A Résumé Evaluation System Based on Text Mining

Yi-Chi Chou
Digital Institute
Institute for Information Industry
Taipei, Taiwan ROC
gracechou@iii.org.tw

Chun-Yen Chao
Digital Institute
Institute for Information Industry
Taipei, Taiwan ROC
chunyenchao@iii.org.tw

Han-Yen Yu
Digital Institute
Institute for Information Industry
Taipei, Taiwan ROC
yuhanyen@iii.org.tw

Abstract— **This study explored the application of interview robots on recruitment process. By adopting techniques including web crawling, text mining, and natural language processing, this study developed an effective system that matches job candidates with recruiters. The designed system analyzed electronic résumés in Traditional Chinese, on which the words were graded according to the job market on the Internet and implemented with techniques related to big data. The results demonstrated that the designed system identified the current demand on talent-seeking and quickly presented candidate rankings for a specific position, thereby fulfilling the needs of both job-hunting candidates and talent-seeking recruiters.**

Keywords—Artificial intelligence (AI), big data, text mining, web crawler

I. INTRODUCTION

Artificial intelligence (AI) technology is developing rapidly and is quickly becoming a part of daily life. AI can be adopted to help people in the workplace. For example, AI can be used to assist interviewers. Applying AI in interviews is advantageous because an AI interviewer does not treat interviewees differently because of personal, mental, or physical traits or other external conditions, unlike human interviewers [1].

During interviews, interviewers tend to make unscientific or irrational decisions because of their subjective views and personal emotions; consequently, the opportunity to hire talented individuals can be missed. Because hiring an excellent staff is critical for the success of a firm, all enterprises strive to discover and hire people with considerable talent and potential. Additionally, when job applicants contact any human resources (HR) department or employer, they can generally recognize whether the company is concerned about fairly treating each applicant. The perceived level of fairness can create an impression, good or bad, of the company in the mind of an applicant. Such impressions can lead to acceptance or rejection of an offer of a second-round interview, and thus affect the opportunity for the company to recruit and hire top candidates [2].

Furthermore, candidates, including top candidates, can be affected by their physical and mental status on the date of interview. They can be nervous and underperform or experience stage fright; consequently, they may be overlooked by interviewers despite their considerable abilities. Besides, a conventional interview is limited by time and location, leading to the waste of resources by employers and interview rejections by potential candidates.

To solve this HR problem, businesses have begun to incorporate AI into HR tasks, giving rise to AI-based job matching. Gartner, a global research and advisory firm, indicated that roughly 1.8 million jobs will be replaced by AI by 2020; however, AI will also create 2.3 million jobs that expand the labor market [3]. Similar to major past labor revolutions, AI may lead to technology-related unemployment, but it may also prompt industrial transformation. Although millions of low–mid level jobs might be replaced by AI, AI will likely create more positions, including high-tech jobs, management positions, and even entry-level and low-tech jobs of a different nature.

This study developed an AI-based interviewing system to reduce the loss of talent caused by the emotional reactions and subjectivity of interviewers when viewing résumés. The designed system performs the function of résumé assessment and explores the personality traits of candidates by classifying them into four dimensions of soft power, namely dominance, influence, steadiness, and compliance (DISC) after assessing the submitted electronic résumés. This system also assesses three dimensions of competence, namely education and experience, skills, and personality traits, which are indicated by the information contained in a résumé (e.g., education, experience, specialties, and autobiography). The system examines the aforementioned data by collecting the current job market demands on the internet, performing Chinese natural language processing, and analyzing the big data relevant to the position in question. The results of this examination can help determine the quality of the match between job applicants and a business. Finally, the designed system quantifies the aforementioned DISC data and three competency dimensions by scoring each résumé. The results are then compiled into a report that contains the personal analysis, ranking, and distribution forecast for the candidate in question.

This study aimed to effectively score résumés. The resulting reports can then serve as a reference for HR departments or employers. Furthermore, job applicants can benefit because the system can prevent unfair treatment during an interview. The system not only saves businesses personnel costs but also mitigates the limitations of time and space using the Internet. The designed system produces a final report that is delivered to both the job recruiting company and job

applicant and can serve as a reference for both parties to understand each other's needs, thereby facilitating achievement of a win–win situation.

Section II briefly summarizes the application of interviewing robots and the concept of a web crawler; Section III introduces the functions, architecture and process, and interface specifications of the proposed system; Section IV describes the development environment and results; and Section V offers an analysis of the experimental results.

