performed with three commonly used job portals: Naukri.com, Indeed, and LinkedIn Jobs. According to Table 1, NaukariCraft demonstrates clear strength points where major players have no or minimal provisions — specifically in bilingual interaction, resume reading and interpretation based on ATS, and career insight generation supported by LLM.

Addition of large language models allows NaukariCraft to read and understand resume content at a higher level than mere keyword matching, improving job relevancy and candidate direction. In addition, inclusion of Hindi-language availability places the site within reach of a broader demographic and closes an important gap for usability in the Indian job marketplace.

Table 1: Feature Comparison with Popular Job Portals

Feature	NaukariCraft	LinkedIn	Naukari.com	Indeed
Bilingual (Hindi/English) Support	Yes	No	No	No
Resume Parsing	Yes	Yes (Limited)	Yes	Yes
ATS Score Analyzer	Yes	No	No	No
Job Matching via LLM	Yes	No (Keyword-based)	No	No
Career Insight Generator	Yes	No	No	No
Resume Builder	Yes	Yes (Template-based)	Yes	Yes
Real-Time Response	Yes	Yes	Yes	Yes
Multilingual AI-based Guidance	Yes	No	No	No

Each of the major functional modules was measured for average response time under normal conditions of usage (standard internet connection, Flask-based server, and LLM APIs). As shown on Table 2, all of the core modules responded within a usable time for real-time applications, with resume parsing and job matching taking less than 7 seconds. The trend and insights module, which entails sequential processing of several agents querying external content and summarizing it, was the most latency-incurring process.

The system also showed consistent performance on repeated trials, and no failure of the backend and no considerable delay were faced while testing. These response times uphold the suitability of deploying NaukariCraft as a responsive web app capable of serving users effectively without any perceivable delays.

Table 2: Module-wise Average Latency

DOI: https://doi.org/10.54216/FPA.210111
Received: January 27, 2025 Revised: May 16, 2025 Accepted: July 07, 2025

Module	Average Latency (seconds)	Description
Resume Parsing	6.8	Gemini LLM parsing of uploaded document
Job Matching	5.2	Skill-based scraping and cosine ranking
ATS Score Analyzer	7.4	Full feedback generation and analysis
Trend & Insight Generator	4.3	Multi-agent summarization from web content

7. Conclusion

NaukariCraft represents a significant advancement in providing access to employment opportunities via the integration of bilingual support (Hindi and English) and advanced AI capabilities. By integrating key features such as ATS friendly resume optimization; finding relevant job postings, agent system to provide real-time insights and trends relevant to user's domain, the platform meaningfully reduces the manual burden on users while enhancing the quality, relevance and access to job postings. The platform's bilingual design particularly addresses the needs of Hindi-speaking users, an often under-represented demographic in the digital employment ecosystem. It not only improves access for non-English speakers but also refines the overall job search experience for all users by offering real-time resume feedback, career guidance and a streamlined application workflow. This holistic system marks a step towards making job searches smarter, faster and more equitable through technology-driven intervention.

To expand on this strong foundation, several enhancements are planned for the next phase of development. The expansion to additional regional languages will widen NaukariCraft's reach and inclusivity across India. A personalized job application tracker can be introduced to let users monitor their applications, interview statuses and feedback that will effectively create a complete career management suite. In future, NaukariCraft will be made available as a mobile application on both iOS and android platforms to cater to users who prefer job hunting on mobile devices and to those users who do not have access laptops or monitors and live in remote locations.

Funding: "This research received no external funding"

Conflicts of Interest: "The authors declare no conflict of interest."

References

[1] V. Bhatia, P. Rawat, A. Kumar, and R. R. Shah, "End-to-end resume parsing and finding candidates for a job description using BERT," arXiv preprint, arXiv: 1910.03089, 2019.

[2] E. Kaygin, "Comparative analysis of ML (Machine Learning) and LLM (Large Language Models) in resume parsing: A paradigm shift in talent acquisition," 2023.

[3] Y. Shavit, S. Agarwal, M. Brundage, S. Adler, C. O'Keefe, R. Campbell, and T. Lee, "Practices for governing agentic AI systems," OpenAI Research Paper, 2023.

[4] W. Huang, Z. Gu, C. Peng, J. Liang, Z. Li, Y. Xiao, L. Wen, and Z. Chen, "AutoScraper: A progressive understanding web agent for web scraper generation," in Proc. of EMNLP 2024, pp. 2371–2389, 2024.

[5] Y. Lin, H. Lei, P. C. Addo, and X. Li, "Machine learned resume-job matching solution," arXiv preprint, arXiv: 1607.07657, 2016.

[6] J. Chen, L. Gao, and Z. Tang, "Information extraction from resume documents in PDF format," *Electronic Imaging*, vol. 28, pp. 1–8, 2016.

[7] C. H. Ayishathahira, C. Sreejith, and C. Raseek, "Combination of neural networks and conditional random fields for efficient resume parsing," in Proc. 2018 IC4, pp. 388–393, 2018.

[8] U. M. Digerås and S. A. Svepstad, "AI in job search: An exploratory study of job seeker's adoption of AI and its implications for HR and recruitment," Master's thesis, Handelshøyskolen BI, 2024.

DOI: https://doi.org/10.54216/FPA.210111
Received: January 27, 2025 Revised: May 16, 2025 Accepted: July 07, 2025

[9] M. Howison, W. O. Ensor, S. Maharjan, R. Parikh, S. H. Sengamedu, P. Daniels, et al., "Extracting structured labor market information from job postings with generative AI," *Digital Government: Research and Practice*, 2024.

[10] C. Chawla, S. Chatterjee, S. S. Gadadinni, P. Verma, and S. Banerjee, "Agentic AI: The building blocks of sophisticated AI business applications," *Journal of AI, Robotics & Workplace Automation*, vol. 3, no. 3, pp. 1–15, 2024.

[11] V. S. Kumaran and A. Sankar, "Towards an automated system for intelligent screening of candidates for recruitment using ontology mapping (EXPERT)," *International Journal of Metadata, Semantics and Ontologies*, vol. 8, no. 1, pp. 56–64, 2013.

[12] M. Bronzi, V. Crescenzi, P. Merialdo, and P. Papotti, "Extraction and integration of partially overlapping web sources," *Proc. VLDB Endowment*, vol. 6, no. 10, pp. 805–816, 2013.

[13] K. Ma, H. Zhang, H. Wang, X. Pan, W. Yu, and D. Yu, "LASER: LLM agent with state-space exploration for web navigation," arXiv preprint, arXiv:2309.08172, 2023.

[14] T. Shi, A. Karpathy, L. Fan, J. Hernandez, and P. Liang, "World of bits: An open-domain platform for web-based agents," in Proc. ICML, pp. 3135–3144, 2017.

[15] B. Kutlu, A. Rabea, and C. R. Udeozor, "LinkedIn for recruitment: An examination of recruiters' use of 'Apply' and 'Easy Apply' features," 2018.

[16] P. Dhandre, L. Yadav, and N. Jiwane, "Create an automated job application by using LinkedIn," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 12, no. 5, 2024.

[17] K. S. Kumar, P. Srihari, and C. J. Raman, "AI for career growth: Advanced resume analysis and LinkedIn scraping for personalized job recommendations," in Proc. 2nd ICSSAS, IEEE, India, 2024.

DOI: https://doi.org/10.54216/FPA.210111
Received: January 27, 2025 Revised: May 16, 2025 Accepted: July 07, 2025