Adding value in the DWPI by correcting bibliographic errors in the original documents

Correcting errors in the original patent content is a key aspect of the DWPI value add, providing searchers with the most comprehensive and accurate value added patent content in one file – the Derwent World Patents Index.

In 2009 over 3.13 million patents were loaded into DWPI, which is a 14.5% increase on the previous year. Processing this volume of data is a huge task, particularly as the original patent data can contain errors in the bibliographic data, which could impact the accuracy in searching.

At Thomson Reuters, the DWPI production teams use both programmatic rules and human endeavour to identify and correct as many of these mistakes as possible, to ensure the accuracy of key value add elements such as the DWPI patent family and Patent Assignee codes, and also improve searching through the addition of IPCs.

Each week data from the various patent offices is loaded (receipted) into the DWPI production system and conversion/validation programs are run to standardize all of the bibliographic data fields into a standard DWPI format. Through this process records with errors or missing data are identified and are routed to experts in the production team to be manually corrected. Typical errors includes incorrect priority or application numbers, or wrong patent assignee names, or missing data such as the International Patent Classification (IPCs).

In the first two months of this year the DWPI production teams made corrections to 6198 records with incorrect or missing priority information, and have manually applied IPCs to 450 PCT records that were published initially without any classification. Corrections to the priority data improve the accuracy of the DWPI patent family, as this is a key data element in bringing together patents relating to the same invention.

The example shows an incorrectly spelt patent assignee name on a US Patent Application. The company name was manually corrected in the DWPI record, and the relevant DWPI Patent Assignee code was applied.