

# Báo cáo Research về cách kéo CV từ các nguồn về System pool

## I. Tổng quát:

- **Mục đích:** Tìm hiểu cách để có thể lấy data về CV của các ứng cử viên từ nhiều nguồn về system pool của bên mình để có thể sử dụng tùy vào nhu cầu.
- **Phương hướng tiếp cận:** Tìm hiểu về những trang web liên quan tới việc làm, xem xét những chính sách hỗ trợ cũng như dịch vụ của họ để xem liệu có cách nào thu thập CV của các ứng cử viên đã sử dụng dịch vụ của họ hay không.
- **Thông tin về các nguồn tìm hiểu:**

Nguồn	Người nghiên cứu
Base vn	Hoàng Thiên Ân
Gmail/Email	Nguyễn Hải Quân
TopCV	Nguyễn Hải Quân

## II. Chi tiết về các nguồn:

---

### Nguồn 1: Gmail/Email – Thu thập CV qua địa chỉ email ứng tuyển

---

#### 1. Mô tả Nguồn (Source Description)

- Gmail/Email là gì?**  
Gmail là nền tảng thư điện tử phổ biến nhất hiện nay với giao diện đơn giản, khả năng tích hợp cao, đặc biệt trong hoạt động tuyển dụng. Ứng viên thường gửi hồ sơ/CV tới email tuyển dụng của doanh nghiệp (vd: [hr@hiptechvn.com](mailto:hr@hiptechvn.com)).
  - Đối tượng sử dụng chính:**
    - Ứng viên (ứng tuyển qua email)
    - Doanh nghiệp (nhận CV, phỏng vấn)
    - Nhà tuyển dụng dùng ATS (Applicant Tracking Systems)
- 

#### 2. Khả năng tích hợp (Integration Feasibility)

- API chính thức:**  Có
    - Gmail API từ Google Workspace cung cấp quyền truy xuất email, đính kèm, tiêu đề...
    - Docs: [Gmail API - Google Developers](#)
  - Phương pháp tích hợp thay thế:**
    - IMAP Protocol (chuẩn mực):** Hầu hết các hệ thống email đều hỗ trợ IMAP.
    - Email Parsing:** Trích xuất thông tin từ nội dung email hoặc tệp đính kèm (PDF/DOCX).
  - Cơ chế xác thực:**
    - OAuth 2.0** (Google Cloud Console, Gmail API)
    - IMAP/SMTP yêu cầu tài khoản và mật khẩu ứng dụng (App Password)
- 

#### 3. Phương pháp tích hợp đề xuất (Proposed Integration Method)

Thành phần	Mô tả
Luồng dữ liệu	Gmail → Email Parser → CV Extractor → System Pool (DB hoặc ATS)
Tần suất xử lý	Theo thời gian thực (webhook) hoặc Cron job 5–15 phút/lần
Định dạng dữ liệu	PDF, DOCX, RTF, hoặc đơn giản là nội dung text

Công nghệ gọi ý	Python (IMAP/SMTP), Node.js (Nodemailer), PDFMiner, python-docx, spaCy
Công cụ hỗ trợ	Zapier, Integromat, Mailparser.io, AWS Lambda, Google Cloud Functions

## 4. Ưu điểm (Advantages)

Tiêu chí	Đánh giá
Quy mô nguồn ứng viên	⭐⭐⭐⭐ – Gần như toàn bộ ứng viên đều dùng email để gửi CV
Chất lượng dữ liệu	⭐⭐⭐⭐ – CV thường đầy đủ thông tin, đính kèm chuẩn
Dễ tích hợp	⭐⭐⭐⭐⭐ – Có thư viện và tài liệu sẵn, dễ gỡ lỗi và bảo trì
Chi phí	⭐⭐⭐⭐⭐ – Miễn phí nếu dùng IMAP/SMTP; Gmail API miễn phí trong mức giới hạn

## 5. Hạn chế / Nhược điểm (Drawbacks / Limitations)

Vấn đề	Mô tả
Giới hạn rate/quota	Gmail API có hạn mức (vd: 1 triệu request/ngày cho Gmail API free tier)
Trích xuất khó khăn	Một số CV bị scan sai hoặc không đúng cấu trúc dễ trích xuất
Trùng lặp dữ liệu	Gửi lại nhiều lần → cần chống trùng (hash tệp, email ID...)
Cần parser CV phức tạp	OCR, NLP để hiểu và phân loại nội dung

