Research on the Construction and Dynamic Evolution Analysis of Job and Competency Graphs Driven by Multi-Source Heterogeneous Data

Huaruo Mao*, Zexiang Zhang, Yifei Zhai, Zihao Wang, Yuexin Wang, Ruian Yan
School of Artificial Intelligence and Big Data, Henan University of Technology, Zhengzhou, Henan, China
**Corresponding Author*

Abstract: With the rapid development of emerging information technologies such as artificial intelligence and big data, industrial structures and talent demands are continuously changing. Traditional talent analysis methods based on fixed job classification and keyword retrieval are difficult to meet the demands of the employment market. This paper designs and implements an AI job intelligent analysis platform driven by multi-source heterogeneous data. The system focuses on four core aspects: job data collection, job knowledge construction, job dynamic analysis, and intelligent person-job matching. Through Scrapy and Playwright, the system realizes automatic multi-source data collection, and combines natural language processing, large language models, and knowledge graph technologies to achieve job information analysis, skill entity extraction, and dynamic updates of job competencies. The system uses MySQL and Elasticsearch for structured data storage and full-text retrieval, utilizes Neo4j to construct job-skill-industry knowledge associations, and implements RAG-based job definition generation through LangChain and ChromaDB. The platform realizes functions including multi-source job data management, new job discovery and definition, job competency dynamic evolution analysis, job panorama graph visualization, intelligent resume parsing, and person-job matching recommendation. The system adopts a front-end and back-end separated architecture, with the front-end built based on Vue3 to construct visual interactive pages and the back-end based on FastAPI to provide interface services. The test results show that the platform can effectively support job data collection, job knowledge management, job competency analysis, and talent matching requirements, providing intelligent decision support for enterprise recruitment and individual career development.

Keywords: Multi-Source Heterogeneous Data; Job Competency Graph; Large Language Model; Knowledge Graph; Job Dynamic Evolution; Person-Job Matching

1. Introduction

1.1 Research Background

In recent years, intelligent technologies such as artificial intelligence and big data have developed rapidly, and the industrial structure of the digital economy has been continuously adjusted. The demand for enterprise talent capabilities has shown characteristics of rapid change. Zhao Peng [1] pointed out that the current job allocation in the public sector still relies on managers' experience-based judgments, while personnel information is often stored in paper archives without data-driven and systematic support for precise matching, resulting in prominent person-job mismatch problems. In the field of emerging information technologies, jobs are updated rapidly and have complex technical associations. New types of jobs, such as large language model application engineers and AI agent development engineers, continue to emerge, but clear standards have not yet been established in traditional occupational classification systems. Current recruitment platforms mainly rely on job titles, keyword matching, and manual screening to identify candidates. However, these approaches have limitations when dealing with emerging technology-related jobs. On the one hand, different enterprises may use different names to describe the same job, making it difficult for traditional methods to identify their relationships. On the other hand, job competency requirements are continuously adjusted with technological development, while traditional job databases lack update capabilities and cannot reflect market changes in a timely manner. In addition, job seekers often find it difficult to accurately determine the gap between their own capabilities and the requirements of target jobs. To address the above issues, this paper designs an AI job intelligent analysis platform that utilizes technologies such as natural language processing, large language models, and knowledge graphs to achieve new job discovery, job definition generation, dynamic job competency updates, and person-job matching analysis, forming a complete analysis process from job data acquisition to talent capability matching.

1.2 Research Content

This paper focuses on the design and implementation of an AI job intelligent analysis system. The main research contents include: (1) Multi-source heterogeneous data collection and processing: Job information from different sources is obtained through automated data collection, and natural language processing is used to complete text analysis and data standardization. (2) New job discovery and definition: By analyzing job text features, skill relationships, and industry development trends, emerging job types are identified, and large language models combined with knowledge bases are used to generate standardized job definitions. (3) Dynamic evolution of job competencies: Job data from different time periods is compared to demonstrate changes in skill requirements, and knowledge graphs are used to establish associations among jobs, skills, and industries. (4) Person-job matching analysis: User resumes are analyzed to generate personal competency profiles, and the matching degree is evaluated based on job competency models, providing competency gap analysis and optimization recommendations. This paper conducts research around multi-source job data collection, knowledge construction, new job discovery, competency evolution, and person-job matching. The overall research content framework is shown in Figure 1.

