# GLOBAL PATENT INTELLIGENCE ENGINE

A Framework for National and Global Patent Search, Classification, Family Mapping, and Claim-Overlap Analysis


| | |
|---|---|
| **AUTHOR:** | Hon. Tyree J. Mason I, Director of Bureau of Computum Analysis |
| **FRAMEWORK TYPE:** | Patent Intelligence & Prior-Art Retrieval Algorithm |
| **SCOPE & FUNCTION:** | National, regional, and international systems \| Discovery, normalization, classification, comparison, and analytical ranking |


## 1. SYSTEM DEFINITION

The Global Patent Intelligence Engine (GPIE) is a computational framework designed to conduct systematic patent searches across national, regional, and international intellectual-property databases. Rather than treating a patent search as a conventional text query, GPIE models the search as a graph-based information-retrieval problem:

$$P_{search} = \Phi(K, C, F, J, T, Q, R)$$

where:

- $K$ = technical keywords and terminology
- $C$ = patent classification structures
- $F$ = patent-family relationships
- $J$ = jurisdictional information
- $T$ = temporal relationships
- $Q$ = claims and technical limitations
- $R$ = relevance and similarity metrics

**Core Objective:** The objective is not merely to find patents containing similar words, but to determine: Which previously disclosed inventions, patent applications, patent families, claims, publications, and technical disclosures are materially relevant to the subject invention, in which jurisdictions, during which periods, and with what degree of technical and legal overlap?

## 2. CORE ARCHITECTURE

GPIE consists of nine principal layers:

$$GPIE = L_1 + L_2 + L_3 + L_4 + L_5 + L_6 + L_7 + L_8 + L_9$$

| Layer | Name | Primary Operation |
|---|---|---|
| Layer 1 | Invention Decomposition | Convert the proposed invention into discrete technical elements. |
| Layer 2 | Terminology Expansion | Generate synonyms, historical terminology, abbreviations, scientific terminology, and alternative descriptions. |
| Layer 3 | Classification Discovery | Identify relevant CPC, IPC, USPC, Locarno, and other applicable classification systems. |
| Layer 4 | Global Retrieval | Search national, regional, and international patent repositories. |
| Layer 5 | Patent-Family Reconstruction | Determine whether apparently separate documents represent the same underlying invention. |
| Layer 6 | Claim Intelligence | Extract independent and dependent claims and decompose them into technical limitations. |
| Layer 7 | Prior-Art Temporal Analysis | Determine publication dates, priority dates, filing dates, continuations, divisionals, grants, and expiration/status information. |
| Layer 8 | Similarity & Overlap Analysis | Compare the proposed invention against identified disclosures. |
| Layer 9 | Legal-Analytical Output | Produce a jurisdiction-aware patent landscape and identify potential novelty, overlap, freedom-to-operate, and further-search issues. |

## 3. INVENTION DECOMPOSITION ENGINE

The first operation should never be a keyword search. The invention should first be represented as a structured vector:

$$I = \{S, F, M, E, C, P, O\}$$

where: $S$ = system components, $F$ = functional relationships, $M$ = mechanisms, $E$ = inputs/outputs, $C$ = control architecture, $P$ = process steps, and $O$ = intended operational result.

*Example:* An AI-controlled energy system might be decomposed into: energy-input subsystem, sensing subsystem, computational subsystem, control algorithm, conversion mechanism, energy-storage subsystem, feedback mechanism, and output subsystem. This decomposition prevents the search engine from becoming dependent upon the inventor's particular terminology.

## 4. TECHNICAL ONTOLOGY GENERATOR

GPIE creates a terminology graph $G_T = (V, E)$ where each node represents a technical concept and each edge represents a semantic relationship. For every technical element, the engine generates: exact terminology, synonyms, abbreviations, acronyms, historical terminology, engineering terminology, scientific terminology, alternative spellings, component-level terminology, functional descriptions, result-oriented descriptions, broader concepts, and narrower concepts.

