

# Compte rendu — Workflow n8n : Recherche de stages à l'étranger

Leo Torres — DO3

19 janvier 2026

## Table des matières

<b>1 Contexte et objectifs</b>	<b>2</b>
1.1 Objectif du projet . . . . .	2
1.2 Périmètre et hypothèses . . . . .	2
<b>2 Architecture du workflow n8n</b>	<b>2</b>
2.1 Vue d'ensemble du workflow . . . . .	2
2.2 Description des étapes . . . . .	2
2.2.1 Étape 1 — Génération des combinaisons ville × technologie . . . . .	2
2.2.2 Étape 2 — Recherche web (Tavily) . . . . .	2
2.2.3 Étape 3 — Enrichissement scoring via MCP . . . . .	2
2.2.4 Étape 4 — Extraction d'informations via LLM . . . . .	3
2.2.5 Étape 5 — Génération des résumés via LLM . . . . .	3
2.2.6 Étape 6 — Filtrage personnalisé . . . . .	3
2.2.7 Étape 7 — Export des résultats . . . . .	3
2.2.8 Étape 8 — Synthèse finale . . . . .	3
2.2.9 Étape 9 — Notification . . . . .	3
<b>3 Exemples de résultats</b>	<b>3</b>
3.1 Offres extraites (5 à 10 exemples) . . . . .	3
3.2 Fichier CSV exporté (ou capture) . . . . .	4
3.3 Workflow n8n (JSON) dans le repo . . . . .	4
<b>4 Bonus — Méthode de scoring</b>	<b>5</b>
4.1 Méthode utilisée . . . . .	5
4.2 Choix de scoring et justification . . . . .	5
4.3 Évaluation : est-ce que ça fonctionne comme attendu ? . . . . .	5
<b>5 Analyse critique</b>	<b>5</b>
5.1 Difficultés rencontrées . . . . .	5
5.2 Critères de filtrage choisis et justification . . . . .	5
5.3 Apports du workflow et axes d'amélioration . . . . .	6
<b>6 Conclusion</b>	<b>6</b>
<b>A Annexes</b>	<b>6</b>
A.1 Prompts LLM (extraits) . . . . .	6
A.2 Exemple de sortie JSON complète . . . . .	7

# 1 Contexte et objectifs

## 1.1 Objectif du projet

Le workflow automatise la recherche d'offres de stage à l'étranger à partir d'une matrice  $\text{ville} \times \text{technologie}$ . Les résultats sont enrichis par un score de ville (MCP), normalisés par LLM, dédoublonnés, puis exportés vers Google Sheets et résumés dans une synthèse envoyée sur Discord.

## 1.2 Périmètre et hypothèses

- Villes : Berlin, Stockholm.
- Technologies : crypto, devops, cybersecurity, iot (8 combinaisons).
- Source principale : recherche web via Tavily (1 résultat par requête).
- LLM : Google Gemini (gemma-3-4b-it pour extraction, gemma-3-27b-it pour résumé, gemini-robotics-er-1.5-preview pour synthèse) (à changé en fonction de la disponibilité des quotas gratuits).
- Limites : dépendance aux API, champs souvent absents (salaire, remote), redondances d'URLs.

# 2 Architecture du workflow n8n

## 2.1 Vue d'ensemble du workflow

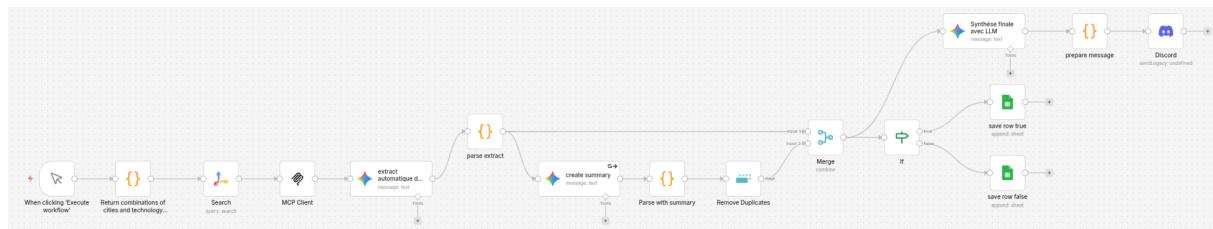


FIGURE 1 – Workflow n8n complet.