## 6. Pháp lý và Bảo mật (Legal & Privacy Considerations)

Vấn đề	Đề xuất xử lý
Yêu cầu sự đồng ý người dùng	<input checked="" type="checkbox"/> Cần đưa thông báo rõ ràng trong JD hoặc email xác nhận
Điều khoản dịch vụ Gmail	Phải tuân thủ <a href="#">Google API Services User Data Policy</a>
Tuân thủ bảo vệ dữ liệu	Áp dụng tiêu chuẩn GDPR (nếu hoạt động quốc tế), lưu log truy cập, mã hóa CV

## 7. Chi phí & Bản quyền (Cost / Licensing)

Mục tiêu	Chi tiết
Gói miễn phí	<input checked="" type="checkbox"/> Có. Gmail API miễn phí với hạn mức phù hợp nhu cầu vừa và nhỏ
Chi phí nâng cao	Google Workspace Business: từ \$6/user/tháng (tùy doanh nghiệp)

Cần hợp đồng doanh nghiệp?

✗ Không bắt buộc (trừ khi yêu cầu SLA cao hoặc tích hợp quy mô lớn)

## 8. Độ phức tạp khi tích hợp (Integration Complexity)

- **Mức độ: Trung bình (Medium)**
- **Giải thích:**
  - Cần xử lý OAuth 2.0, email MIME, parser file đính kèm, bảo mật nội dung
  - Tuy nhiên có rất nhiều công cụ và thư viện hỗ trợ mạnh

Ví dụ kiến trúc kỹ thuật gợi ý:

```
[Ứng viên]
↓
[Gửi CV tới Gmail hr@hiptechvn.com]
↓
[Email Crawler Module] ← Gmail API hoặc IMAP Polling
↓
[CV Attachment Parser] → Trích xuất text, thông tin cá nhân
↓
[NLP + Cleanser] → Làm sạch thông tin, chuẩn hóa định dạng
↓
[System Pool Database] → PostgreSQL / MongoDB / Elasticsearch
```

*Kết luận:*

Gmail là một trong những nguồn thu thập CV linh hoạt nhất, tiết kiệm chi phí nhất, dễ triển khai và có thể mở rộng dễ dàng bằng cách kết nối nhiều email tuyển dụng. Rất nên triển khai sớm trong bất kỳ hệ thống ATS hay Talent Pool nội bộ nào.

---

## Nguồn 2: E-Hiring của Base.vn – Thu thập CV ứng tuyển từ nhiều jobsite

---

### 1. Mô tả Nguồn (Source Description)

- E-Hiring của Base.vn là gì?**  
E-Hiring là hệ thống quản lý tuyển dụng cho phép các doanh nghiệp quản lý quy trình nhận và xử lý CV  
Nhà tuyển dụng thông qua E-Hiring có thể thu thập CV từ email và các jobsite đã được tích hợp vào hệ thống của E-Hiring như là VietnamWorks, TopCV.
  - Đối tượng sử dụng chính:**  
Chuyên gia nhân sự, các nhà quản lý tuyển dụng, các công ty doanh nghiệp vừa và nhỏ ở Việt Nam
- 

### 2. Khả năng tích hợp (Integration Feasibility)

- API chính thức:** **Có**
    - Base Public API được phát triển để hỗ trợ việc tích hợp và trao đổi dữ liệu giữa Base và các hệ thống khác.  
Docs: [Base Public API, E-Hiring](#)
  - Cơ chế xác thực:**
    - Về phía Base.vn: Access Token được tạo bằng cách nhập App ID
    - Về phía các website tích hợp: Tùy thuộc vào từng bên jobsite khác nhau mà sẽ có các yêu cầu khác nhau về các Access Token, Secret Key hoặc Oauth Token.
- 

### 3. Phương pháp tích hợp đề xuất (Proposed Integration Method)

Thành phần	Mô tả
Luồng dữ liệu	CV ứng tuyển từ Jobsites → E-Hiring → API Calls → System Pool
Tần suất xử lý	Hệ thống có 4 thời điểm quét thông tin định kỳ: 07:00, 13:00, 20:00, 23:45
Định dạng dữ liệu	Kết quả trả về của API sẽ bao gồm đường link dẫn đến file CV của candidates. Tùy thuộc vào định dạng của file CV được upload lên dữ liệu (docx, pdf,...)

