World Patent Information 34 (2012) 8–18

Contents lists available at SciVerse ScienceDirect

World Patent Information

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

Review Article

It takes all kinds to make a world – Some thoughts on the use of classification in patent searching

Bernd Wolter

Siemens AG, P.O. Box 22 16 34, D-80506 Munich, Germany

Keywords:
Patent Classification
IPC
ECLA
DeKla
File Index
F-term
Derwent manual codes
USPC
Review
Classification systems comparisons
Wind motors

A B S T R A C T

This article addresses patent classification as practiced by different patent offices and one private company. It highlights strengths and weaknesses of the individual systems. To this end it compares the representation of the concept of controlling wind motors in the classification systems of WIPO, EPO, DPMA, JPO, USPTO and Derwent's World Patent Index. As the author is a searcher in the mechanical and electronic arts, specific statements and examples will also relate to these technologies. A comparision of search results obtained by using different classification schemes will show how additional subject matter may be found. The incompatibilities of IPC-based and non-IPC systems and how they can be overcome are exemplified. Possible reasons for differences in application of the classifications are discussed and what these may mean for the searcher. In the conclusion reasons for expending the additional effort of applying several classifications will be given and the author argues why it is a good thing there are more than just one or two classification schemes.

1. Introduction

In the perception of the general public the language of patenting has a bad reputation. It is perceived as obscure and more obfuscating than enlightening. Reading, writing and interpreting patents remains a task for specialised professionals who feel an obligation toward their clients not to give too much away in their applications while at the same time expecting everybody else to use concise language in theirs. Patent searchers are caught between a rock and a hard place here.

The last years have seen lots of promise for patent searching - development of intelligent search tools that no longer just do what you say but also what you mean. Although one can already foresee that the developers of those systems will grossly underestimate the ingenuity of human (and lawyer's) minds in formulating patent applications, these directions in computer linguistics are nevertheless interesting and should be closely watched and tested.

While waiting for fulfillment of these promises professional patent searchers will have to resort to predigested and readily available knowledge: it lies in the answer to the question why one would want to use more than one patent classification system. While the general use of the International Patent Classification (IPC) is uncontested there is also a general wondering and doubt as to the wisdom of having several systems - not only in managerial and lawyer circles but also sometimes among patent searchers.

2. Classification systems

Only a few years ago patent searching was a highly specialised and costly activity only to be trusted in the hands of specially trained people. With the ease of access to either publicly available sources like EspaceNet, the national offices websites and the likes of Google Patents or in-house databases many more people feel they are now also patent searchers. Apart from missing experience with the intricacies of fully fledged patent databases and their multitude of data fields and operators they often have only a rudimentary knowledge of the different patent classification systems available in those databases.

2.1. IPC

The best known and most widely spread classification - both in application and usage - is of course IPC. For those interested in its history I recommend the World Intellectual Property Organization (WIPO) website [1] and also the archives of this publication. This article will not try to be a classification primer - if you want to learn how to use it in patent searching you will find starting points at database hosts like STN [2] or at patent offices [3]. For in-depth

E-mail address: bernd.wolter@siemens.com.

0172-2190/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wpi.2011.08.001

additional reading on individual systems that cannot be the subject of this article see also the list of articles at the end [15–23].

What is of interest in the context of the present article are the changes - and the reasons for them - in the last five years. As the interested searcher knows, the two gravest birth defects of the original IPC were its long revision cycles and the failure to reclassify the back-file after each revision. These were due to the technological limitations at the time of its inception. So we have to hold it to the WIPO as the keeper of the IPC that they have modernised it when these limitations no longer applied (or shortly thereafter). After some deliberations it was decided that the then upcoming 8th edition of the IPC in 2005 should be the first one after the new fashion. Its major changes were quarterly revisions, back-file reclassification and the splitting of the classes into core and advanced level for smaller and bigger patent offices.

With the shorter revisions new developments in technology could now be included much faster into the classification scheme (e.g. hybrid internal combustion/electrical drives (B60W) made it into IPC in record time). Patents that already dealt with that technology but were applied for before its inclusion into IPC, would now get the most accurate class possible when reclassified. Not all of the original ideas turned out to be of universal use and again WIPO has to be lauded for correcting those shortcomings: quarterly revision turned out to be impractical and therefore will as of 2011 be yearly. Membership of the club of those patent offices that classified into advanced groups, turned out to be a matter of national pride and only very few offices remained classifying according to core level, so this distinction that led to unnecessary confusion has also fallen with the 2011 revision.

