

Look-Alike Inventors

Methodology & Statistical Approach

How IDiyas finds inventors with similar research and technical interests

What it does

Two inventors can work at different companies, on different continents, and never have co-authored a single patent — yet still work on remarkably similar problems. The Look-Alike Inventors feature surfaces those hidden peers: for any inventor, it returns a ranked list of other inventors whose body of work points in the same technical direction.

In one sentence

We describe every inventor as a vector of patent counts across technology categories (CPC codes), then rank all other inventors by how closely their vector correlates with that profile.

Step 1 — Build each inventor's research 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 of an inventor's patents and tallying their CPC codes, we build a profile: a dictionary mapping each technology category to the number of times that inventor has worked in it.

```
Inventor profile (illustrative):
{
  "G06N3-08": 34,    // neural network training
  "G06N3-04": 28,    // neural network architectures
  "G06F18-21": 19,   // pattern-recognition design
  "H04L9-08": 6,     // key distribution
  ... more CPC codes ...
}
```

This dictionary is the inventor's vector. Each CPC code is one dimension; the count is how strongly the inventor leans into that area. A generalist has a long, spread-out vector; a deep specialist has a sparse one concentrated in a few codes. Using counts rather than a simple present/absent flag preserves **intensity** — an inventor with 34 neural-network patents is treated very differently from one with 3.

Step 2 — Measure similarity between inventors

Two inventors are alike if their profiles rise and fall together: where one has many patents, 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 inventor 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 inventor A (0 if absent)

y = count of the same CPC code for inventor B

n = number of distinct CPC codes across both

The score r runs from -1 to +1:

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

Step 3 — Rank and publish

- Restrict to inventors with enough patents to form a stable profile — a minimum activity threshold filters out noise, since a handful of patents can make two inventors look identical by chance.
- Compare each inventor'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 inventor'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 inventor works on, not where they work or who they know. That is why the feature can connect a corporate R&D engineer with an academic researcher across the world who has never appeared in the same patent — 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 inventors, 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, highly prolific inventors carry large magnitudes; future refinements (share-based or TF-IDF-style weighting that down-weights ubiquitous CPC codes) could further isolate specialization from sheer output.
- **Positive matches only.** Negatively correlated inventors are explicitly dissimilar and are not surfaced as suggestions.

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

— IDiyas Patent Intelligence · Internal Technical Documentation —