## 2.2 Description des étapes

### 2.2.1 Étape 1 — Génération des combinaisons ville × technologie

Un nœud Code génère les couples ville x technologie à partir de deux tableaux statiques.

### 2.2.2 Étape 2 — Recherche web (Tavily)

Chaque couple déclenche une recherche Tavily avec la requête type `company <tech> <city> internship`. La recherche est en `basic`, avec `max_results=2`.

### 2.2.3 Étape 3 — Enrichissement scoring via MCP

Le MCP Client appelle l'outil `find_city_score` avec le nom de ville. Si la ville est connue, un score est renvoyé et ajouté aux données d'entrée ; sinon le score est absent et le MCP indique `found=false`.

#### **2.2.4 Étape 4 — Extraction d'informations via LLM**

Le LLM retourne un JSON strict, sans texte additionnel, conforme à un schéma fixe (company, job\_title, city, country, remote\_policy, contract\_type, salary, currency, duration, application\_deadline, skills, languages, source\_url, source\_title, city\_score). Les valeurs manquantes sont forcées à null.

#### **2.2.5 Étape 5 — Génération des résumés via LLM**

Un second LLM produit un résumé FR de 2 à 3 phrases par offre. Les sorties non-JSON sont nettoyées et parsees dans un node Code.

#### **2.2.6 Étape 6 — Filtrage personnalisé**

Le filtrage est minimal : dedoublonnage par source\_url, puis routage selon la présence d'un salaire pour séparer deux onglets de stockage.

#### **2.2.7 Étape 7 — Export des résultats**

Les offres sont écrites dans Google Sheets (équivalent CSV), dans deux onglets en fonction du filtre (pour les offres ok selon le filtre et pour les autres). Colonnes : company, job\_title, contract\_type, city, country, salary, currency, duration, application\_deadline, skills, languages, source\_url, source\_title, city\_score, summary, remote\_policy.

#### **2.2.8 Étape 8 — Synthèse finale**

Un LLM génère une synthèse lisible avec un titre, le nombre d'offres, un top 3 argumenté et des statistiques globales (villes, contrats, remote, salaires).

#### **2.2.9 Étape 9 — Notification**

La synthèse est envoyée sur Discord via webhook. Un node Code tronque le message à 1900 caractères pour respecter la limite.

### **3 Exemples de résultats**

#### **3.1 Offres extraites (5 à 10 exemples)**

Exemple de 5 à 10 offres extraites (toutes les infos dans le JSON) :

FIGURE 2 – Exemple de 5–10 offres extraites (JSON).

### 3.2 Fichier CSV exporté (ou capture)

internships																	summary
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
	company	job_title	city	country	remote_policy	contract_type	salary	currency	duration	application_deadline	skills	languages	source_url	source_file	city_score	summary	
1	Cryptid	Web3 Internship Stockholm	Sweden	Sweden	unknown	internship	\$30k/year to \$120k/year	€	1 year	2024-06-30	[Blockchain, cryptid]		<a href="https://github.com/Cryptid-Team/cryptid-internship">https://github.com/Cryptid-Team/cryptid-internship</a>	https://github.com/Cryptid-Team/cryptid-internship	0.78932030	Cryptid propose un stage d'internat en Blockchain et cryptomonnaie à Stockholm. Il s'agit d'un contrat de stage avec	
2	CertiGlobe	Web3 Internship Barcelona	Spain	Spain	unknown	internship	\$30k/year to \$120k/year	€	1 year	2024-06-30	[Blockchain, cryptid]		<a href="https://github.com/CertiGlobe/certi-globe-web3-internship">https://github.com/CertiGlobe/certi-globe-web3-internship</a>	https://github.com/CertiGlobe/certi-globe-web3-internship	0.98383	CertiGlobe propose des stages Web3 à Barcelone, en Espagne. Il s'agit d'un contrat de stage avec	
3																	
4																	
5																	
6																	
7																	
8																	
9																	
10																	
11																	
12																	
13																	
14																	
15																	
16																	
17																	
18																	
19																	
20																	
21																	
22																	
23																	
24																	
25																	
26																	
27																	
28																	
29																	
30																	
31																	
32																	
33																	
34																	
35																	
36																	
37																	
38																	
39																	
40																	
41																	
42																	
43																	
44																	
45																	
46																	
47																	
48																	

