

THE UNIVERSITY OF HONG KONG
DEPARTMENT OF COMPUTER SCIENCE 2025-26



COMP4801
Final Year Project
Detailed Project Plan

BAE Junyoung (3035716464)
Supervisor: Dr. Chim Tat Wing

Topic
Automated Job Application Tracker

Table of Contents

1. Project Background.....	3
2. Project Objective.....	5
3. Project Methodology.....	7
4. Timeline and Milestones.....	9
4. References.....	11

1. Project Background

With the increasing competitiveness of the job market, students and professionals often apply to dozens of positions across different companies and platforms. Tracking the progress of each application (applied, test, interview, offer, rejection) is difficult when information is scattered across email inboxes.

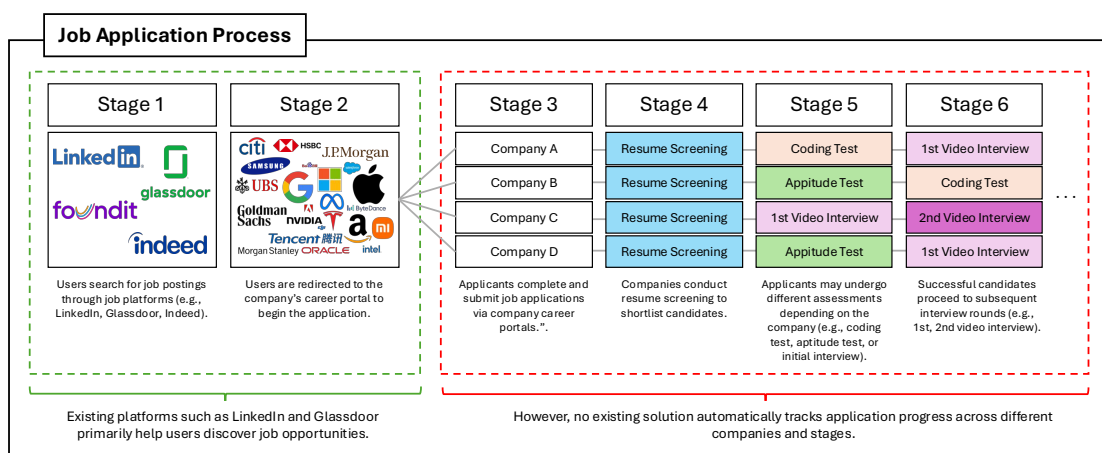


Figure 1. Job Application Process

As shown in Figure 1, the typical job application process begins with applicants searching for opportunities on platforms such as LinkedIn, Glassdoor, and Indeed (Stage 1) (Amanova, Ozgur, Msilu, & Demir, 2021). These platforms help users discover jobs but do not manage the full application process. Once a suitable job is found, applicants are redirected to the company's career portal to submit their application (Stage 2).

After submission, each company follows its own recruitment process, which often differs significantly. Some may begin with resume screening followed by an aptitude test, while others may move directly to a coding challenge or interview. This results in multiple parallel processes with varying entry points and formats, from single interviews to multi-round assessments across video, on-site, and technical formats. The fragmented nature of these processes forces applicants to track dozens of applications across inconsistent portals, often leading to confusion and missed updates.

Existing job portals such as LinkedIn or Glassdoor provide limited tracking features, but only for applications submitted through their platforms (LinkedIn, n.d.). They do not cover applications made via independent company portals or email, leaving a gap once candidates move beyond the job discovery stage.

This project addresses that gap by proposing a web application that integrates with Gmail to automatically track job-related emails. By extracting and classifying messages into recruitment stages (Applied, Screening, Test, Interview, Offer/Rejection), the system will visualize the application pipeline across companies. The focus will be on the red-marked area (Stages 3 onward), where no automated solution currently exists to help applicants keep track of progress.

With the Gmail API and AI, the system will retrieve and classify emails, generate summary statistics such as conversion rates between stages, and visualize application progress to highlight bottlenecks. In addition, a conversational chatbot will assist users with common queries, and an intelligent adviser will provide tailored recommendations and interview preparation tips.

