

From loading dock to desktop: the journey of a patent to a *Derwent World Patents Index* record

Bob Stembridge
Thomson Scientific
July 2007

Derwent World Patents Index[®] (*DWPI*SM) is the world's premier patents research resource for technology alerting, analysis, competitive intelligence, prior art and patentability research, and infringement and invalidity studies. But what actually goes into preparing a record for *DWPI*? This article goes behind the abstract to examine the work carried out by Thomson Scientific analysts.

Introduction

For over fifty years, Thomson Scientific has been helping researchers keep up-to-date with technological developments by providing concise summaries of inventions reported each week in patent documents published around the world. These summaries are collected together in a database which now comprises over 15 million inventions covering all technology reported in patents from over 40 different authorities worldwide. This together with the classification and indexing applied by Thomson Scientific scientists and engineers has developed into *Derwent World Patents Index*.

This article examines the work carried out by Thomson Scientific analysts in preparing patent information for inclusion in the *DWPI* database. It is based on the presentation delivered at the 2006 PIUG meeting in Minneapolis by Andrew McFarlane, Content and Quality Manager for Chemistry, Thomson Scientific.

In the beginning . . .

. . . there is the patent publication!

The image shows a collage of patent documents. The top left is a Japanese patent document (JP 2006-090600). The top center is an European Patent Office document (EP 1 645 618 A1) for 'EUROPÄISCHE PATENTANMELDUNG'. The top right is a partial view of a document with 'ation = GB 2 342 597 ... A'. The bottom left is a United States Patent Application Publication (US 2006/0074267 A1) for 'Cas et al.' regarding 'ALKYLIMINOACETATE MELAYLAMINYL ESTERS'. The bottom right shows a graph labeled 'Fig. 2' with a vertical axis 'A' and a horizontal axis 'B', and a reference to 'GB 2 342 597 A'.

Around 38,000 patent publications arrive at Thomson Scientific each week. Printed on paper, that would equate to a stack about 75 metres high! Fortunately, much of this information is now supplied electronically, although still in a bewildering array of data formats and delivery media.

Preliminary processing

In order to manage this volume of patent information, a team of data analysts performs a series of processes to convert the data to a uniform format, identify and correct systematic data errors, apply name standardization and identify new inventions or assign existing inventions to existing patent families.

Error checks for systematic data errors are conducted on priority dates & numbers, application dates & numbers, and invalid or missing International Classification Symbols where correction is made by intellectually assigning the IPC at the Subclass level.

Accuracy is enhanced through identifying and correcting errors in company and inventor names which can occur through transliteration of non-roman character names, or through misspelling or incorrect formatting. Company names are checked against an internal register, and the Derwent company code added for recognised companies; for new company names, a new code is assigned and added to the internal register.

On receipt of a patent document, a Basic/Equivalence search is performed against an internal register of patent priority data. If the priority data searched from the newly arrived document do not match the priority data of any previously processed document, then the document describes a new invention, is designated "Basic", and a new *DWPI* record is created with a new unique accession number. If the document matches the priority data of an invention already listed in the internal register, this is designated "Equivalent" and is added to the existing *DWPI* record. That record now has at least two specifications as part of its patent family.

The *DWPI* patent family is a collection of documents related to the same invention published in different countries and in various languages, which separates out significant departures from original filing of the invention. An algorithm identifies the closely related members of the family through direct priority matches and also indirectly linked family members and creates links between related families. A dedicated team of experts make the final decisions and also add to the family the non-Convention equivalents. These are documents without priority information which have been filed outside the terms of the Paris Convention (after the 12 month priority period or by non-signatory countries).

This intellectual clean-up of raw patent data during *DWPI* processing is one of the significant ways in which Thomson Scientific adds value to patent information.

Sectioning, classification and routing

In order to receive attention from the appropriate expert scientists and engineers, patents describing new inventions are categorised according to their technical content, and routed to the relevant editorial department for review and analysis by an expert in that field. Inventions covering multiple technical areas are routed to each appropriate area for attention.

Inventions are divided into 21 broad subject areas or Sections:

- A-M (Chemical);
- P-Q (Engineering); and
- S-X (Electronic and Electrical).

Within each section, inventions are further categorized in one or more Derwent Class.

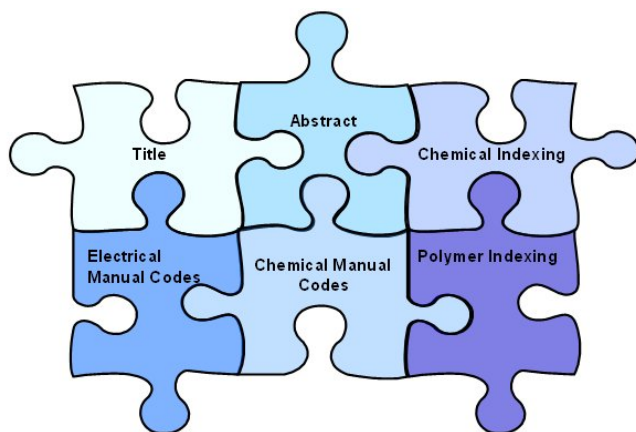
Each Class is comprised of the Section letter, followed by two digits, for example:

- X22 is the class designation for Automotive Electrics; and,
- C04 is the class for all Chemical Fertilisers.

