

■ 求人原稿自動フィルタリングシステム — フィルタリング条件詳細ドキュメント

作成者：イ・ジェリョン

(filter_core_v2.py 基準 / Core Logic 説明書)

本ドキュメントは、filter_core_v2.py において実行されるすべての検証項目および判定基準を整理した開発者向け説明資料である。

各行に対して合計9つの検証項目をチェックし、個別結果を総合して最終判定 (OK / 要確認 / NG) を決定する。

1. 全体判定ロジック概要

各行ごとに以下の9つのチェック項目を個別に判定する。

- (1) 必須項目 (check_required)
- (2) メールアドレス (check_email)
- (3) 雇用形態 (check_employment)
- (4) 職種 (check_job_title)
- (5) 就業先会社名の特殊記号 (check_company_special)
- (6) 紹介元会社名の特殊記号 (check_intro_company_special)
- (7) 非公開ルール (check_private_intro)
- (8) 市区町村の文字化け (check_city_garbled)
- (9) 最低賃金 (judge_min_wage)

判定優先順位は以下の通りとする。

- (1) 1つでも NG が存在する場合 → 最終判定 NG
- (2) NG は存在しないが 要確認 が存在する場合 → 最終判定 要確認
- (3) すべての OK の場合 → 最終判定 OK

2. 必須項目存在チェック

チェック対象カラムは以下の通り。

- (1) 就業先会社名
- (2) 雇用形態
- (3) 職種
- (4) 応募先メールアドレス
- (5) 市区町村
- (6) 給与形態
- (7) 給与下限

判定基準は以下の通り。

- (1) すべての必須カラムに値が存在する場合 → OK
- (2) 1つでも空欄が存在する場合 → NG

3. 応募先メールアドレス形式チェック

チェック内容は以下の通り。

- (1) メールアドレスカラムの空欄チェック
- (2) 複数のメールを分離可能 (, ; の空白と改行はすべて許可)
- (3) 各メールアドレスが正規表現に一致するか確認

判定基準は以下の通り。

- (1) すべてのメール形式が正常な場合 → OK
- (2) メールアドレスが空欄の場合 → NG
- (3) 1つでも形式エラーが存在する場合 → NG

4. 雇用形態許可チェック

許可されている雇用形態は以下の通り。

"正社員", "契約社員", "派遣社員", "パート", "アルバイト", "アルバイト・パート", "アルバイト/パート", "アルバイト / パート", "アルバイト、パート", "パート・アルバイト", "パート/アルバイト", "パート / アルバイト", "パート、アルバイト", "業務委託"

判定基準は以下の通り。

- (1) 許可リストと完全一致する場合 → OK
- (2) 空欄の場合 → NG
- (3) 許可されていない表記の場合 → NG

5. 就業先（採用先）会社名の特殊記号チェック

禁止されている特殊記号は以下の通り。

"(株)", "(株)", "(株)", "(有)", "(有)", "(有)"

判定基準は以下の通り。

- (1) 特殊記号を含まない場合 → OK
- (2) 空欄の場合 → NG
- (3) 特殊記号を含む場合 → NG

6. 紹介元会社名の特殊記号チェック

判定基準は以下の通り。

- (1) 空欄の場合 → OK
- (2) 特殊記号を含まない場合 → OK
- (3) 特殊記号を含む場合 → NG

7. 就業先会社名が「非公開」の場合のルールチェック

基本ルールは以下の通り。

就業先会社名が「非公開」の場合、紹介元会社名は必須とする。

判定基準は以下の通り。

- (1) 就業先会社名が「非公開」でない場合 → OK
- (2) 就業先会社名が「非公開」かつ紹介元会社名が空欄の場合 → NG
- (3) 就業先会社名が「非公開」かつ雇用形態が「派遣社員」、紹介元会社名が「非公開」の場合 → NG
- (4) 就業先会社名が「非公開」かつ雇用形態が「派遣社員」、紹介元会社名が空欄でなく「非公開」でもない場合 → OK
- (5) 就業先会社名が「非公開」かつ雇用形態が「派遣社員」以外で、紹介元会社名が空欄でない場合 → OK (ただし、紹介元の会社名が非公開の場合はNGです)