*Example:* "Quantum sensor" may expand into concepts involving: quantum sensing, quantum measurement, quantum detector, quantum-enhanced sensing, quantum metrology, atomic sensor, superconducting sensor, spin-based sensing, and photon-based detection. The expansion engine should distinguish between true semantic equivalents and merely related concepts.

## 5. SEARCH VECTOR GENERATION & 6. SEARCH HIERARCHY

The system generates several independent search vectors rather than one massive Boolean query:

$$V = [V_K, V_C, V_F, V_Q, V_T, V_H]$$

(Keyword, Classification, Functional, Claim-language, Temporal, Hybrid semantic)

GPIE executes searches in 9 progressively broader layers:

1. **Exact Terminology:** Search exact technical phrases.
2. **Synonym Expansion:** Replace terms with recognized equivalents.
3. **Functional Search:** Search for what the invention does, rather than what it is called.
4. **Component Search:** Search individual subsystems independently.
5. **Classification Search:** Search CPC/IPC and other classification structures.
6. **Citation Search:** Follow backward, forward, examiner, applicant, and NPL citations.
7. **Family Search:** Expand discovered documents into their complete patent families.
8. **Inventor/Assignee Search:** Search related work by inventor, applicant, assignee, subsidiary, predecessor, or successor.

9. **Semantic Similarity:** Use embeddings/semantic-retrieval methods to locate technically similar disclosures with different vocabulary.

## 7. GLOBAL JURISDICTION MATRIX

Every retrieved document is assigned to a jurisdictional node: $J_i = (country, region, office, status)$. The system distinguishes among:

- **National Offices:** United States, Canada, United Kingdom, Germany, France, Italy, Spain, Japan, China, South Korea, India, Australia, Brazil, Mexico, Israel, Singapore, and other national offices.
- **Regional Systems:** Applicable regional patent organizations (e.g., EPO, EA, ARIPO, OAPI).
- **International System:** The PCT/WIPO publication layer is treated as a global routing layer rather than automatically interpreted as a worldwide granted patent.

## 8. PATENT-FAMILY RECONSTRUCTION

Determining whether multiple documents represent the same underlying invention via family graph $G_F = (P, E)$ (where $P$ = patent documents, $E$ = priority/continuation/divisional/national-phase relationships). The system identifies: earliest priority document, priority chain, PCT application, national-phase applications, continuations, CIPs, divisionals, grants, abandoned applications, expired patents, and pending applications. This prevents a single invention from being incorrectly counted as dozens of independent prior-art references.

## 9. CLAIM DECOMPOSITION & 10. CLAIM OVERLAP MATRIX

$$C_i = \{L_1, L_2, ..., L_n\}$$

Patent claims are converted into structured limitations. For an independent claim: Each limitation is classified as structural, functional, relational, temporal, numerical, material, algorithmic, procedural, environmental, or control-related.

*Example:* "System comprising A configured to communicate with B to generate C" becomes $C = \{A, B, A \leftrightarrow B, f(A,B)=C\}$. The engine compares limitations rather than entire sentences.

For candidate patent $P$ and proposed invention $I$, the claim overlap is calculated as:

$$O(I, P) = \sum_{i=1}^{n} (w_i M_i) / \sum_{i=1}^{n} w_i$$

where $M_i \in [0, 1]$, $M_i = 1$ when limitation $i$ is materially disclosed ($M_i = 0$ if absent), and $w_i$ is the importance weight.

## 11. RELEVANCE SCORING

Each retrieved document receives a composite relevance score:

$$R(P) = w_k K + w_c C + w_s S + w_q Q + w_f F + w_t T + w_x X$$

where weights are adjustable by search mode:

- **Prior-Art Mode:** Increase $w_q$, $w_f$
- **Patent-Landscape Mode:** Increase $w_c$, $w_s$, $w_f$
- **Competitive-Intelligence Mode:** Increase $w_f$, $w_x$, $w_j$

## 12. TEMPORAL, 13. CITATION-GRAPH, & 14. SEMANTIC INTELLIGENCE

**Temporal Intelligence:** Maintains four distinct dates $D = \{D_p, D_f, D_{pub}, D_g\}$ (priority, filing, publication, grant). The engine never treats these dates as interchangeable.

