Image not found or type unknown

# How the matching engine works

## PUQ Sanctions Checker module **WHMCS**

Order now | Download | Community

For every client the module screens two name candidates against the local lists: `firstname lastname` and `companyname` (plus the profile country against the blocked-countries list).

# Normalization

Both the client name and every list entry name/alias go through the same pipeline: lowercase → Cyrillic transliterated to Latin (`Иванов` → `ivanov`) → remaining accents transliterated to ASCII (`Müller` → `muller`) → everything except letters/digits/spaces removed → multiple spaces collapsed.

# The three methods

Each pair is scored by up to three configurable methods (Configuration → Sanctions sources → Matching engine); the best score wins and the method that produced it is shown in the match details.

## 1. Full name (`local_match_threshold`, default 85%)

PHP `similar_text` over the whole normalized strings: `2 × common_characters / (len1 + len2) × 100`. Repeated with alphabetically sorted words, higher of the two wins, so `Ivan Ivanov` = `Ivanov Ivan` = 100%. Example: `dmitry ivanov` vs `dmitrii ivanov` ≈ 89%.

# 2. Trigrams (default on / 65% / min length 6)

Both names are cut into 3-letter chunks sliding from start to end (`iv`, `iva`, `van`, `ano`, `nov`, …) and the share of common chunks is computed (Dice: `2 × common / (n1 + n2) × 100`). Very tolerant to typos and letter swaps: `Vladimir` vs `Vladimr` ≈ 81%. The method runs only when **both names have at least `trigram_min_length` letters** (spaces not counted) — short names like `Li Wei` produce so few chunks that unrelated names would cross the threshold by coincidence.

# 3. Words (default on / minimum 1 / word similarity 85%)

Every client word is matched to the closest entry word (words shorter than 3 letters and generic company suffixes like LLC/Ltd/GmbH are ignored); a pair counts when the two words are ≥ `word_similarity_threshold` similar (`Ivanov`/`Ivanoff` counts). The entry is reported when at least `word_match_min_words` pairs matched **and at least one matched word is the client SURNAME or a company word — a first-name-only match never flags the client**. The score is scaled by the fraction of client words that matched: a surname-only hit on "Ivan Ivanov" scores ~50%, a full-name hit ~100%.

With the default minimum of 1, clients whose surname (or one company word) matches a listed person are flagged **for review** with a proportionally lower score — the admin inspects the match (each one links to the OpenSanctions entity profile) and either acts or marks the client as Verified.

# Reference behaviour (default settings)

| Client vs list entry | Result |
|---|---|
| `Vladimr Putin` vs `Vladimir Putin` (typo) | full 96% |
| `Дмитрий Петров` vs `Dmitriy Petrov` (Cyrillic) | full 93% |

| Client vs list entry | Result |
| --- | --- |
| `Dmytro Kravchenko` vs `Denis Borisovich KRAVCHENKO` (surname only) | words 1/2, score 50% |
| `Denis Petrenko` vs `Denis Borisovich KRAVCHENKO` (first name only) | **no match** |
| `Roskosmos LLC` vs `State Corporation Roskosmos` (company word) | words 1/1, score 100% |
| `Alpha Trading LLC` vs `Bravo Trading Limited` (suffixes only) | **no match** |
| Unrelated names | no match |

# The US CSL API

When enabled, each name candidate is sent to the trade.gov API with `fuzzy_name=true`. The government's own fuzzy algorithm handles typos and transliteration variants and returns a match score; results below **CSL minimum match score** are dropped.

**Important:** a match is a signal for manual review, not proof. Fuzzy matching by name always produces some false positives — that is why the ticket is the primary action and the verification workflow exists.

---

Revision #7
Created 11 July 2026 21:33:28 by Ruslan
Updated 15 July 2026 23:51:25 by Ruslan