Another wise decision was to pass the actual implementation of the classification effort into the hands of the European Patent Office (EPO) who had already considerable experience with this kind of work (as will be seen shortly). The EPO became the keeper of the Master Classification Database (MCD) to care for the vast repository of classified patent documents and vouch for application of classification to the documents and its correction should the need arise. The massive effort of both the 2006 and on-going reclassification of older documents nevertheless is shared by a number of patent offices around the world. To ease the workload they introduced the concept of the simple patent family (all family members sharing exactly the same priorities) which inherit the classification symbols from their founding member. With these measures in place, the former unwieldy and often error-prone system has become flexible and rather quick to respond to changes. Where before - not to miss documents - you had to follow sub-groups back in time through several revisions to catch possible changes in the scheme, you can now be much more confident to find most of the state of the art with one classification symbol.

The classification definitions can be found in French and English at the WIPO website (and for Germans at the Deutsches Patent- und Markenamt (DPMA) website). Novice and experienced searchers may wish to read through the FAQs there also - they are a source of very valuable background and practical information on the IPC!

Now, what are the advantages of all this? Every patent document published in the 59 states in WIPO's list [4] will carry at least one IPC symbol - notice that not only the signatories of the Strasbourg Agreement use IPC but as WIPO holds "the IPC is used in more than 100 countries in the world". That means that you can search subject matter covered by the IPC in the patent literature of all those countries without speaking their languages, or even being able to read their scripts. And it means no matter what patent database you use (EspaceNet, some national office's website, commercial abstract or full-text databases, even Google Patents) IPC will be there to be searched - although you will have to watch out for differences in format. For the purpose of this article I will stick to the F03D7/00 notation of most of the office's websites, although WIPO Standard 8 (ST.8) calls for zero-padded four digit main groups – some database providers go with that, others use just three digits, hyphens or no padding at all. This often is a source of confusion: you should take care to check the customs of every new provider you use.

But this ubiquity comes at a price: to make it easier for patent offices to classify patents into IPC, it is not as detailed as it could be. Now you might say that 70,000 sub-groups is quite a lot of choice, but in view of the massive increase of publications the patent world has seen in the last 20 years and the increasingly detailed and small-scale improvements that are nowadays the subject of applications, it is often not enough. If you are searching in an area where the state of the art is advancing in very small increments you will often find yourself wishing for sub-groups not containing thousands of documents. In some art units IPC is hardly ever used for searching (most notably the chemical and pharmaceutical arts). Another problem is the application of IPC by a huge and very diverse body of examiners (and other PTO personnel) that are not always really convinced of the necessity of a thorough and meticulous effort in classification (you may have heard of PTOs using (semi-) automatic systems).

Nevertheless IPC remains one very important arrow in a patent searcher's quiver. For the purpose of this article let's assume you were asked to find patents pertaining to the following question: "How to remotely control and regulate the temperature of an offshore gear-less wind turbine generator?" Let's only pick the "control of wind turbine" aspect and consider the place it may be found in the different classification systems, starting with IPC (see Fig. 1). Note that this example was mainly chosen for its clarity and brevity - so it fits all into one figure.

2.2. ECLA

Simply speaking the European Classification (ECLA) is the EPO's incarnation of the IPC. While sharing the section, class, sub-class and group symbols, it about doubles the number of sub-groups (140,000 instead of 70,000). But the country coverage differs - whereas we have seen that IPC is used by over 100 PTOs around the world, ECLA is only applied by a single authority to a subset of those, ie EP, DE, FR, GB, NL, BE, WO and US (their family members then inherit the classes). A list of sources and a short description can be found on the EspaceNet website [5]. I think the quote "Other documents may be occasionally classified at the examiner's initiative." from that site's description is key to the utility of ECLA: while IPC is applied to patent documents because national PTOs are obliged to do so, here (and in some other offices as well) it is done because the examiner has an immediate benefit from doing so. It is not done for publications nor for external use, but for re-use in future cases. For the EPO examiner there is a real labor-saving incentive in applying pertinent and consistent classification to patent applications!

While coverage of patent documents is smaller then that of the IPC, the level of detail is superior. Have a look at the same group that we have seen in the IPC in Fig. 2 below. In the layout you find on the EspaceNet website under "Classification Search" you can easily distinguish the ECLA from the IPC into which it is integrated: while the IPC sections, classes and groups remain unchanged in notation, there are sub-groups that contain only numbers (IPC) and others that also contain letters. The ones with letters also have their definitions start with "[N:]" which is the second clue that these are ECLA codes. You will notice that the original three sub-groups have been expanded into 17 giving more detail to subgroup 02 "the wind motors having rotation axis substantially in wind direction".

