

# I2E Patent Mining Solution

---

## Overview

Linguamatics I2E Patent Mining Solution brings the power of I2E text mining to three major full-text patent collections (WIPO, EPO and USPTO). The solution allows users to generate arbitrary queries for patent search and analysis, on a secure cloud-based platform via I2E OnDemand. A library of queries is available for common life sciences specific analysis. Powerful proprietary indexes identify chemical and biological entities wherever they appear in the patent, coupled with natural language processing (NLP) and recognition of numerical information means more precise searches, and better clustering of results.

I2E Patent Mining Solution includes the detailed (and hierarchical) Cooperative Patent Classification (CPC) which makes it possible to carry out highly specific searches. Linguamatics' normalized company name ontology and organization type identifier makes searching for competitors and potential partners reliable and efficient.

## Who should use Linguamatics I2E Patent Mining Solution?

### R&D

Before starting a new R&D project, using I2E will enable you to configure technically robust patent landscapes, white space analyses and state of the art reviews for confident decision making. Patents can be grouped into families for viewing using priority information, or viewed separately.

### Competitive Intelligence

In competitive intelligence off-the-shelf queries can be utilised to track new inventions in a particular

therapeutic area. Queries can be saved and automated to run regularly, always providing the most up-to-date information first to identify trends over time.

### Legal IP

Those working in legal IP departments will appreciate the ability to perform highly subject specific prior art searches and technically detailed invalidity searches at the document level.

First stage freedom to operate searching is made more efficient, by allowing searches to be restricted to individual claims, and/or by only querying granted patents from the start.

## Key Benefits

- ◆ Configure technically robust patent landscapes, white space analyses and state-of-the-art reviews for confident decision making
- ◆ More efficient first stage freedom to operate searching
- ◆ Track new inventions in a particular therapeutic area, using off-the-shelf queries
- ◆ Find new chemical entities in patents using structure, substructure & similarity search
- ◆ Find numerical information e.g. amounts, units, concentrations and percentages
- ◆ Find companies, drugs, diseases, genes located in any region of a WO, EP or US patent

## What's covered?

Fully searchable text from 3 authorities:

**EP** Applications (A) and Grants (B) documents. 1978 to date - 5 million documents, approx. 300,000 added per year.

Language of collection: English, French, German

**US** A (Applications) and A and B (Issued) documents 1976 to date - 12.5 million documents, approx 900,000 added per year.

Language of collection: English

**WO** A (Applications)

1978 to date - 3 million documents, approx 210,000 added per year.

Language of collection: English, French, German, Japanese, Russian, Spanish

Documents from additional patent authorities will be made available in future releases.

## Update frequency

The latest year's patent collection is updated on a weekly basis. Historical years are updated monthly, with complete indexing of each patent publication.

The patent collection is provided by IFI Claims Direct from Fairview Research and is licensed to Linguamatics for commercial use.

## Indexes and terminologies

Each patent publication is indexed individually with broad terminologies, NCI Thesaurus and NLM's MeSH (Medical Subject Headings), plus additional specialized terminologies for chemicals, diseases and genes and proteins. I2E also provides chemical structure recognition software and uses SMILES strings to enable structure search and visualization.

The terminologies contain hundreds of thousands of concepts and the different ways they are typically expressed (synonyms). In addition to concept searching, this allows searches for any members of a class e.g. any patent that mentions a disease (the patent may mention only a specific disease such as "diabetes"), providing alternative filters to the patent classification codes.

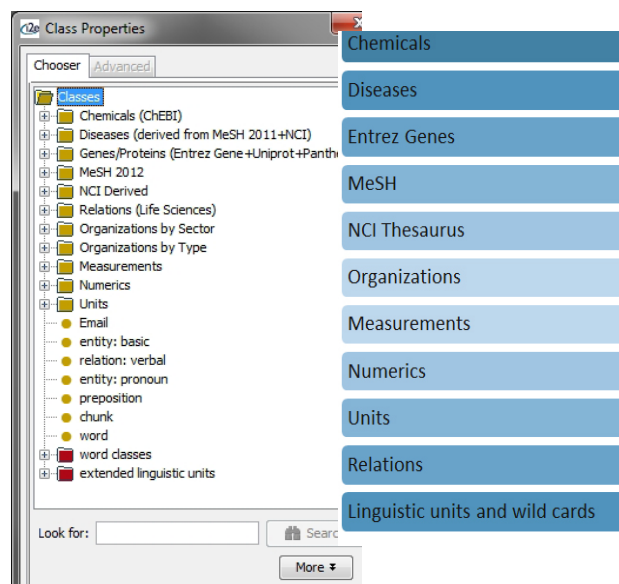


Figure 1. Examples of available terminologies

## Document structure

Each patent document is split into over 100 'regions' to provide consistent, highly detailed search and display options. Dates, tables, cited patents and non-patent literature from the search report and classifications are all separated for accuracy of retrieval.

The CPC is hierarchical, whereas IPC, US class and ECLA can all be searched at the single level.

Colorful highlighting of the patent document allows for easy visualization of important terms and concepts.

I2E arranges and structures the underlying patent and provides both a link back to original patent source as well as a full text copy that can be tailored to bring the most important information quickly to the attention of the viewer. See Figure 3 below.