---

### 4. Ưu điểm (Advantages)

Tiêu chí	Đánh giá
Quy mô nguồn ứng viên	Hỗ trợ tích hợp với hầu hết các nguồn jobsites tại Việt Nam như là Vietnamworks, TopCV, ITViec, Timviecnhanh,...
Quản lý dữ liệu	Thông tin về các ứng cử viên được gọi về từ API có thông tin về jobsite nơi mà họ ứng tuyển
Dễ tích hợp	Có thư viện và tài liệu của các API endpoints cho từng dịch vụ của Base.vn. Đồng thời sử dụng hệ thống Base.vn có để giảm bớt việc cài đặt tích hợp từ nhiều nguồn khác nhau

## 5. Hạn chế / Nhược điểm (Drawbacks / Limitations)

Vấn đề	Mô tả
Giới hạn rate/quota	Hiện tại Base.vn chưa cung cấp thông tin về giới hạn về việc sử dụng API nên cần phải cân nhắc thêm.
Dữ liệu thiếu nhất quán	Vì E-Hiring hỗ trợ tích hợp nhiều jobsites khác nhau nên dẫn đến việc kết quả trả về cũng sẽ khác nhau.
Giới hạn về API endpoint	Có thể sẽ không hỗ trợ một số API đáp ứng các yêu cầu cụ thể hơn.

## 6. Pháp lý và Bảo mật (Legal & Privacy Considerations)

Nội dung	Yêu cầu
Yêu cầu sự đồng ý người dùng	Yêu cầu người sử dụng chấp nhận chia sẻ dữ liệu theo chính sách của họ
Điều khoản dịch vụ Base.vn	Phải tuân thủ <a href="#">Điều khoản sử dụng Base Enterprise</a>
Tuân thủ bảo vệ dữ liệu	Không đăng tải nội dung bất hợp pháp và xâm phạm quyền sở hữu trí tuệ

## 7. Chi phí & Bản quyền (Cost / Licensing)

Mục tiêu	Chi tiết
Gói miễn phí	Base E-Hiring không cung cấp dịch vụ miễn phí
Chi phí	Tối 850k một tháng cho gói dịch vụ cơ bản
Phí API của các trang việc làm khác	Tuy E-Hiring hỗ trợ tích hợp nhiều jobsites để tổng hợp thông tin, nhưng mỗi trang jobsites sẽ có thể phát sinh những chi phí riêng cho dịch vụ của họ

## 8. Độ phức tạp khi tích hợp (Integration Complexity)

- Mức độ: Trung bình (Medium)**
- Giải thích:**
  - Cần phải đăng ký dịch vụ tại nhiều website dẫn đến việc xử lý, cài đặt API cho từng nguồn sẽ khác nhau để có thể tích hợp vào E-Hiring
  - Tuy nhiên vì Base.vn hỗ trợ việc tích hợp nhiều cho nhiều jobsites nên chỉ cần thiết kế hệ thống phù hợp để kéo data về từ Base.vn thay vì nhiều nguồn.

---

### Ví dụ kiến trúc kỹ thuật gợi ý:

[Ứng viên]

↓  
[Gửi CV tới các jobsites được công ty đăng tuyển]

↓  
[Base.vn]

↓  
[CV Attachment from API]

↓  
[System Pool Database]

---

### Kết luận:

**Base E-Hiring** là một dịch vụ dùng để quản lý nhân sự, có hỗ trợ cho việc sử dụng API để thu về CV của các ứng cử viên đã gửi vào các trang web khác nhau mà bên phía công ty đã đăng tải. Nhờ vào khả năng tích hợp với nhiều jobsites khác nhau của Base.vn, việc thiết xử lý API trong hệ thống có thể được rút gọn.

---

## Nguồn 3: TopCV.vn – Khai thác CV từ hệ thống tuyển dụng lớn tại Việt Nam

---

### 1. Mô tả Nguồn (Source Description)

- TopCV là gì?**  
TopCV là một nền tảng tuyển dụng nổi bật tại Việt Nam, nổi tiếng với hệ sinh thái cho phép người dùng tạo CV chuyên nghiệp, nộp đơn, và xây dựng hồ sơ cá nhân. Tính đến năm 2024, TopCV có hơn **6.000.000 ứng viên** và hàng ngàn doanh nghiệp đang sử dụng dịch vụ tuyển dụng.
  - Đối tượng sử dụng chính:**
    - Ứng viên (cá nhân đăng CV, tìm việc)
    - Nhà tuyển dụng (doanh nghiệp dùng TopCV for Business)
    - Doanh nghiệp/tổ chức cần khai thác dữ liệu nhân sự
- 

### 2. Khả năng tích hợp (Integration Feasibility)

Tiêu chí	Đánh giá
API chính thức	<input checked="" type="checkbox"/> Có – TopCV cung cấp API qua nền tảng <a href="#">TopCV for Business</a>
Phương thức tích hợp khác	Có thể kết hợp <b>webhook, RSS feed việc làm</b> , hoặc <b>manual export</b>
Yêu cầu xác thực	Cần tài khoản doanh nghiệp, <b>API Key</b> hoặc OAuth (theo mô hình client-server)
Tài liệu kỹ thuật	Không công khai toàn bộ online. Cần liên hệ trực tiếp TopCV để được cung cấp API docs.

