World Patent Information 31 (2009) 193–198

Contents lists available at ScienceDirect

World Patent Information

journal homepage: www.elsevier.com/locate/worpatin

A practical approach to the reformed IPC

Susan E. Cullen

Director of IP Consulting, Scientific, Thomson Reuters, 3501 Market Street, Philadelphia, PA 19104, United States

ARTICLE INFO

Keywords:
Reformed IPC
Backfile classification
IPC term frequencies
Search strategies

ABSTRACT

In a global collaborative effort, the Version 8 International Patent Classification Codes (IPC-8) have been applied to patent documents published prior to 2006 (backfile). This unprecedented effort employed multiple techniques, including use of earlier IPC versions, use of ECLA codes, and propagation of codes from one document to another within Inpadoc patent families. The result is that most records bear a relatively high number of codes, and family members from different authorities are coded relatively uniformly, at least at present. It is also notable that patent documents are not easily differentiated based on selective use of inventive codes because the vast majority of codes have been designated as inventive even where they cover subject matter that would be likely to figure as additional information in a substantial number of cases. Creating high precision search strategies relying on codes may be helped by requiring 2 or more codes to be present on individual documents or families. Sample searches presented are exemplary, and search strategies must be tested individually for optimizing searching in each subject matter area. Whether the reclassification effects and the propensity to make most codes "inventive" makes the IPC more akin to an indexing system is a question only time will answer.

1. Reclassification of the backfile

After introduction of the new version (IPC-8, IPC-R) of the International Patent Classification codes in 2006 [1–4], a large cooperative effort between patent offices was launched to reclassify pre-2006 publications (the "backfile"). This effort employed machine-aided conversions of intellectually applied classifications, effectively using ad hoc concordances to assign IPC-8 codes to documents bearing the older code versions. The ECLA codes were heavily relied on to assist reclassification, and the new IPC codes were propagated across invention families (Inpadoc). With this task largely completed, it is possible to examine the effects of the reclassification on searching pre- and post-2006.

Because of the quantity of patent data, the examination reported here[1] is both selective and eclectic, and is meant only to demonstrate some of the results that are encountered using searching strategies that rely on IPC-8. Several systems were used to perform the searches described, and included both single document searches (full text systems) and patent family searches (bibliographic abstract systems). The results with the single-document systems are presented, but both kinds of systems offer similar results.

Four IPC code areas that underwent considerable change in 2006 were selected for testing [see Table 1 for definitions]. The codes areas were:

(1) Personal Care: A61K 8/ which replaced A61K 7/, and which may be accompanied by A61Q, a newly established subclass covering applications of A61K 8/ technology.
(2) Vehicle control: B60W.
(3) Combinatorial Chemistry: C40B.
(4) Plant and Insect Control and Growth Regulators: A01P.

Two of these codes (A61K 8/ and B60W) are code areas that are of the "inventive" type and are permitted to be applied as a sole code. The other codes examined (A61Q, C40B and A01P) may not be applied as sole codes and can be included as additional information by examiners or classifiers to describe the invention more fully. They are also permitted to be used as "inventive" codes.

In a series of four figures covering the filing period 1995–2007 (Figs. 1–4) the number of EP documents bearing the codes in question was plotted along with comparator datasets retrieved by using the codes shown. Where possible, the retired code was added as a second comparator for the newly introduced codes. In cases where there was no unitary correspondence between the old and new codes, a relevant surrogate was used as a comparator.

In the cases of A61K 8/ and B60W, the number of documents retrieved was quite substantial. The result for A61K 8/ and A61Q (Fig. 1) mirrored the application of A61K 7/ codes in the backfile, with A61K 7/ trailing off as expected after 2006. Notably A61Q, which is a non-obligatory code, gave yields very similar to A61K 8/.

The result for B60W (Fig. 2) showed quite parallel changes over time between B60W and a related code B60L 11/, but the yield between the two was dissimilar due to incomplete overlap between

E-mail address: sue.cullen@thomsonreuters.com

[1] The results in this article were reported in part at the 2008 Annual Meeting of the Patent Information Users Group [PIUG] in Arlington VA, USA May, 2008.

0172-2190/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wpi.2008.12.008

Table 1
Brief Definitions for Code Areas from WIPO <http://www.wipo.int/classifications/ipc/ipc8/?lang=en>.