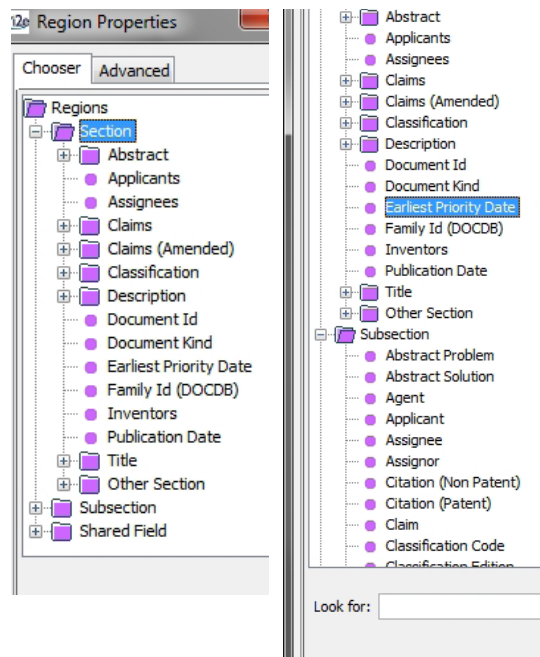


Figure 2. Granular searching over 100 regions

### Combination of a **GPR119** Agonist and the **DPP-IV** Inhibitor **Linagliptin** for Use in the Treatment of **Diabetes** and Related Conditions

#### Abstract (EN)

The present invention relates to combinations of DPP-4 inhibitors with GPR119 agonists, as well as to the use of these combinations for treating and/or preventing metabolic diseases, particularly diabetes (especially type 2 diabetes mellitus) and conditions related thereto.

<b>Inventors:</b>	EICKELMANN PETER (Country: DE ); LUIPPOLD GERD (Country: DE ); MARK MICHAEL (Country: DE ); THOMAS LEO (Country: DE ) (Format: EPO) EICKELMANN PETER, LUIPPOLD GERD, MARK MICHAEL, THOMAS LEO (Format: Intermediate) Eickelmann, Peter (Mittelbiberach Country: DE ); Luippold, Gerd (Warthausen-Birkenhard Country: DE ); Mark, Michael (Biberach an der Riss Country: DE ); Thomas, Leo (Biberach an der Riss Country: DE ) (Format: Original)
<b>Assignees:</b>	BOEHRINGER INGELHEIM INTERNATIONAL GMBH (Ingelheim am Rhein Country: DE ) (Format: Original)
<b>Applicants:</b>	BOEHRINGER INGELHEIM INT (Country: DE ); EICKELMANN PETER (Country: DE ); LUIPPOLD GERD (Country: DE ); MARK MICHAEL (Country: DE ); THOMAS LEO (Country: DE ) (Format: EPO) BOEHRINGER INGELHEIM INTERNATIONAL GMBH; EICKELMANN PETER, LUIPPOLD GERD, MARK MICHAEL, THOMAS LEO (Format: Intermediate) Eickelmann, Peter (Mittelbiberach Country: DE ); Luippold, Gerd (Warthausen-Birkenhard Country: DE ); Mark, Michael (Biberach an der Riss Country: DE ); Thomas, Leo (Biberach an der Riss Country: DE ) (Format: Original)
<b>Publication Id:</b>	US-20130109703-A1 (Publication Patent Office: US Publication Number: 20130109703 Publication Kind: A1 Publication Language: EN)
<b>Publication Date:</b>	20130502 (Publication Year: 2013 Publication Month: 05 Publication Day: 02)
<b>Application Id:</b>	US-201113634886-A (Application Patent Office: US Application Number: 201113634886 Application Kind: A Application Language: EN)
<b>Application Date:</b>	20110318 (Application Year: 2011 Application Month: 03 Application Day: 18)
<b>Family Id (DOCDB):</b>	44210260
<b>Earliest Priority Date:</b>	20100318 (Earliest Priority Year: 2010 Earliest Priority Month: 03 Earliest Priority Day: 18)
<b>International Convention Data</b>	
PCT or Regional Filing Data:	
<ul style="list-style-type: none"> <li>Country: EP Document Number: 1154169 Date: 20110318</li> </ul>	
<b>Reassignments</b>	
<b>Assignee(s):</b>	BOEHRINGER INGELHEIM INTERNATIONAL GMBH (BINGER STRASSE 173 INGELHEIM AM RHEIN 55216 Country: DE )
<b>Assignor(s):</b>	MARK, MICHAEL; THOMAS, LEO; EICKELMANN, PETER; LUIPPOLD, GERD
<b>Correspondence Address:</b>	BOEHRINGER INGELHEIM CORPORATION 900 RIDGEBURY ROAD P O BOX 368 RIDGEBURY CT 06872

Figure 3. Example Patent showing indexed terms linked to ontologies for intelligent retrieval

## Pre-formed search templates

Four search templates are available for you to get started. Options are:

1. Diseases landscaping;
2. Claims analysis - based on disease and/or pharmacological substance occurrence(s) in specific claims;
3. Patent Classification based analysis of a company's patent portfolio;

Specific keyword search (automatically matched to synonyms through the indexes) this can be restricted to a region of the patent e.g. claim(s), description, abstract, title.

## Display and download options

One of the strengths of I2E is the ability to configure exactly how you want the results to look. Export results in a range of output formats, including Excel for efficient post-processing.

Results can be arbitrarily configured to suit a particular task e.g.

- ii) Brief citation (publication number, title, applicant, first inventor, publication year)
- ii) Segment (publication number, relevant text from a region)
- iii) List of relevant normalized entities e.g. gene identifiers or chemical structures

## Summary

Linguamatics I2E Patent Mining Solution with NLP and term extraction makes it possible to find:

**New Chemical Entities** in patents using - structure, substructure and similarity search - to find novel and known compounds (trade names, IUPAC names, familiar names, laboratory code names) are all recognised and matched to structures where possible)

**Numerical information** - including amounts, units, concentrations, percentages and over 7 different measurements that encompass over 50 unit types that could be mentioned anywhere in the description or claims

**Companies, drugs, diseases, genes** located in any region of a WO, EP, or US patent

For more information  
or a demonstration,  
please call us on  
**+44 1223 651910**  
enquiries@linguamatics.com