8. 市区町村文字化けチェック

チェック対象

市区町村

文字化け判定基準は以下の通り。

- (1) 文字列に空白を含む場合 → NG
- (2) 制御文字 (\x00 ~ \x1F) を含む場合 → NG
- (3) 文字化けと判断される特定文字 ("闐|驥|伴") を含む場合 → NG

9. 職種適合性チェック

チェック内容は以下の通り。

- (1) 空欄チェック
- (2) 都道府県名の含有有無
- (3) 地名形式の含有有無
- (4) 募集条件・勤務条件の混入有無

- (5) 数字の含有有無

判定基準は以下の通り。

- (1) 空欄の場合 → NG
- (2) 地名または都道府県を含む場合 → NG
- (3) 募集・勤務・雇用条件キーワードを含む場合 → 要確認
- (4) 数字を含む場合 → 要確認
- (5) 上記以外 → OK

定義済み都道府県名一覧

“北海道, 青森, 岩手, 宮城, 秋田, 山形, 福島, 茨城, 栃木, 群馬, 埼玉, 千葉, 東京, 神奈川, 新潟, 富山, 石川, 福井, 山梨, 長野, 岐阜, 静岡, 愛知, 三重, 滋賀, 京都, 大阪, 兵庫, 奈良, 和歌山, 鳥取, 島根, 岡山, 広島, 山口, 徳島, 香川, 愛媛, 高知, 福岡, 佐賀, 長崎, 熊本, 大分, 宮崎, 鹿児島, 沖縄”

地名形式

区/市/町/村/駅

職種文字列が極端に短い場合（長さ3未満）は、単一漢字（区・市・町・村・駅）による誤検知を防ぐため、地名形式チェックを行わない。これは意味のない単一文字や破損データが地名含有職種として誤判定されるのを防ぐための安全対策である。

募集・勤務・雇用条件キーワード一覧

"募集", "急募", "大募集", "積極採用", "オープニング", "新規", "正社員", "契約社員", "派遣社員", "アルバイト", "パート", "業務委託", "夜勤", "日勤", "深夜", "早朝", "交替", "シフト", "残業", "未経験", "経験不問", "学歴不問", "資格不問", "〇〇不問", "歓迎", "優遇", "高収入", "日払い", "週払い", "即日", "短期", "長期", "寮", "社宅", "在宅", "リモート", "テレワーク", "マネージャー", "リーダー", "部長", "課長", "係長", "主任", "候補"

10. 最低賃金判定

使用カラム

- (1) 都道府県 (addressRegion) → col_pref
- (2) 勤務地住所 → col_address
- (3) 市区町村 (addressLocality) → col_city
- (4) 職種 → col_job
- (5) 就業先会社名 → col_work_company
- (6) 給与形態 (unitText) → col_wage_unit
- (7) 給与下限 (minValue) → col_wage_lower

10-(1) 都道府県決定ロジック

優先度1：都道府県カラムが MIN_WAGE に正確に存在する場合、その値を使用

優先度2：以下テキストから PREF_RE により都道府県を抽出

- 住所(addr)
- 市区町村(city)
- 職種(job)
- 会社名(comp)

10-(2) 給与単位 (unitText) から時給換算

現在unitTextマッピングは、次のようだ

UNIT_MAP = {1:"HOUR", 2:"DAY", 3:"MONTH", 4:"YEAR", 5:"WEEK"}

換算定数 (固定値)

- DAY : 1日 = 8時間
- MONTH : 1か月 = 160時間
- YEAR : 1年 = 1920時間

換算ルール

- HOUR : そのまま
- DAY : $\text{amount} / 8$
- MONTH : $\text{amount} / 160$
- YEAR : $\text{amount} / 1920$
- WEEK : 換算不可 → hourly = None
- その他 (UNKNOWN) : hourly = None

10-(3) 最低賃金判定分岐

- step1 : 都道府県取得失敗 → NG
- step2 : unitText 欠損またはパース失敗 → NG
- step3 : minValue 欠損またはパース失敗 → NG
- step4 : 週給の場合 → 要確認
- step5 : その他換算不可 → NG
- step6 : $\text{hourly} \geq \text{最低賃金}$ → OK
- step7 : $\text{hourly} < \text{最低賃金}$ → NG