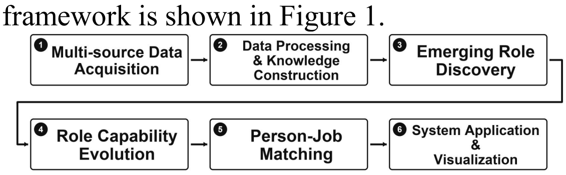

Figure 1. Overall Research Content Framework of the Platform

2. Multi-Source Heterogeneous Data Source Collection and Analysis

2.1 Multi-Source Job Data Collection

Job data serves as the foundation of system operation. Due to the limited coverage and insufficient update frequency of data from a single recruitment platform, this paper adopts a multi-source heterogeneous data collection approach to uniformly acquire recruitment platform data, enterprise recruitment information, and technical ecosystem data. The system combines Scrapy and Playwright to achieve automated data acquisition: Scrapy is responsible for rapid static webpage crawling, while Playwright handles dynamic webpage environments to improve adaptability to complex recruitment websites. The collected data is stored in the MySQL database, while Elasticsearch is utilized to establish full-text retrieval capabilities. Shao Xintian et al. [2] pointed out that traditional relational databases have limited retrieval performance in large-scale data scenarios, while Elasticsearch can significantly improve the query efficiency of large-scale text data through its distributed indexing architecture. This approach is consistent with the retrieval requirements of this system for massive job description texts. The system obtains job data from different sources through Scrapy and Playwright, and stores the collected results in the database. The specific job data collection process is shown in Figure 2.

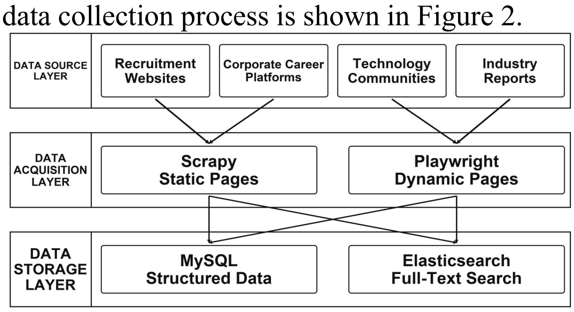

Figure 2. Multi-source Job Data Collection Process

2.2 Job Data Cleaning and Standardization Processing

Job data from different sources has issues such as inconsistent formats, different expressions of job titles, and variations in skill descriptions, requiring data cleaning and standardization processing. The system removes webpage tags, invalid characters, and duplicate content, and performs semantic analysis through natural language processing technology to normalize jobs with different names but similar competency requirements. In terms of skill information processing, tools such as HanLP are utilized to perform entity recognition on job descriptions, extracting key information including technical skills, tool frameworks, and business domains. Combined with large language models, the system analyzes complex semantics and establishes corresponding skill labels. The standardized data includes information such as job titles, responsibilities, skill requirements, industry categories, and data sources, providing a data foundation for subsequent job analysis. To improve the standardization and usability of job data, the system sequentially performs data cleaning, text parsing, skill extraction, and job normalization. The specific process is shown in Figure 3.

