Image not found or type unknown

Automation and cron

How the automatic screening engine and scheduled tasks work.

- How the matching engine works
- Cron automation & check queue

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%
`Dmytro Kravchenko` vs `Denis Borisovich KRAVCHENKO` (surname only)	words 1/2, score 50%

Client vs list entry	Result
`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.

Cron automation & check queue

PUQ Sanctions Checker module **WHMCS**

Order now | Download | Community

The module uses the standard WHMCS cron. No extra cron entries are needed:

- `DailyCronJob` — list synchronization, mass-check scheduling, queue housekeeping;
- `AfterCronJob` (every cron tick, normally ~5 minutes) — **check queue processing**.

The check queue

Checks can be slow: the CSL API is an external HTTPS call per name, FraudRecord hashes every field 32 000 times. Running that inline during client registration or checkout can cause visible delays and timeouts. The queue solves this:

Setting	Default	Effect
Run trigger checks on the next cron (queue)	**off**	Registration / profile update / order checkout only **add the client to the queue** (instant) and the cron performs the check within ~5 minutes. Off = checks run inline as before.
Queue the mass check (process on cron)	**on**	The "Check all clients now" button and the scheduled mass check put clients into the queue; the cron drains it in portions. Off = the button runs realtime AJAX batches with a progress bar.
Queue items processed per cron run	50	Portion size per cron tick. Lower it when the CSL API is enabled.

Manual checks always run in real time — the Check Now button on the client card and Recheck on the Clients page never use the queue.

The queue lives in `puq_sanctions_checker_queue` (client, type sanctions/fraud, context, status pending → processing → done/error). The Home page shows the number of pending items; every processed portion is logged (`queue_processed`), failures per client are logged as `check_error`. Finished rows are purged after 7 days. If the queue table is unavailable, trigger checks automatically fall back to real-time so registration never breaks.

What runs daily

1. **Sanctions lists synchronization** — the enabled OpenSanctions mirrors (EU FSF, Canada DFATD-SEMA) are downloaded and re-imported into the local database. The files are ~2 MB each and the URLs are static, so this is fast and reliable.
2. **Periodic mass check** — when **Periodic mass check (cron)** is set to Daily, Weekly, or Monthly, the module re-screens clients (Active only or all, per the **Clients to check in mass check** setting) once per period — via the queue by default. The last run timestamp is shown on the Home page.

Notes

- Tickets are deduplicated, so a weekly mass check does not spam the helpdesk: a client with an already-open sanctions ticket does not get a second one.
- All cron activity is written to the module **Log** page (`list_sync`, `mass_check`, `queue_processed` events).

Manual equivalents

Everything the cron does can be triggered from the **Home** page: **Sync sanctions lists now** and **Check all clients now** (queued by default; with the queue disabled — realtime batches with a progress bar).