[8]	A01N	PRESERVATION OF BODIES OF HUMANS OR ANIMALS OR PLANTS OR PARTS THEREOF; BIOCIDES, e.g. AS DISINFECTANTS, AS PESTICIDES, AS HERBICIDES
[8]	A01P	BIOCIDAL, PEST REPELLANT, PEST ATTRACTANT OR PLANT GROWTH REGULATORY ACTIVITY OF CHEMICAL COMPOUNDS OR PREPARATIONS
[8]	A23K	FODDER
[8]	A23L	FOODS, FOODSTUFFS, OR NON-ALCOHOLIC BEVERAGES, THEIR PREPARATION OR TREATMENT, e.g. COOKING, MODIFICATION OF NUTRITIVE QUALITIES, PHYSICAL TREATMENT
[7]	A61K 7/	Cosmetics or similar toilet preparations
[8]	A61K 8/	Cosmetics or similar toilet preparations
[8]	A61Q	USE OF COSMETICS OR SIMILAR TOILET PREPARATIONS
[8]	B01J 19/	Chemical, physical, or physico-chemical processes in general
[8]	B41M	PRINTING, DUPLICATING, MARKING, OR COPYING PROCESSES;
[7]	B60K 41/	Conjoint control of drive units; Conjoint control of at least two sub-units thereof
[8]	B60L 11/	Electric propulsion with power supplied within the vehicle
[8]	B60W	CONJOINT CONTROL OF VEHICLE SUB-UNITS OF DIFFERENT TYPE OR DIFFERENT FUNCTION; CONTROL SYSTEMS SPECIALLY ADAPTED FOR HYBRID VEHICLES; ROAD VEHICLE DRIVE CONTROL SYSTEMS FOR PURPOSES NOT RELATED TO THE CONTROL OF A PARTICULAR SUB-UNIT
[8]	C07H	SUGARS; DERIVATIVES THEREOF; NUCLEOSIDES; NUCLEOTIDES; NUCLEIC ACIDS
[8]	C07K	PEPTIDES
[8]	C12N	MICRO-ORGANISMS OR ENZYMES; COMPOSITIONS THEREOF
[8]	C40B	COMBINATORIAL CHEMISTRY; LIBRARIES, e.g. CHEMICAL LIBRARIES, IN SILICO LIBRARIES
[8]	G01N	INVESTIGATING OR ANALYSING MATERIALS BY DETERMINING CHEMICAL OR PHYSICAL PROPERTIES
[8]	G03F	PHOTOMECHANICAL PRODUCTION OF TEXTURED OR PATTERNED SURFACES, e.g. FOR PRINTING, FOR PROCESSING OF SEMICONDUCTOR DEVICES

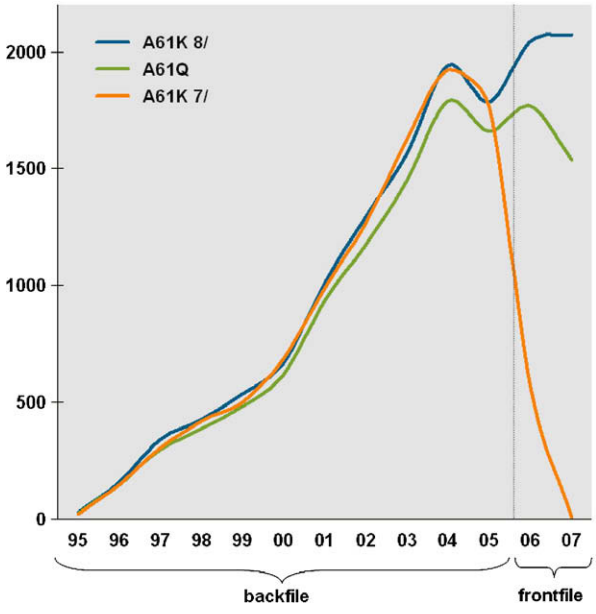

Fig. 1. Comparison of three different code limitations run against EP documents during the publication period Jan 1, 1995–May 2, 2008. The documents retrieved were publication-stage deduplicated to avoid double counting. The retired code A61K 7/ is attached to documents with dates spanning 1995–2005 in the backfile, but of course this code has not been applied to documents filed after 2006. The two new codes A61K 8/ and A61Q are applied to documents both pre- and post-2006. The number of documents in the backfile that have all three codes is quite similar, probably reflecting well-defined relationships between old and new codes. The total number of documents examined was about 15,000.