II. RELATED WORK

A. Web Crawler

Because information technologies (ITs) such as AI, big data, and cloud computing are thriving, IT jobs are particularly competitive in the current employment market. Firms must understand the current employment market. Adopting effective approaches that enable firms to quickly acquire targeted information is essential because online information is increasingly abundant. Researchers can perform language processing and analysis on the data collected by a web crawler. For example, Peng et al. [4] adopted the web-crawling framework Scrapy and designed a web-crawling program that was used to collect recruitment data concerning the shipping industry. Such a program is an ideal tool for researchers to conduct relevant analyses in the future.

Mbah et al. [5] used text mining to collect, analyze, and visualize local job data. The authors used web crawlers, keyword analysis, and a program written in the R language for data analysis and visualization to match job applicants with businesses. These technologies could help vocational and educational institutions to develop talent.

B. Interviewee and Interviewer

For employers, the recruitment process consumes considerable time and personnel resources. A study proposed a simplified recruitment model in which a test of mental stress was automated, and text mining was applied to create a list with applicant scores for specific jobs [6]. This approach proved to be effective in matching the personality traits, skills, and personal qualities of applicants with the positions for which they applied.

Being interviewed can be particularly difficult for recent graduates because of introversion or insufficient experience. Stanica et al. [7] incorporated virtual reality and a chatbot into an interactive system to help software engineers improve their interview skills. The results demonstrated that the system helped job candidates improve their interview performance.

AI has also been applied for HR-related positions. In this regard, AI has primarily been used to provide professional assistance to HR personnel and improve their work efficiency. For example, FirstJob, a company focusing on the recruitment of millennials, deployed an AI recruiting tool known as Mya. This tool has simplified the originally complex recruitment process and improved the efficiency of recruitment by 75% [8]. The Japanese firm Softbank Corp. introduced an AI system developed by IBM—Watson—2017. This AI system was trained to understand the corporation's recruitment standards using previous HR cases. In one second, Watson can review an average of more than 100 résumés. Its efficiency is considerably higher than that of the human recruiters who spend three minutes on average to review just one résumé [9].

III. RESEARCH METHOD

Fig. 1 depicts the architecture of the designed system, which operates in four stages: (1) The system collects phrases related to the DISC traits and three competency dimensions to build a basis for quantification and scoring in later stages. (2) The Jieba system, which is capable of performing Chinese character segmentation and filtering out stop words, is applied to preprocess the data collected in the previous stage and the text of the résumés submitted by candidates. (3) The frequency with which words appear that are related to the three competencies is calculated and the scores are corrected using a normalization formula, resulting in weights pertinent for the current job market on the Internet. These phrases and scores are then stored in a database constructed in a virtual cloud server. (4) The designed system delivers visualized data reports to users after performing big data computation. These reports can help both the job applicant and the recruiters understand the attributes and potential of the job applicant, thereby increasing the degree of appropriate job matching.
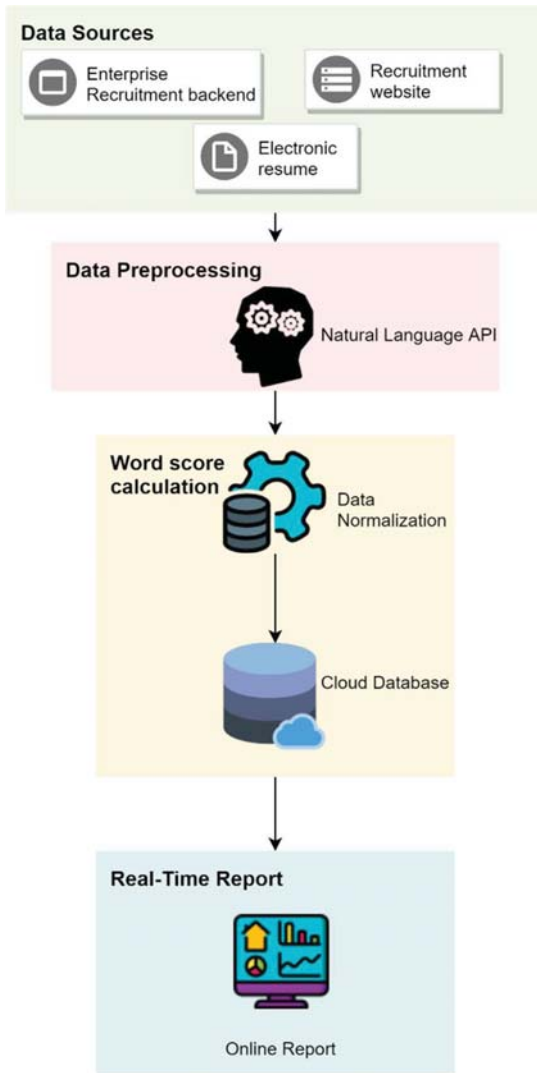
Fig. 1. System framework diagram

A. Data sources and collection

1) Sources of the DISC phrasess: In the designed system, DISC phrases were collected from the résumés received by the recruitment backend of a technology company.