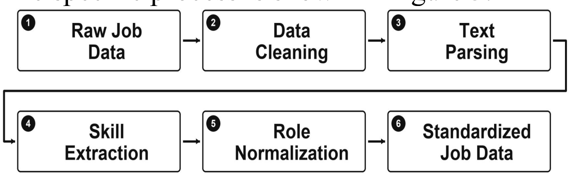

Figure 3. Job Data Cleaning and Standardization Processing Flow

3. New Job Discovery and Definition

3.1 New Job Identification Based on Multi-Source Data

Traditional job classification systems have long update cycles and are difficult to cover emerging occupations in a timely manner. The system designs a new job discovery module based on multi-source data analysis, which comprehensively evaluates job skill requirements, industry application directions, and changes in job quantity. When a certain type of job continuously appears in multiple data sources and its corresponding skill system shows a significant growth trend, the system marks it as a potential new job. For example, under the rapid development of artificial intelligence technology, a large number of job descriptions containing skill requirements such as "large language model applications", "Prompt engineering", and "RAG application development" have emerged. Although these jobs have different names, the system can discover potential new job categories through semantic analysis and job clustering, and perform aggregation analysis to form unified job categories. The system can identify emerging jobs based on multi-source job data and display information such as job popularity, discovery time, industry distribution, and data sources. The new job discovery page is shown in Figure 4.

Figure 4. New Job Discovery Page

3.2 Job Definition Generation and Optimization

To address the problem that emerging jobs lack unified descriptions, this paper designs a job definition generation module based on large language models and retrieval-augmented knowledge techniques. The research of Range et al. [3] shows that combining large language models with structured knowledge graphs can effectively constrain the generation space of models and improve the accuracy and traceability of generated content. Inspired by this, the system builds a job knowledge retrieval process through LangChain, using existing job databases, skill knowledge bases, and industry-related materials as knowledge sources. Combined with large language models, the system generates structured job profiles containing job titles, descriptions, core responsibilities, required skills, additional skills, and typical industry application scenarios. Since the generation process of large language models may involve information bias, the system designs a job definition optimization mechanism. Administrators can manually review and adjust the content automatically generated by AI, and the modification results are fed back into the job knowledge base for subsequent optimization. The system combines large language models with job knowledge bases to generate structured job definitions and supports administrators in reviewing and optimizing the generated results. The job definition generation and optimization page is shown in Figure 5.

4. Dynamic Evolution Analysis of Job Competencies

4.1 Job Competency Change Detection

A static job model alone cannot meet the requirements of talent demand analysis. Therefore, the system designs a dynamic evolution module for job competencies, which continuously collects job data from different time periods and conducts comparative analysis of skill requirements for the same job at different stages. The system identifies newly added competencies, reduced competencies, and changes in competency importance, and presents the basis for these changes by combining data sources. For example, with the development of artificial intelligence technology in recent years, traditional software development positions increasingly require capabilities in AI-assisted development, cloud-native deployment, and intelligent application development. This function helps enterprises understand the changing trends of job requirements and assists job seekers in adjusting their skill structures in a timely manner. To analyze the changes in job skill requirements over time, the system compares job data from different stages and displays newly added skills, weakened skills, and historical change trends. The analysis page is shown in Figure 6.