Derwent Classes are consistently applied by subject experts. When used in combination with other online search terms e.g. Keyword Search, these Classes enable you to precisely and effectively restrict your search to the relevant subject area. Cross-classification ensures that all the patents of interest are located.

The value add jigsaw

In addition to the value added by intellectual clean-up of the raw patent data, there are a number of additional pieces of intellectual enhancement applied by Thomson Scientific analysts which together go to complete the jigsaw of added value.



The title

The title for a *DWPI* record is written to cover the scope, use and novelty of an invention: in other words what it's for, how it's used and what's new about it. Regardless of the source language, the *DWPI* title will be in English. Titles are designed to be easy to scan and to provide a high level summary of the invention, enabling searchers to quickly identify which patents are important to their work and hence save time and money by making the best use of their resources.

Examples:

- **Author Title:** Phenol derivatives for treating multiple sclerosis
DWPI Title: Use of 4-aminoalkyl-phenol derivatives, 4-hydroxybenzamide derivatives and (4-hydroxy-phenyl)-alkanamide derivatives for treatment of multiple sclerosis
- **Author Title:** A method for conditioning organic pigments.
DWPI Title: Pigmentation composition for macromolecular substances, coatings and inks contains an organic pigment conditioned with a surfactant

The abstract

Depending on the technology, Thomson Scientific subject experts write abstracts detailing the claims and disclosures of the inventions, and highlighting in more detail the main uses and advantages of the technology. These are written in English regardless of the original language of the patent. Details are provided under various headings which include:

- Novelty
- Detailed Description
- Use
- Advantage
- Activity
- Mechanism of Action
- Description of Drawing(s)
- Technology Focus
- Independent claims
- Specific compounds
- Preferred features as indicated in subordinate claims
- Specific example

Standard abstracts (Alert abstracts) are produced for Electrical Engineering Sections S-X. The most detailed abstracts (Documentation abstracts) are produced for Chemical patents in Sections A-M. The abstract is designed to provide more information than the author abstract

and to cover all relevant information so there is no initial need to go to the original patent specification.

Manual coding and deep indexing

In addition to producing the title and abstract, Thomson Scientific experts apply a number of manual codes relating to the classification and where appropriate deep indexing to provide uniquely powerful and precise search and retrieval capabilities to enable customers to focus on information relevant to their research.

For electrical engineering inventions including areas such as computing, communications, power generation, automotive electronics and domestic electrical appliances, a system called the *Electrical Patents Index* Manual Codes (EPI Manual codes) is used to further classify technology by the claimed novelty of the invention, as well as the application described in the main body of the patent document. This is a hierarchical system, intended for use as an online retrieval tool for abstracts of Electrical and Electronic engineering patents. It was introduced in 1980 and applies to all *DWPI* engineering records from that time. The classification was extended in 2006 to include mechanical transportation inventions and currently comprises around 9,000 codes.

For chemical inventions a corresponding system, the *Chemical Patents Index* (CPI) Manual codes, has been applied to chemical inventions in *DWPI* for over 35 years and currently comprises around 8,700 codes. These are applied on the basis of the title, novelty and use/advantage sections of the abstract and are used to describe the novel features of the invention.

Additional deep indexing tools are available for chemical patents that cover Pharmaceutical, Agrochemical and General chemistry (Chemical indexing) and for Polymer chemistry (Polymer indexing). Deep indexing covers more comprehensive retrieval systems where the compounds and concepts that are important in the patent are associated with the functions/activities they have and/or describe the processes carried out on them.

Further details of these classification and indexing systems are available at <http://scientific.thomson.com/support/patents/dwpi/ref/reftools/classification/>

Journey's end

Once the preliminary processing to . . .

- convert and correct the raw patent data
- standardize company names
- distinguish new inventions from patent family members
- provide a value-added title and detailed abstract

- classify and index to provide rich and comprehensive search and retrieval capability

. . . is complete, the final stage is to provide the *DWPI* record to professional search platforms including the online hosts and *Delphion*. Here the record is added to update the *DWPI* database where users can search and precisely retrieve patents information relevant to their research needs from over 40 patent issuing authorities spanning over 40 years.

Who can imagine what technological developments will take place over the next 40 years? Whatever they may be, Thomson Scientific will be there to abstract and index the inventions for inclusion to the *DWPI* database to provide customers with access to the world's most comprehensive and valuable source of value-added patent information.