FIGURE 3 – Aperçu de l'export CSV / Google Sheets.

### 3.3 Workflow n8n (JSON) dans le repo

```
1  repo/
2    workflow-final.json
```

## 4 Bonus — Méthode de scoring

### 4.1 Méthode utilisée

- Source : je me suis appuyee sur le document de reference present dans `docs/` (Kaklauskas et al., 2018, colonne `Sj`).
- Analyse : j'ai derive un score comparable entre villes a partir de `Sj` pour obtenir une echelle 0–100 (plus lisible).
- Acces via MCP : outil `find_city_score` expose par le serveur MCP.
- Normalisation : suppression d'accents, casse, ponctuation et alias courants (ex. *Krakow*).
- Valeurs manquantes : si la ville n'est pas trouvée, le MCP renvoie `found=false` et aucun score.

Calcul utilise :

$$\text{city\_score} = \left( 1 - \frac{Sj - \min(Sj)}{\max(Sj) - \min(Sj)} \right) \times 100$$

Ainsi, un `Sj` faible (forte concordance) donne un score plus eleve, tandis qu'un `Sj` eleve est penalise.

Je me suis aidé de l'IA afin de trouver le calcul pour rendre le calcul plus lisible, précis et cohérent.

### 4.2 Choix de scoring et justification

- Pertinence : indicateur externe pour comparer rapidement l'attractivite des villes.
- Simplicite : table pre-calculée, sans besoin de recalcul complexe.
- Robustesse : normalisation et aliases limitent les erreurs de correspondance.

### 4.3 Évaluation : est-ce que ça fonctionne comme attendu ?

Pour les villes presentes dans la table, le score est coherent et stable. Les limites viennent des villes absentes ou des libelles atypiques (ex. quartiers, regions). Les aliases ameliorent le taux de correspondance, mais un enrichissement geographique serait utile.

## 5 Analyse critique

### 5.1 Difficultés rencontrées

- Choix d'un LLM gratuit : trouver un modele avec un plan gratuit, assez rapide et suffisamment performant pour des extractions fiables.
- Parsing et fusion : difficultes a parser les sorties LLM puis a merger les flux pour n'obtenir qu'un seul objet exploitable.
- Webhook Discord : configuration et tests pour respecter les contraintes de taille des messages.
- Google Sheets : configuration Google Cloud API (OAuth) plus longue que prevu avant de pouvoir ecrire dans un tableur.

### 5.2 Critères de filtrage choisis et justification

Le filtrage est volontairement leger pour maximiser la couverture : dedoublonnage par URL et separation selon `salary` present ou non. Cela permet d'identifier rapidement les offres les plus informatives sans exclure trop de candidats.

### 5.3 Apports du workflow et axes d'amélioration

- Gain de temps important sur la collecte et la mise en forme des offres.
- Ameliorations : augmenter `max_results`, ajouter des sources, filtrer par `city_score`, enrichir la deduplication, meilleure gestion d'erreurs et retries.

## 6 Conclusion

Le workflow automatise efficacement la recherche et la synthese d'offres de stage. L'extraction structuree et l'export permettent une analyse rapide, tandis que le scoring MCP apporte un signal supplementaire. Les principaux axes d'amelioration concernent la couverture des sources et la robustesse face aux donnees incompltes.

