

Look-Alike Attorneys

Methodology & Statistical Approach

How IDiyas finds patent practitioners with similar technical practice areas

What it does

Two patent attorneys may sit at different firms in different cities, yet quietly prosecute the same kinds of inventions — the same semiconductor filings, the same biotech claims, the same software architectures. The Look-Alike Attorneys feature surfaces those peers: for any practitioner, it returns a ranked list of other attorneys whose body of prosecuted work points in the same technical direction.

In one sentence

We describe every attorney as a vector of patent counts across technology categories (CPC codes) drawn from the patents they have prosecuted, then rank all other attorneys by how closely their vector correlates with that profile.

Step 1 — Build each attorney's practice fingerprint

Every patent carries one or more CPC codes (Cooperative Patent Classification — the shared USPTO/EPO system) describing its precise technical subject matter. By collecting all the patents on which an attorney or firm is the practitioner of record and tallying their CPC codes, we build a profile: a dictionary mapping each technology category to how often that attorney has worked in it.

```
Attorney profile (illustrative):
{
  "H01L21-02": 187,    // semiconductor processing
  "H01L29-78": 142,    // transistor structures
  "G06F30-39": 61,     // circuit design / EDA
  "H03K19-00": 44,     // logic circuits
  ... more CPC codes ...
}
```

This dictionary is the attorney's vector. Each CPC code is one dimension; the count is how heavily the attorney's docket leans into that area. A boutique specialist has a sparse vector concentrated in a few codes; a full-service practice has a long, spread-out one. Using counts rather than a present/absent flag preserves **intensity** — a practitioner with 187 semiconductor filings is treated very differently from one with a single incidental matter.

Step 2 — Measure similarity between attorneys

Two attorneys are alike if their profiles rise and fall together: where one has many filings, the other tends to as well, and where one is thin, the other is thin too. This co-movement is exactly what a Pearson-style correlation coefficient captures. For two profiles aligned over the union of their CPC codes (any code missing from one attorney is filled with zero):

$$r = \frac{n \cdot \Sigma(xy) - \Sigma x \cdot \Sigma y}{\sqrt{[n \cdot \Sigma x^2 - (\Sigma x)^2] \cdot [n \cdot \Sigma y^2 - (\Sigma y)^2]}}$$

x = count of a CPC code for attorney A (0 if absent)

y = count of the same CPC code for attorney B

n = number of distinct CPC codes across both

The score r runs from −1 to +1:

- **r near +1:** dockets move together strongly — a strong practice look-alike.
- **r near 0:** no meaningful overlap — unrelated practice areas.
- **r below 0:** opposite focus — where one specializes, the other is absent.

Step 3 — Rank and publish

- Restrict to attorneys with enough prosecuted patents to form a stable profile — a minimum activity threshold filters out noise, since a handful of filings can make two attorneys look identical by chance.
- Compare each attorney's profile against every other using the correlation above.
- Keep only positive correlations, sort them descending, and retain the strongest matches.
- Store the ranked peers on the attorney's record so the profile page loads them instantly — no live computation at request time.

The core insight

Look-alike status is about what an attorney prosecutes, not which firm they belong to or where they practice. That is why the feature can connect a large-firm partner with a solo practitioner across the country who has never appeared on the same matter — because their technical fingerprints align.

Notes & limitations

- **Pearson-family profile measure.** The correlation is computed over the set of CPC codes shared and unshared between two attorneys, so it behaves as a centered, count-weighted similarity score rather than a textbook correlation over independent paired samples — but it ranks look-alikes effectively.
- **Volume vs. specialization.** Because profiles use raw counts, high-volume practices carry large magnitudes; future refinements (share-based or TF-IDF-style weighting that down-weights ubiquitous CPC codes) could further isolate specialization from sheer docket size.
- **Attribution scope.** Profiles reflect the patents attributed to an attorney or firm of record in the corpus; coverage and naming consistency in the underlying data shape the result.
- **Positive matches only.** Negatively correlated attorneys are explicitly dissimilar and are not surfaced as suggestions.

Look-Alike Attorneys turns the full IDiyas patent corpus into a per-attorney technical fingerprint (CPC profile), then ranks every attorney against every other by Pearson-style correlation of those fingerprints — filtered to active practitioners and to positive matches. The result is a fast, precomputed list of practice peers powering both attorney profile pages and partner data feeds.

— IDiyas Patent Intelligence · Internal Technical Documentation —