2) Sources of phrases related to the three competency dimensions: This study adopted a web crawler written in Python to extract the job description and requirements posted by enterprises on the three major recruitment websites in Taiwan. The data (phrases) collected mainly comprise the recruitment information for computer engineering positions (e.g., network administrators, system engineers, data analysts, and database administrators).

B. Data preprocessing:

1) DISC and the three competency dimensions: Generally, Chinese text segmentation is more complex than English segmentation. English words are separated by spaces and are thus easier to distinguish. This study collected Chinese text related to DISC and competency dimensions and performed text segmentation and stop-words filtering using the Jieba system. The Jieba system can process a considerable amount of text data and was thus used to improve the segmentation performance. In addition to punctuation marks, numerous irrelevant words and phrases in the Chinese text must be deleted to save storage space and increase processing efficiency.

2) Electronic résumés: The received résumés were read and transformed into text using python-docx and pdfminer packages. Natural language processing was then performed. The processed phrases were compared with the phrases in the database.

C. Scoring of phrases

The frequency with which the processed phrases appeared in the recruiting market was calculated to acquire a count for each phrase. This study only retained phrases with a count of 20 or more and were relevant to the requirements for computer engineers as specified by businesses. These phrases were categorized into education and work experience, skills, and personality traits (i.e., competency dimensions). Subsequently, normalization was performed on these dimensions using Formula (1). Normalization enabled the model data to achieve convergence. Thus, the weight of each phrase under its dimension could be recognized.

$$1 + \frac{\left(x - \min(x)\right) \cdot 9}{\max(x) - \min(x)} . \qquad (1)$$

Each value in a row was made to subtract the minimum value in the row, multiplied by nine, and subsequently divided by the difference between the maximum and minimum values. The value one was then added to the resulting value. Accordingly, we obtained a new row of figures ranging between one and ten. Thus, all the eigenvalues remained under the same scale range.

D. Real-time reports

Figure 2 presents the backend architecture that was developed. A MySQL database was built in the cloud server, and concatenation was performed in hypertext preprocessor (PHP). Job applicants submitted résumés to the designed system in which Python-based algorithm was adopted for quantifying each index. Next, the quantified user data were processed using big data techniques. The resultant visualized reports were presented on the frontend interface, enabling the candidates to identify their own attributes and recruiters to recognize the attributes and potential of the candidates concerned. Therefore, the matching level between positions and suitable candidates could be enhanced.
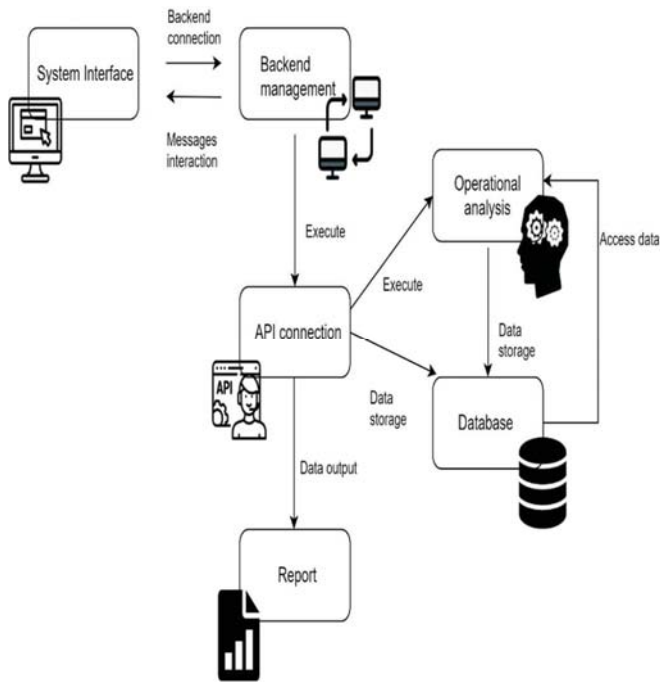
Fig. 2. Backend architecture diagram