## A Annexes

### A.1 Prompts LLM (extraits)

```
1 Prompt d'extraction (Gemini) :
2 You are a job-offer information extractor.
3
4 STRICT RULES:
5 - Output ONLY valid JSON. No markdown, no commentary, no surrounding
   text.
6 - Do NOT invent. If a value is not explicitly present, use null.
7 - Empty lists must be [].
8 - Extract ONE output object per offer in the input array, preserving the
   same order and length.
9
10 OUTPUT FORMAT:
11 Return a JSON array of objects. Each object must follow EXACTLY this
   schema:
12 {
13     "company": string|null,
14     "job_title": string|null,
15     "city": string|null,
16     "country": string|null,
17     "remote_policy": "onsite"|"hybrid"|"remote"|"unknown"|null,
18     "contract_type": "internship"|"apprenticeship"|"full_time"|
19         "part_time"|"unknown"|null,
20     "salary": string|null,
21     "currency": string|null,
22     "duration": string|null,
23     "application_deadline": string|null,
24     "skills": string[],
25     "languages": string[],
26     "source_url": string|null,
27     "source_title": string|null,
28     "city_score": number|null
29 }
30
31 MAPPING RULES:
32 - source_url = offer.url if present, else null
33 - source_title = offer.title if present, else null
34 - city_score = offer.score if present, else use input.search_score if
   provided, else null
```

```

35 - remote_policy:
36   - "remote" if clearly remote
37   - "hybrid" if clearly hybrid
38   - "onsite" if clearly onsite/on-site
39   - otherwise "unknown"
40 - contract_type:
41   - "internship" if stage/internship/intern
42   - "apprenticeship" if alternance/apprenticeship
43   - otherwise "unknown"
44 - salary: keep original text
45 - currency: "EUR", "USD", "GBP" if clearly identifiable, else null
46
47 Prompt de resume (Gemini) :
48 Generate a short readable summary for each extracted offer.
49
50 STRICT RULES:
51 - Output ONLY valid JSON. No markdown, no commentary.
52 - Return a JSON array with the same length/order as the input array.
53 - Do not include any other keys.
54
55 OUTPUT SCHEMA (per item):
56 {
57   "source_url": string|null,
58   "summary": string
59 }
60
61 SUMMARY RULES:
62 - 2 to 3 sentences maximum.
63 - Include: company (if known), role, location, remote policy, contract
   type, and salary if present.
64 - If a field is missing, omit it (do not guess).
65 - Write the summary in French.
66
67 Prompt de synthese finale (Gemini) :
68 Tu es un assistant charge de produire un rapport clair et synthetique
69 a partir d'une liste d'offres de stage.
70
71 Ta tache :
72 1) Un titre : "Synthese des offres de stage"
73 2) Le nombre total d'offres pertinentes trouvees
74 3) Le Top 3 des meilleures offres avec pour chacune :
75   - entreprise
76   - poste
77   - ville
78   - pourquoi elle est interessante (1 a 2 phrases max)
79 4) Des statistiques globales :
80   - Villes les plus frequentes
81   - Types de contrats dominants
82   - Remote vs onsite vs hybride
83   - Salaires observes (si presents)
84 5) Une conclusion courte (2-3 phrases) sur la qualite globale des offres

```

## A.2 Exemple de sortie JSON complète

Exemple synthetique (format complet) :

```

1 {
2   "company": "Example Corp",

```

```
3 "job_title": "DevOps Intern",
4 "city": "Berlin",
5 "country": "Germany",
6 "remote_policy": "hybrid",
7 "contract_type": "internship",
8 "salary": "1200 EUR/month",
9 "currency": "EUR",
10 "duration": "6 months",
11 "application_deadline": null,
12 "skills": ["linux", "docker", "kubernetes"],
13 "languages": ["english"],
14 "source_url": "https://example.com/job",
15 "source_title": "DevOps Internship - Example Corp",
16 "city_score": 92.68,
17 "summary": "Example Corp propose un stage DevOps à Berlin en mode
18     hybride..."}
```