---

### 3. Phương pháp tích hợp đề xuất (Proposed Integration Method)

Thành phần	Mô tả
Luồng dữ liệu	TopCV API → Job Post Responses → Trích xuất CV → System Pool
Tần suất xử lý	Gợi ý: cron job mỗi 15 phút / webhook sự kiện nộp CV
Định dạng dữ liệu	JSON (profile), PDF/DOCX (CV), metadata (position, location...)
Công cụ gọi ý	Node.js + Axios, Python Requests, AWS Lambda, PostgreSQL
API endpoint mẫu	/api/candidate/{id}/cv (giả định, cần xác minh sau khi có API doc)

---

### 4. Ưu điểm (Advantages)

Yêu tố	Đánh giá
Số lượng ứng viên lớn	 – Hơn 6 triệu hồ sơ từ mọi ngành nghề tại Việt Nam
Dữ liệu chất lượng cao	 – Hồ sơ chuẩn hóa, nhiều người tạo CV trực tiếp trên TopCV
Để tích hợp	 – Cần có tài khoản và hợp tác nhưng có hỗ trợ kỹ thuật
Chi phí linh hoạt	 – Có gói miễn phí giới hạn, gói trả phí nâng cao tính năng truy xuất

## 5. Nhược điểm / Giới hạn (Drawbacks / Limitations)

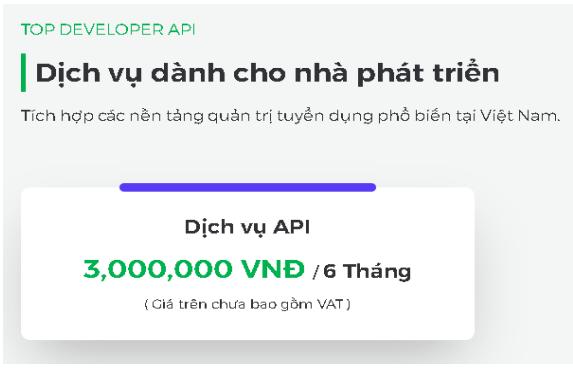
Nhược điểm	Ghi chú
Giới hạn truy cập CV pool	Chỉ được xem hồ sơ công khai hoặc ứng viên nộp đơn vào job của bạn
Cần gói doanh nghiệp trả phí	Truy cập API và CV pool cần gói tuyển dụng trả phí hoặc hợp tác B2B
Chưa công khai API đầy đủ	Phải liên hệ trực tiếp để lấy tài liệu kỹ thuật
Rủi ro trùng lặp CV	Cần có cơ chế hash/check ID hồ sơ để chống duplicate khi sync nhiều nguồn

## 6. Pháp lý và Bảo mật (Legal & Privacy Considerations)

Nội dung	Mức độ
Yêu cầu sự đồng ý người dùng	<input checked="" type="checkbox"/> Có – Ứng viên cần đồng ý hiển thị thông tin với nhà tuyển dụng
Tuân thủ TOS	<input checked="" type="checkbox"/> Phải tuân thủ <a href="#">Điều khoản sử dụng TopCV</a>
Bảo vệ dữ liệu cá nhân (PDPA)	<input checked="" type="checkbox"/> Cần cam kết không chia sẻ, bán lại dữ liệu khi sử dụng CV Pool

## 7. Chi phí & Bản quyền (Cost / Licensing)

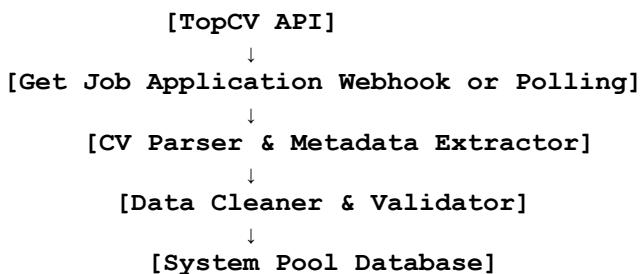
Mục tiêu	Thông tin
Gói miễn phí	Có – Nhưng hạn chế số lượt xem và hồ sơ ứng viên
Gói trả phí phổ biến	Từ khoảng <b>4.000.000–15.000.000 VND/tháng</b> (theo số lượng job/CV view)
Truy cập API	Chỉ trong gói doanh nghiệp → cần liên hệ tư vấn TopCV
Cần hợp đồng B2B?	<input checked="" type="checkbox"/> Có – Nhằm truy cập API và hỗ trợ SLA kỹ thuật