The user interface consists of dialog boxes. When a candidate runs the system, a dialog box pops out and requests that the candidate submit a résumé. Next, a link to the analysis result is provided. The analysis result is presented in a report form (Fig. 3). Written in hypertext markup language (HTML), JavaScript, and PHP, the report contains several blocks in a top-down sequence. These blocks provide the DISC results and the ranking with a distribution forecast that accounts for the skills, education and work experience, and personality traits of the candidate (Fig. 4).
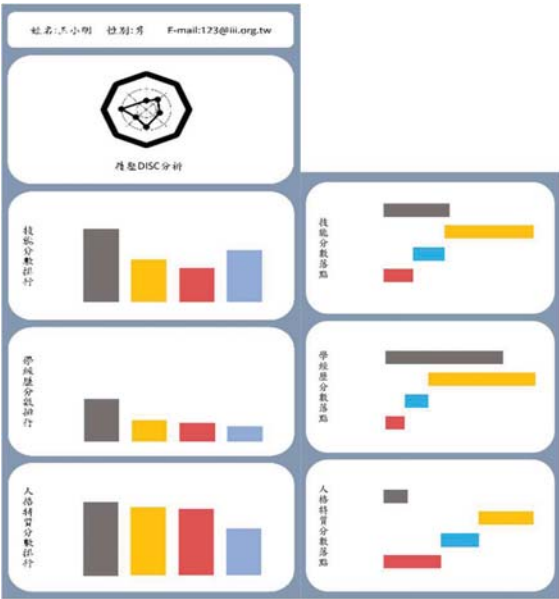

Fig. 3. Online report

Fig. 4. Report designs

E. System flow chart

Fig. 5 presents the flow chart of the designed system. After the backend receives a résumé uploaded by a candidate, the designed system uses the algorithm developed by this study to perform text mining and big data computing. The resultant report is presented to the candidate on the user interface.
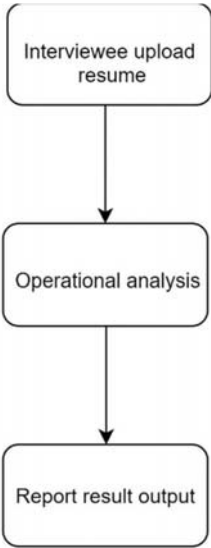

Fig. 5. System flow chart

IV. RESEARCH RESULT

A. Software development environment and software platform resources

To develop the cloud-based system, the Windows Server 2016 DataCenter was used as the development environment and XAMPP was used to develop the web server. The MySQL

database is for information collection during the process of the job applicants using the system. The data flow was developed in PHP and concatenated using JSON API. The data were connected on Chatfuel by running JSON API. At the backend, cloud computing was performed using a text-mining model in Python and text scanning techniques, which served as the tools for logical operations. By adopting the aforementioned techniques and statistical analysis of big data, an AI system was constructed. The candidate report was then presented using HTML, CSS, JavaScript, and PHP.

TABLE I. DEVELOPMENT ENVIRONMENT AND TOOLS

Development environment	Windows Server 2016 DataCenter
Web server	XAMPP v3.2.2
Database	MySQL
Development tool	Chatfuel, Sublime Text3
Backend development language	Python, PHP
Front-end development language	HTML, CSS, JavaScript