**Citation-Graph Expansion:** Recursively explores until the marginal discovery rate falls below threshold $\varepsilon$:

$$P_0 \rightarrow C_{backward} \rightarrow C_{forward} \rightarrow F_{family} \rightarrow I_{inventor} \rightarrow A_{assignee}$$

$$(\Delta N_{relevant} / \Delta N_{searched}) < \varepsilon$$

**Semantic Search Layer:** Supplemented by vector retrieval. Embeddings $e_P, e_I \in \mathbb{R}^d$ are compared via cosine similarity: $S(P, I) = cos(e_P, e_I)$. Used for retrieval only; legal conclusions require claim-level and jurisdiction-specific analysis.

## 15. NPL, 16. STATUS, & 17. SEARCH CONFIDENCE

**Non-Patent Literature (NPL):** Includes scientific papers, conference proceedings, dissertations, technical manuals, standards, archived websites, engineering publications, regulatory documents, and public demonstrations. Classified as Source $\in$ {Patent, Application, NPL, Product, Standard, Archive, Other}.

**Jurisdictional Status Engine:** Normalized state Status($P$) ∈ { Pending, Granted, Abandoned, Expired, Lapsed, Revoked, Withdrawn, Unknown } while preserving original office status.

**Search Confidence Score:** Conf = $f$(source quality, claim coverage, family certainty, temporal certainty, jurisdictional certainty).

| Score Range | Interpretation |
|---|---|
| 0.90 - 1.00 | Very high confidence |
| 0.75 - 0.89 | High confidence |
| 0.50 - 0.74 | Moderate confidence |
| 0.25 - 0.49 | Low confidence |
| 0.00 - 0.24 | Insufficient evidence |

## 18. SEARCH-STATE MACHINE

The algorithm operates as a state sequence $S_0 \rightarrow S_1 \rightarrow S_2 \rightarrow S_3 \rightarrow S_4 \rightarrow S_5 \rightarrow S_6 \rightarrow S_7$ with feedback loop $S_5 \rightarrow S_2$ to incorporate newly discovered terminology into earlier search stages.

## 19. CORE ALGORITHM

**GPIE EXECUTION ALGORITHM**

*INPUT:* Invention Description | Relevant Date | Target Jurisdictions | Search Objective

1. Parse invention.
2. Decompose invention into technical elements.
3. Generate technical ontology.
4. Generate synonyms and functional descriptions.
5. Identify CPC/IPC and applicable classification codes.
6. Execute exact-keyword searches.
7. Execute expanded keyword searches.
8. Execute classification searches.
9. Execute claim-focused searches.
10. Rank initial results.
11. Expand high-relevance patent families.
12. Traverse backward citations.

13. Traverse forward citations.
14. Search inventors and assignees.
15. Search related non-patent literature.
16. Extract claims and technical limitations.
17. Normalize dates.
18. Normalize jurisdictional status.
19. Construct patent-family graph.
20. Construct citation graph.
21. Calculate semantic similarity.
22. Calculate technical claim overlap.
23. Apply temporal filters.
24. Deduplicate family members.
25. Rank results.
26. Identify search gaps.
27. Repeat searches for unresolved gaps.
28. Stop when marginal discovery falls below threshold.
29. Generate patent landscape.
30. Generate prior-art matrix.
31. Generate claim-overlap matrix.
32. Generate jurisdictional matrix.
33. Generate confidence scores.
34. Produce final analytical report.