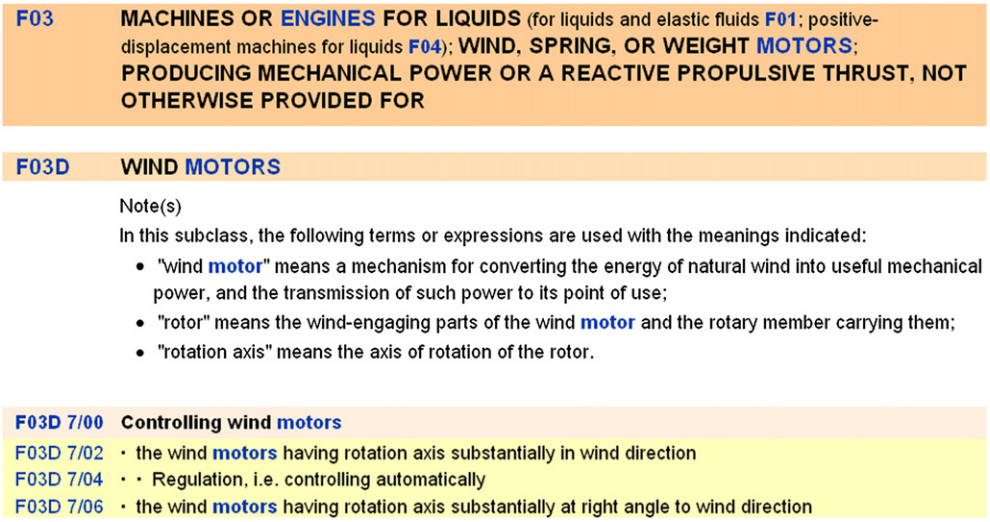

Fig. 1. F03D7/00 from the official WIPO IPC website.

The website is also a good starting point if you are not sure where to start your classification search: if you enter e.g. “control wind turbine” into the search field at the top of the page the keywords (ten at most) will be searched against the site’s patent database and the first thousand or so results will be statistically analysed for the most common classes. The top ten results will be marked with little squares: the more squares the more appropriate the class. From the presentation of the classification you can also directly accept symbols into a search by simply checking the boxes behind the sub-groups. With the notable exception of some national websites and Google Patents, ECLA can nowadays be searched in most patent databases - just check their data-sheets!

Nevertheless there have to be some words of caution: ECLA is not always completely in line with IPC, e.g. see the notes section of A61F. As we have seen above not every publication is automatically classified. Classification lags behind publication - there can be several months between publication and addition of ECLA symbols to the first family member. The definitions are subject to change without notice - you should not use them for long term surveillance projects without regularly checking whether the codes might not have changed. Also ECLA will not add much in the way of indexing to well-established technologies - see e.g. metallurgy of iron “C21” or mechanical metal-working “B21” which add very little in the way of deeper indexing when compared to other areas of technology.

Despite these uncertainties with regard to coverage and timeliness ECLA codes are subject to a rigid EPO internal regimen: there is a governing body within the office that watches over their definitions and application, they are systematically fitted into the IPC framework, redundancies are avoided and as soon as a subgroup