B. The system-generated final report

Fig. 6 illustrates a final report generated by the designed system, which consists of three parts: (1) The DISC distribution is indicated by a radar chart. (2) A bar chart compares the scores of the candidate concerned with those of the top three candidates among the system users with respect to skills, education and work experience, and personality traits. (3) The candidate's scores of the three competency dimensions are calculated and categorized into three groups, namely high (top 25%), moderate (26%–75%), and low (76%–100%), according to their distribution among all users.
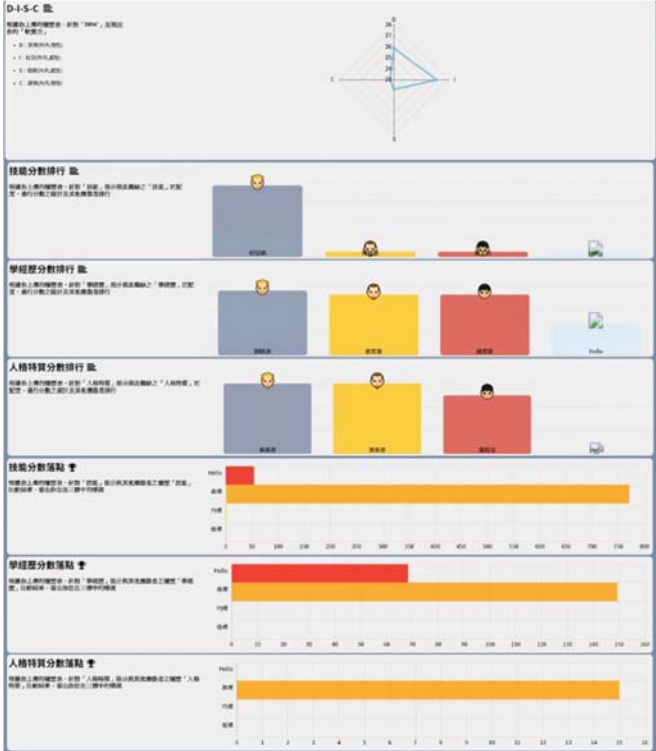

Fig. 6. Report result

V. CONCLUSION AND FUTURE WORK

In the system designed in this study, computing is performed on the basis of two models, namely DISC and the three competency dimensions. After a résumé is processed using these two models, the system produces a real-time online report that informs candidates of their soft power attributes (i.e., DISC dimensions) and competency ranking and shortcomings; this is a useful tool for self-evaluation. Recruiters can also understand job candidates through these online reports; the reports can serve as a reference for talent selection and evaluation.

In the future, to produce such reports, the designed system should not only consider recruitment websites as the source for collecting résumés and the requirements of large companies as the only reference for weighting. Therefore, further research should be conducted on how to customize reports according to the requirements of a specific firm. Thus, the company concerned can adjust the weighting standard pursuant to its required skills, personality traits, or even educational attainment to discover candidates who most closely meet its requirements.

ACKNOWLEDGEMENTS

REFERENCES

[1] Huang Jingjing, "My interviewer is AI," Jan. 1, 2018. [Online]. Available: https://www.chinatimes.com/newspapers/2018012100 0370-260209. [Accessed Nov. 12, 2018].

[2] Cevich, "Trying to get talented people on the train, the attitude of the interviewer is very important!," Aug. 4, 2016. [Online]. Available: https://www.managertoday.com.tw/columns/view/529 81. [Accessed Nov. 12, 2018].

[3] STAMFORD, Conn., "Gartner Says By 2020, Artificial Intelligence Will Create More Jobs Than It Eliminates," Dec. 13, 2017. [Online]. Available: https://www.gartner.com/newsroom/id/3837763.

[4] Dongcheng Peng, Tieshan Li, Yang Wang, and C. L. Philip Chen. "Research on Information Collection Method of Shipping Job Hunting Based on Web Crawler." In 2018 Eighth International Conference on Information Science and Technology (ICIST), 2018, pp. 57-62.

[5] Raymond Blanch Mbah, Manjeet Rege, and Bhabani Misra. "Discovering Job Market Trends with Text Analytics." In 2017 International Conference on Information Technology (ICIT), 2017, pp. 137-142.

[6] Sawleshwarkar Shreya, Rangnani Nisha, Mariwalla Vijeta, and Halbe Aparna. "Simplified Recruitment Model Using Text-Mining on Psychometric and Aptitude Tests." In 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2018, pp. 586-589.

[7] Iulia Stanica, Maria-Iuliana Dascalu, Constanta Nicoleta Bodea, and Alin Dragos Bogdan Moldoveanu. "VR Job Interview Simulator: Where Virtual Reality Meets Artificial Intelligence for Education." In 2018 Zooming Innovation in Consumer Technologies Conference (ZINC), pp. 9-12.

[8] Khari johnson, "Recruitment chatbot Mya automates 75% of hiring process," Jul. 11, 2016. [Online]. Available: https://venturebeat.com/2016/07/11/recruitment-chatbot-mya-automates-75-of-hiring-process/. [Accessed Nov. 12, 2018].

[9] Jingyuan Gao, "Will automatically detect the lie, the enterprise AI interviewer - the era of human-computer cooperation is coming," May. 6, 2018. [Online]. Available: https://www.bnext.com.tw/article/49123/lie-detectors-with-artificial-intelligence-are-future-of-border-security. [Accessed Nov. 12, 2018].