From a business perspective, the application fills a clear market gap by offering centralized, automated tracking of applications across companies, giving job seekers a competitive edge. From a technical perspective, the project demonstrates the value of automation through Gmail API integration, intelligent filtering, and AI-driven classification, while ensuring scalability, reliability, and secure handling of personal data.

2. Project Objective

The objective of this project is to develop a web system that helps students and professionals efficiently track and manage their job applications across multiple platforms. The application integrates Gmail with AI analysis to classify and visualize the job application process. The following key features are planned to achieve this objective:

1. Gmail Integration and Email Retrieval

The system will connect to a user's Gmail account through OAuth 2.0 authentication. Once authorized, it will automatically retrieve job-related emails using Gmail's advanced query filters. This ensures that users do not need to manually search through cluttered inboxes, saving time and reducing the chances of missing important updates. The platform will work in read-only mode to ensure user privacy and security.

2. Automated Email Classification

Using AI and pattern recognition, the system will classify emails into specific stages of the recruitment pipeline such as Applied, Coding Test, Interview, Offer, or Rejection. By analyzing the sender information, subject lines, and email body content, the system will intelligently distinguish job-related emails from general communication, ensuring accurate categorization.

3. Application Journey Visualization

To make the process more intuitive, the platform will visualize the user's job application progress through timeline or tree-based diagrams. This visual representation allows users to quickly understand which stage they are in for each application and compare progress across different companies. This clarity helps job seekers prioritize their preparation and follow-ups.

4. Performance Analytics and Statistics

The system will generate key performance indicators such as conversion rates between stages (e.g., from Application to Interview) and average response times. These insights will allow users to identify personal bottlenecks, recognize strengths, and improve their job application strategies over time.

5. Conversational Chatbot Support

The application will integrate a chatbot to guide users with common job application queries. For example, the chatbot can provide explanations of recruitment stages, answer FAQs about the system, and assist with navigation within the platform. This feature aims to improve user experience by offering immediate assistance.

6. Intelligent Interview Adviser

Beyond tracking, the application will serve as a career-support tool by offering tailored recommendations for interview preparation. Leveraging AI, the adviser will suggest practice resources, highlight past patterns of user performance, and provide tips aligned with the specific stage of recruitment. This will help candidates improve their readiness and confidence before assessments and interviews.

3. Project Methodology

The methodology focuses on the technical approach to building the Job Application Tracker. It explains how emails will be collected, filtered, classified, and presented, as well as how the system will be implemented and deployed

1. Data Collection

Data collection is essential to construct the dataset for filtering and classification. The primary source of data will be the Gmail API, which provides access to email metadata and message bodies. Advanced Gmail search queries will be used to limit retrieved data to job-related communication (e.g., emails containing keywords such as *apply*, *interview*, *offer*, *recruiter*).

In addition to leveraging Gmail's built-in query filters, the project will analyze actual recruitment emails from different companies to identify domain-specific keywords, tone, and phrasing used in real hiring processes. This keyword analysis will form the basis of the rule-based classification system and will later serve as training or reference data for AI-based classifiers.

Metadata such as sender address, subject line, and date will be retrieved along with email body content. An anonymization process will be applied to store email records in a secure database, ensuring that no personally identifiable information is exposed. The collected and cleaned dataset will be used throughout the project for iterative testing and refinement.

2. Filtering and Classification System

The filtering and classification system forms the core brain of the application, responsible for distinguishing relevant job-related emails and categorizing them into recruitment stages. The first stage of filtering will rely on Gmail queries (q=), which act as a lightweight and efficient mechanism to reduce the number of irrelevant emails before deeper analysis is conducted. Once this initial step is completed, a classification pipeline will be applied. At the baseline level, rule-based matching will be used, relying on keyword detection and sender-domain checks to provide a fast and transparent method of classification. To handle more complex cases, the system will also support an AI-based classifier through integration with models such as HuggingFace transformers or GPT. These models are capable of interpreting subtle variations in email phrasing, for example distinguishing between a rejection message and an interview invitation, thereby improving overall accuracy. By combining rule-based and AI-driven methods, the system aims to balance efficiency with precision, ensuring consistent and reliable classification across the varied communication styles used by different companies.