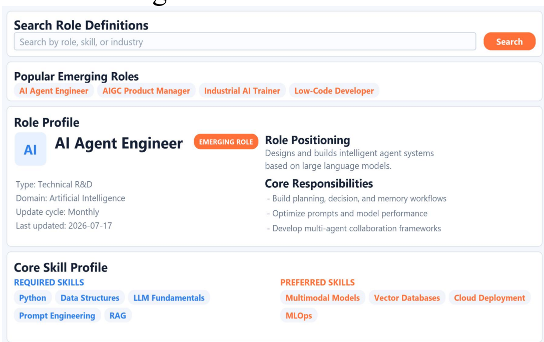

Figure 5. Job Definition Generation and Optimization Page

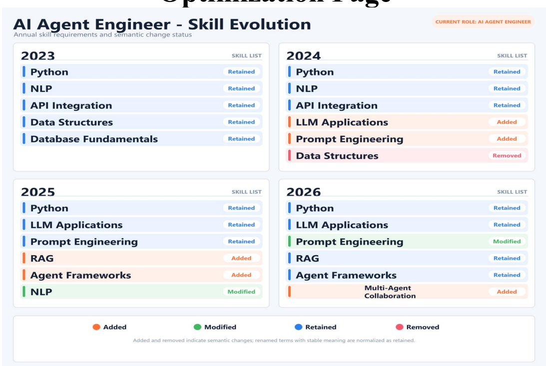

Figure 6. Job Competency Dynamic Change Analysis Page

4.2 Job Competency Graph Construction

To visually demonstrate the relationships among jobs, skills, and industries, the system constructs a job competency panorama graph based on knowledge graph technology. The graph takes jobs, skills, and industries as the main entities and achieves structured representation of job knowledge by establishing relationships among entities. The system adopts Neo4j as the graph database for storing job knowledge relationships and utilizes AntV G6 to implement front-end visualization. It supports users in interactively viewing the skill systems corresponding to jobs or querying related jobs through skill nodes, enabling multi-dimensional analysis. To address the dynamic evolution requirements of graphs, Zhu Yan et al. [4] adopted temporal-aware embedding and graph attention networks to achieve entity evolution representation in knowledge graphs. Inspired by this approach, this paper adds timestamps to nodes in the job competency graph and introduces an incremental update mechanism to track market changes in skill requirements, enabling the graph to support both static relationship visualization and dynamic evolution analysis. Compared with traditional job list presentation methods, the job competency graph can more clearly reflect technical associations and job development trends. The system constructs an association network with jobs, skills, and industries as the main entities, thereby intuitively displaying the relationships among different entities.

5. Person-Job Matching and Competency Gap Analysis

5.1 Resume Parsing and Competency Profile Construction

Traditional resume screening mainly relies on manual reading and keyword searching, which is easily affected by factors such as format differences, variations in description methods, and scattered skill information. Feng Li [5] addressed the problem of Chinese resume information extraction by integrating a named entity recognition method based on RoBERTa dynamic word vectors and BiGRU sequence encoding, which can effectively improve the extraction accuracy of key information such as educational background and work experience. Inspired by this, the system designs an artificial intelligence-based resume parsing and competency profile construction module. The module supports common formats such as PDF and Word, uses OCR technology to complete text recognition, and combines natural language processing to extract key information including educational background, work experience, project experience, and technical skills. During the skill recognition process, the system combines the job competency knowledge base to perform semantic understanding of technical keywords, avoiding the problem that traditional keyword matching can only identify individual skill names. The parsed user information is transformed into a structured competency profile, including skill labels, project experience, technical directions, and competency levels, and is associated with the job competency graph to provide a data foundation for subsequent matching and gap analysis. After users upload resumes, the system can extract skills, project experience, and competency features, and generate structured personal competency profiles.

5.2 Job Matching and Optimization Recommendation Generation

Person-job matching is a core issue in enterprise recruitment, and its effectiveness depends on the comprehensive evaluation of both job requirements and candidate capabilities [6]. The research of Yu et al. [7] further shows that resume-job matching methods based on large language models (LLMs) can effectively improve matching accuracy by ranking candidate resumes and job requirements in a list-wise manner. In addition, Jouanneau et al. [8] pointed out that score calibration is crucial in talent matching scenarios-the same score should have consistent semantic meanings among different job-candidate pairs, providing a quantifiable foundation for competency gap analysis. The system uses the job competency graph as the job requirement model and performs correlation analysis between user resume parsing results and the skills required by jobs. When users possess most of the core competencies required by a job, the matching score is increased; when key skills are missing, the competency gaps are explicitly displayed in the results. For example, for the position of "Large Language Model Application Engineer", the system can analyze the user's existing capabilities such as Python and machine learning, while identifying missing skills including RAG framework application, vector databases, and agent development. It then provides learning recommendations based on job competency development trends. Compared with traditional recruitment systems that only output results such as "matching successful" or "matching failed", the system proposed in this paper further provides competency improvement paths, transforming the job matching process from simple screening into talent development assistance. The system compares user competency profiles with target job requirements, calculates job matching degrees, and provides competency gaps and learning recommendations.