## 8. Độ phức tạp khi tích hợp (Integration Complexity)

- **Mức độ: Trung bình đến cao**
- **Vì sao?**
  - Cần đàm phán và đăng ký doanh nghiệp
  - Không có tài liệu API công khai → phụ thuộc hỗ trợ TopCV
  - Sau khi có quyền, luồng tích hợp tương đối dễ triển khai

### Sơ đồ luồng tích hợp để xuất



### Kết Luận

TopCV là một nguồn rất mạnh nếu muốn truy cập vào “CV pool toàn thị trường”, thay vì chờ ứng viên ứng tuyển vào job. Tuy nhiên, để tích hợp, **bắt buộc phải hợp tác doanh nghiệp và trả phí gói tuyển dụng nâng cao**.

Nếu đang build **CV Aggregator System**, TopCV nên là **một trong những nguồn tích hợp cốt lõi**, bên cạnh Gmail.

### **III. Tổng kết:**

Sau khi nghiên cứu các nguồn chính để thu thập và tích hợp dữ liệu CV về system pool, có thể rút ra một số điểm chính như sau:

#### **1. Nguồn Gmail/Email**

- Là kênh phổ biến và hầu như doanh nghiệp nào cũng đã dùng để nhận CV.
- Chi phí rất thấp (thậm chí miễn phí) và dễ triển khai nhờ Gmail API hoặc IMAP.
- Cần đầu tư xây dựng parser để xử lý các định dạng file đính kèm và làm sạch dữ liệu.
- Rất phù hợp triển khai trước để xây “pipeline” gom dữ liệu cơ bản.

#### **2. Nguồn Base.vn (E-Hiring)**

- Cho phép tập trung CV từ nhiều jobsite lớn tại Việt Nam (VietnamWorks, TopCV, ITViec, v.v.).
- Có API chính thức và lịch quét dữ liệu định kỳ rõ ràng.
- Mất phí thuê bao hàng tháng (~850.000 VNĐ/tháng) nhưng tiết kiệm công sức tích hợp từng jobsite riêng lẻ.
- Phù hợp cho công ty vừa và nhỏ cần quản lý ứng viên từ nhiều nguồn.

#### **3. Nguồn TopCV.vn**

- Là nền tảng lớn với hơn 6 triệu hồ sơ ứng viên.
- Có API cho doanh nghiệp nhưng yêu cầu gói trả phí và hợp đồng B2B.
- Cần phối hợp trực tiếp để lấy tài liệu kỹ thuật và quyền truy cập.
- Giá cao hơn (4–15 triệu/tháng) nhưng tiếp cận kho hồ sơ rất rộng và chất lượng dữ liệu tốt.

#### **Lưu ý:**

- Gmail/Email và Base.vn không hỗ trợ việc thu thập việc kéo CV từ các ứng cử viên của các công ty khác mà chỉ có thể nhân được CV gửi đến phía công ty mình.
- TopCV khả thi trong việc thu thập CV từ các ứng viên của các công ty khác, nhưng có thể sẽ cần phải có thảo luận với bên họ vì dịch vụ API chưa được công khai chi tiết.
- Việc lấy CV của các ứng viên mà không thuộc công ty của mình là khá khó khăn vì vấn đề pháp lý bảo mật thông tin người dùng và rất ít nguồn hỗ trợ cung cấp thông tin này một cách công khai.

#### **Mở rộng:**

- Có thể đổi hướng tiếp cận bằng cách thay vì nhắm tới CV của ứng viên, chúng ta có thể lấy một số thông tin mà họ cung cấp. Tại một số website như là Indeed, Linked, sẽ khuyến khích người sử dụng công khai những thông tin như kỹ năng, kinh nghiệm làm việc, học vấn, giới thiệu

về bản thân,... Nếu có thể tổng hợp những thông tin này từ các nguồn API họ cung cấp, chúng ta sẽ có được một kết quả trả về gần tương đương với bản CV hoàn chỉnh.

- Vẫn còn một số website chưa nghiên cứu nên việc kéo CV từ các nguồn khác vẫn có thể khả dụng ngoài dịch vụ của TopCV. Nên vẫn còn cần phải tìm hiểu thêm.

- Vì vẫn còn hạn chế trong việc truy cập vào các dịch vụ của những nguồn đã đề cập, nên có thể vẫn chưa thể nắm bắt rõ được những dịch vụ mà họ cung cấp.