*OUTPUT:* Global Patent Intelligence Report

## 20. PRINCIPAL OUTPUT MATRICES

### A. Patent Landscape Matrix

| Patent Family | Jurisdiction | Priority | Publication | Status | Technical Relevance | Claim Relevance |
|---|---|---|---|---|---|---|
| FAM-001 | US, EP, JP | 2018-04-12 | 2019-10-18 | Granted | High | High |
| FAM-002 | WO, CN | 2020-01-15 | 2021-07-22 | Pending | Moderate | Low |

### B. Prior-Art Matrix

| Reference | Disclosure | Relevant Date | Technical Overlap | Claim Overlap | Confidence |
|---|---|---|---|---|---|

| | | | | | |
|---|---|---|---|---|---|
| US10123456B2 | Subsystem A + B control | 2017-09-01 | 85% | 70% | 0.92 |
| NPL-IEEE-2019 | Sensor feedback algorithm | 2019-02-10 | 60% | N/A | 0.85 |

### C. Family Matrix

| Root Priority | PCT | US | EP | CN | JP | KR |
|---|---|---|---|---|---|---|
| US201862123456 | WO2019123456 | US10987654 | EP3654321 | CN110987654 | JP6876543 | KR10202012345 |

### D. Claim-Limitation Matrix

| Limitation | Proposed Invention | Patent A | Patent B | Patent C |
|---|---|---|---|---|
| L1: Structural component A | ✓ | | | ✓ |
| L2: Functional coupling A ↔ B | ✓ | | ✓ | |
| L3: Control threshold algorithm | ✓ | ✓ | | ✓ |

## 21. SEARCH OBJECTIVE MODES

- **MODE 01 - NOVELTY SEARCH:** Primary question: Has the claimed subject matter already been disclosed?
- **MODE 02 - INVENTIVE-STEP / OBVIOUSNESS SEARCH:** Primary question: Do combinations of existing disclosures materially converge on the proposed invention?
- **MODE 03 - FREEDOM-TO-OPERATE SCREEN:** Primary question: Which potentially enforceable rights could materially intersect with the contemplated activity?
- **MODE 04 - PATENTABILITY LANDSCAPE:** Primary question: What technical territory remains comparatively open?
- **MODE 05 - COMPETITOR INTELLIGENCE:** Primary question: What organizations and inventors are developing technology in this domain?
- **MODE 06 - TECHNOLOGY LANDSCAPE:** Primary question: How is an entire technological field distributed across patents, families, jurisdictions, and organizations?
- **MODE 07 - PORTFOLIO ANALYSIS:** Primary question: How are individual patent families connected within an organization's intellectual-property portfolio?

## 22. COMPLETENESS METRIC & 23. ANTI-BIAS ARCHITECTURE

**Completeness Metric:** A search is not labeled "complete" merely because a set number of databases were queried. Completeness is calculated as $\Gamma = 1 - (N_{new} / N_{examined})$. As $\Gamma \rightarrow 1$, the search approaches saturation. System coverage is tracked via Coverage = (J, C, T, S, F, N).

**Anti-Bias Architecture:** Enforces independent search paths, blind expansion, family deduplication, citation expansion, and negative search recording (auditable search history).

## 24. AUDIT TRAIL & 25. FINAL INTELLIGENCE ARCHITECTURE

**Audit Trail:** Generates records $A_i = \{ timestamp, query, database, jurisdiction, filters, results, ranking, analyst/system, decision \}$.

**Final Intelligence Architecture:**

$$GPIE = D + O + C + R + F + Q + T + J + N + A$$

[ Invention Ontology ] → [ Classification ] → [ Global Retrieval ] → [ Patent Graph ] → [ Claim Graph ] → [ Temporal Analysis ] → [ Jurisdictional Analysis ] → [ Patent Intelligence ]

## 26. GOVERNING PRINCIPLE

**Governing Principle:** Search the invention by what it is called, what it does, what it contains, how its components interact, how its claims are constructed, where its technological ancestors appear, and how those disclosures propagate through global patent families.