6. System Implementation and Experimental Analysis

6.1 Overall System Architecture Implementation

The AI job intelligent analysis platform designed in this paper adopts a front-end and back-end separated architecture, consisting of a data acquisition layer, data processing layer, data storage layer, intelligent analysis layer, and application presentation layer. Qian Lei and Long Xinyue [9] proposed a three-layer architecture concept of "basic support—technology empowerment—application adaptation". This paper draws on their layered collaborative design concept. The data acquisition layer uses Scrapy and Playwright to achieve automatic acquisition of multi-source job information; the data processing layer utilizes HanLP, OCR, and large language model interfaces to complete text parsing, skill entity recognition, and resume information extraction. The data storage layer adopts multi-database collaboration: MySQL stores basic job information and business data, Elasticsearch implements full-text search, Neo4j stores job-skill-industry knowledge relationships, and ChromaDB stores vectorized knowledge content. Huang Liming [10] pointed out that multi-source heterogeneous data should be assigned differentiated storage resources according to its characteristic differences to optimize performance. Based on this, the system adopts the above multi-database collaborative hierarchical storage strategy to balance the transactional consistency of structured data, the retrieval efficiency of unstructured text, and the specialized storage and access requirements of graph relationships and vectors, avoiding performance bottlenecks caused by a single storage engine. The intelligent analysis layer combines LangChain and large language models to implement functions including new job discovery, job definition generation, competency change analysis, and intelligent question answering. The application presentation layer uses Vue3, Vite, and Pinia to build the interactive interface on the front end, while the back end uses FastAPI to provide interface services. The system adopts a front-end and back-end separated layered architecture, consisting of the data acquisition layer, data processing layer, data storage layer, intelligent analysis layer, interface service layer, and application presentation layer. The overall system architecture is shown in Figure 7.

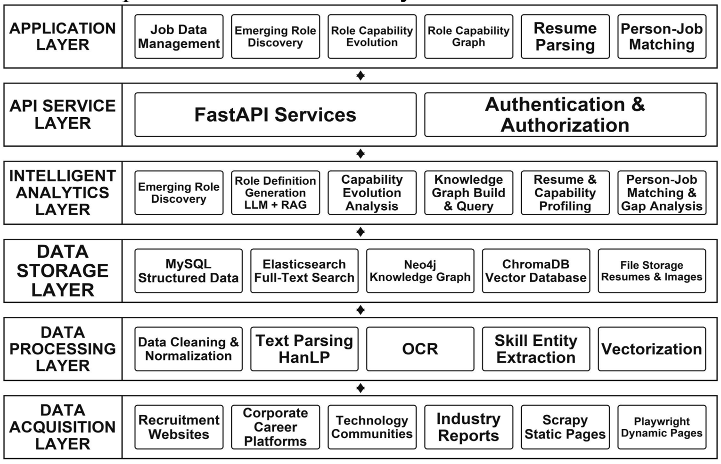

Figure 7. Overall System Architecture

6.2 System Testing and Result Analysis

To verify the functional completeness and practical application effectiveness of the system, this paper conducts tests from aspects including data acquisition, job analysis, job definition, person-job matching, and system interaction. The test results show that the system can successfully complete multi-source recruitment data acquisition; identify jobs with different names but similar competencies and generate corresponding job profiles; display changes in job skill requirements and present relationship associations through graphs; and generate job matching results based on user skills while providing competency gap analysis and learning recommendations. The system realizes a complete business process from job data acquisition and job knowledge construction to talent matching analysis, meeting the functional requirements of new job discovery, dynamic job competency updates, and person-job matching analysis. The system functional test results are shown in Table 1.