3. Web Application Development

The web application will be designed to deliver both backend processing and frontend visualization. The backend will be implemented using Flask in Python, responsible for Gmail API integration, email retrieval, filtering, classification logic, and the provision of RESTful API endpoints for frontend consumption. The frontend will be developed with React.js and D3.js to provide an interactive user interface for visualizing the application journey. Timelines, tree-based diagrams, and summary dashboards will allow users to clearly monitor their progress across companies. For data storage, a relational database will be adopted, using SQLite during testing and PostgreSQL for production. This database will store email records, classification results, and user progress logs, with a schema designed to support scalability and secure data management.

4. Deployment

Deployment will follow a staged approach. During the development stage, Cloudflare Tunnel will be employed to securely expose the application, allowing external testing without the need for full-scale deployment. For the final deployment, the system will be hosted on a cloud service provider such as AWS or Google Cloud in order to ensure scalability, resilience, and global availability. The chosen cloud infrastructure will also support continuous integration and delivery pipelines to enable smooth updates and efficient maintenance.

5. Security

Security and privacy are critical since the application deals with personal email data. To address this, the system will rely on OAuth 2.0 authentication, which allows users to log in with their Google accounts without requiring the application to store their credentials. The Gmail API will be accessed using a minimal read-only scope (gmail.readonly) to ensure that no unauthorized modifications can be made to the user's inbox. In addition, access and refresh tokens will be stored securely with appropriate refresh mechanisms, maintaining usability while safeguarding sensitive information.

4. Timeline and Milestones

Period	Milestone / Deliverable
October 1, 2025	Deliverables of Phase 1 (Inception) <ul style="list-style-type: none"> - Submit detailed project plan. - Initiate project webpage design (basic layout, navigation). - Define scope and requirements with advisor feedback.
October, 2025	<ul style="list-style-type: none"> - Collect job application email data samples (pilot dataset) and perform keyword analysis from real recruitment emails. - Create initial rule-based model to detect job-related messages and status indicators. - Draft system architecture (frontend–backend–database). - Set up cloud infrastructure for development (Cloudflare Tunnel, GitHub repo).
November – December, 2025	<ul style="list-style-type: none"> - Design relational database schema. - Implement backend structure with API handling - Implement backend with Flask: Gmail API integration, email fetching, and classification pipeline. - Implement frontend prototype in React.js with basic visualization (timeline/tree diagram). - Integrate backend and frontend through RESTful API. - Conduct unit testing of core modules (data retrieval, filtering, visualization).
January 12 – 17, 2026	First presentation <ul style="list-style-type: none"> - Present initial visualization of application pipeline. - Collect supervisor feedback for improvement.
January 25, 2026	<ul style="list-style-type: none"> - Preliminary implementation. - Write detailed interim report.
April 19, 2026	Deliverables of Phase 3 (Construction) <ul style="list-style-type: none"> - Finalized tested implementation. - Submission of final report. - Publish project web page (final version).

April 20 – 25, 2026	Final presentation.
April 22, 2026	Poster exhibition Present project poster. - Submit 1-minute video demonstration of the system.

4. References

Amanova, C., Ozgur, S. B., Msilu, F., & Demir, F. (2021). Demystifying the user experience: A case study on online job search engines. *American Journal of Qualitative Research*, 5(2), 233–250. <https://doi.org/10.29333/ajqr/11411>

LinkedIn. (n.d.). *How to use candidate search filters*. LinkedIn Help. Retrieved September 19, 2025, from <https://www.linkedin.com/help/recruiter/answer/a7791507>