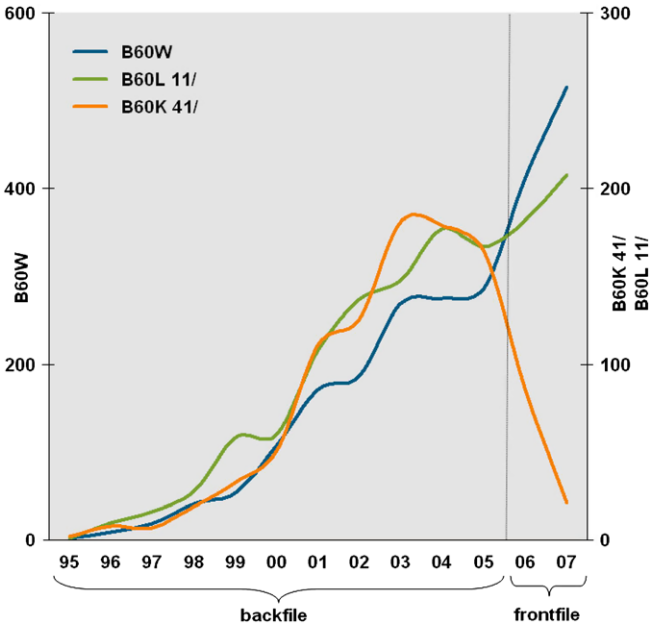

Fig. 2. Comparison of three different code limitations run against EP documents during the publication period Jan 1, 1995–May 2, 2008. The documents retrieved were publication-stage deduplicated to avoid double counting. The retired code B60K 41/ is attached to documents with dates spanning 1995–2005 in the backfile, but of course this code has not been applied to documents after 2006. The two new codes B60W and B60L 11/ are applied to documents both pre- and post-2006. The number of documents in the backfile that have the B60L 11/ and B60W codes is about half the number carrying B60K 41/. This disparity is due to only partial overlap between old and new codes. The total number of documents examined was about 2500.

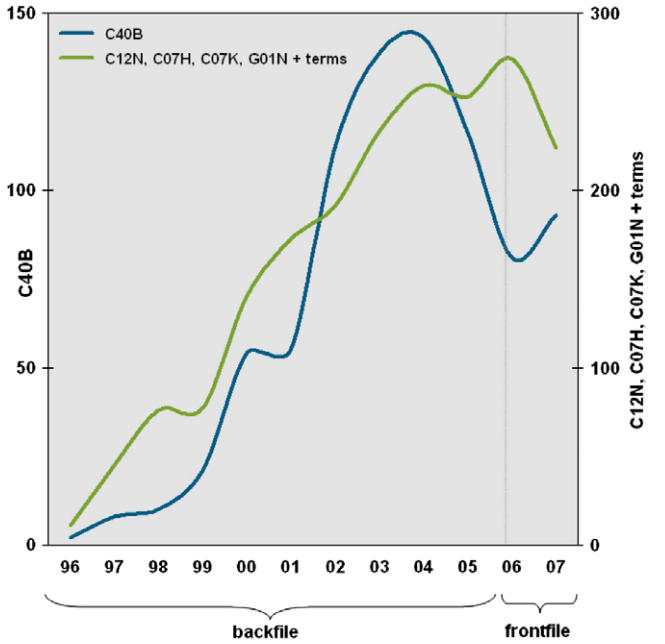

Fig. 3. Comparison of two different code limitations run against EP documents during the publication period Jan 1, 1995–May 2, 2008. The documents retrieved were publication-stage deduplicated to avoid double counting. The new code examined is C40B, but it supplements rather than replaces existing codes. In order to obtain a rough comparator, the C12N or C07H OR C07K OR G01N codes were searched together with key terms (librar* or combinatorial). The number of documents in the backfile that have the C40B code trails that of the comparator. Codes like C40B are potentially very useful, except that they are not permitted to be used as the first-listed code, and thus may not always be applied even when applicable. The total number of documents examined was about 3000.