Table 1. System Functional Test Results

Test Module	Test Content	Test Result
Data Acquisition Module	Automatic Job Data Retrieval	Pass
Emerging Role Discovery Module	Emerging Role Identification	Pass
Role Definition Module	Role Profile Generation	Pass
Knowledge Graph Module	Role Relationship Visualization	Pass
Resume Parsing Module	Skill Information Extraction	Pass
Matching Module	Job Recommendation and Gap Analysis	Pass

7. Conclusion and Future Work

7.1 Work Summary

This paper addresses the issues of rapidly changing jobs and insufficient talent matching efficiency under the background of the digital economy, and designs and implements an AI job intelligent analysis platform. Based on multi-source heterogeneous job data, the system combines natural language processing, large language models, and knowledge graph technologies to realize functions including job information collection, standardized processing, new job discovery, job definition generation, job competency dynamic analysis, and person-job matching. Through multi-source data collection, the coverage of job information is improved; through knowledge graphs, job competency relationships are represented; through large language models, job analysis capabilities are enhanced; and through visual interactive methods, system usability is improved. Compared with traditional job analysis methods, the system proposed in this paper not only focuses on current job information but also analyzes job competency change trends through continuous data updates, enabling dynamic management of job knowledge and providing intelligent support for enterprise recruitment, job planning, and individual career development.

7.2 Future Work

In the future, the system can be further improved from the following aspects: first, expanding the scope of data sources by introducing more industry data, enterprise business data, and technical ecosystem data; second, enhancing job trend prediction capabilities by combining time-series analysis technologies to predict future job development directions; third, improving intelligent Agent capabilities to enable the system to automatically complete job analysis, report generation, and user consultation services. With the continuous development of artificial intelligence technologies, intelligent job analysis systems based on large language models and knowledge graphs will play an increasingly important role in enterprise talent management, vocational education, and individual career planning.

References

[1] ZHAO Peng. Deeply Exploring the Value of Archival Information and Activating Precise Person-Job Matching. Human Resources, 2026, (13): 74-75.

[2] SHAO Xintian, CHEN Junli, ZHANG Hanjv. Design of a Server Cluster Monitoring and Evaluation System Based on Elasticsearch. Computer Technology and Development, 2025, 35(9): 77-84. DOI:10.20165/j.cnki.ISSN1673-629X.2025.0120.

[3] RANGE J, SCHEMBERA B, GÖDDEKE D. Making Mathematical Knowledge Explainable, Accessible and Interoperable Through Large Language Model Integration. arXiv preprint, 2026. arXiv:2607.24512.

[4] ZHU Yan, WANG Hailin. Design of Personalized Recommendation Algorithm for Vocational Education Resources Based on Dynamic Knowledge Graphs and Reinforcement Learning . Journal of Jilin University (Information Science Edition), 1-9 [2026-07-28].

[5] FENG Li. Research on Chinese Resume Information Entity Recognition Based on Deep Learning. Wuhan Polytechnic University, 2024. DOI:10.27776/d.cnki.gwhgy.2024.000493.

[6] CHEN Cui. Research on the Application of Artificial Intelligence in Enhancing Enterprise Recruitment Matching Accuracy. Market Weekly, 2026, 39(19): 187-190.

[7] YU X, XU R Z, XUE C Y, et al. ConFit v3: Improving Resume-Job Matching with LLM-based Re-Ranking. arXiv preprint, 2026. arXiv:2605.09760.

[8] JOUANNEAU W, JOUFFROY E, PALYART M. An Efficient Long-Context Ranking Architecture With Calibrated LLM Distillation: Application to Person–Job Fit. arXiv preprint, 2026. arXiv:2601.10321.

[9] QIAN Lei, LONG Xinyue. Research on Personalized Learning Path Design and Teaching Adaptability Enabled by AIGC. China Computer & Communication, 2026, 38(16): 230-232.

[10] HUANG Liming. Research on Optimization of Multi-source Heterogeneous Data Fusion Storage Architecture. Network Security Technology & Application, 2026, (6): 85-88.