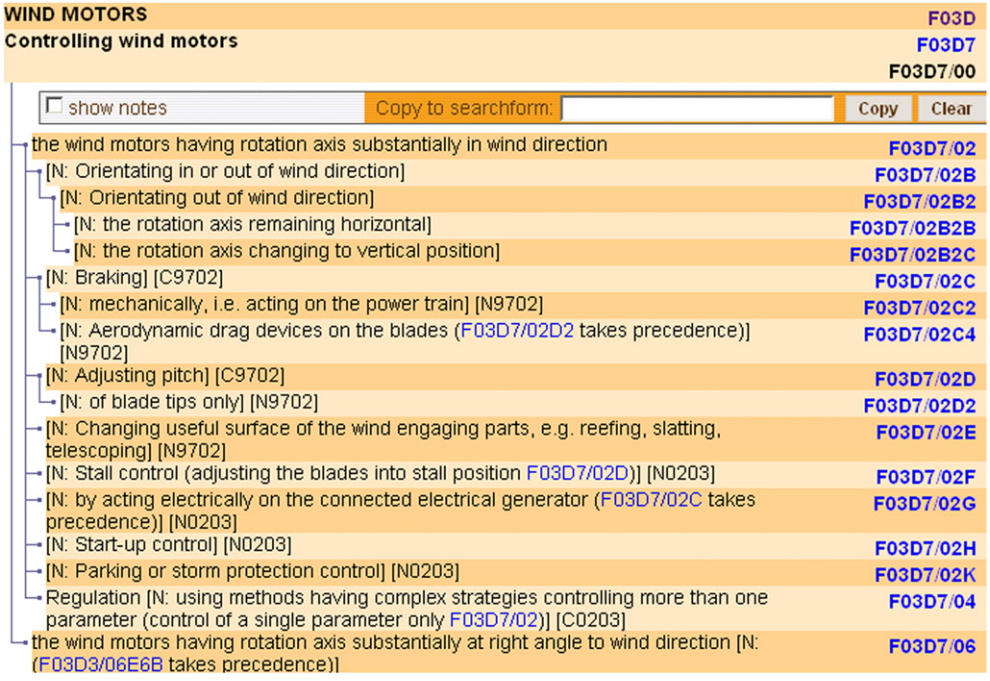

Fig. 2. F03D7/00 in ECLA.

R05B0240:21	NT9	for wind turbines
R05B0240:211	NT10	with vertical axis
R05B0240:212	NT11	of the Darrieus type
R05B0240:213	NT11	of the Savonieus type
R05B0240:214	NT11	of the Musgrove or "H"-type
R05B0240:215	NT11	of the panemone or "vehicle ventilator" type
R05B0240:216	NT11	of the anemometer type
R05B0240:217	NT11	of the crossflow- or "Banki"- or "double action" type [N9507]
R05B0240:221	NT10	with horizontal axis
R05B0240:2211	NT11	of the multibladed, low speed, e.g. "American farm" type [N0504]

Fig. 3. Excerpt from represenation of wind turbine related index terms in ICO R05B (with permission of STN International).

becomes too large in terms of number of publications, further subdivisions are sought. They are much more strictly systematised and applied within the context of the EPO's examination effort than IPC.

2.2.1. ICO

Parallel to ECLA the examiners at the European Patent Office use another classification scheme for their patentability searches: ICO. This acronym used to stand for "In Computer Only" (nowadays it is: "Indexing COdes") and this curious name gives a hint at its inception. Obviously the codes (like ECLA) are not published on any hardcopy patent and exist only in the office's internal databases. Some years ago they have escaped from there and are now available in some commercial provider's databases as well - reason enough to shortly introduce them here.

Although part of the EPO classification, ICO codes are not employed in all technical fields. In contrast to ECLA there is no obligation of the examiner to use them. So they are applied by the examiner for his or her own benefit in later searches - we have the same motivation as with ECLA.

Their systematics is similar to IPC or ECLA, which they sometime closely mirror. Each Section A To H has an equivalent section lettered K,L,M,N,P,R,S and T. The catch is that there is no guarantee that there will be a corresponding class, sub-class or group. Therefor the string of examples breaks at this point: there simply is no R03D7/00 corresponding to the F03D7/00 example above. If on the other hand you text search the classification you will find that there actually is a R05B indexing section that among other things contains details of wind motors (although no control aspects) and which has no F05 main class counterpart – see Fig. 3.

The ICO codes system has been described in more detail in a recent article by Goebel and Hintermaier [24].

So on the one hand there is not always an ICO code for every IPC or ECLA class but on the other hand there is subject matter classified that you might not find anywhere else (for an exmaple of an extensive index see S01N333, for an alphabetical index S06F201). There is no reliable documentation with respect to coverage and there is no knowing whether an examiner will systematically complete or continue to classify cases into ICO once started, or whether he or she - upon changing art units - will continue the work of their predecessor. Nevertheless the classification itself and its application is also monitored by a special EPO department.

Apart from some databases offering ICO code listings or thesauri the only offline compilation that I am aware of can be downloaded from STN's homepage. Just watch the file banner of e.g. InpadocDB and there you will find the place where the latest ECLA and ICO definitions can be downloaded monthly as PDFs. In addition you will find another file giving you last month's changes. But be prepared for 8000+ and 5000 + page files respectively. There has been talk about the EPO themselves putting together classification files for offline use as well.

2.2.2. Tagging codes

The same EPO Patent Information Conference that brought the last details of the IPC reform before its first revision also saw the presentation of the first of the so-called tagging codes. These are additional codes that allow access to certain cross-over

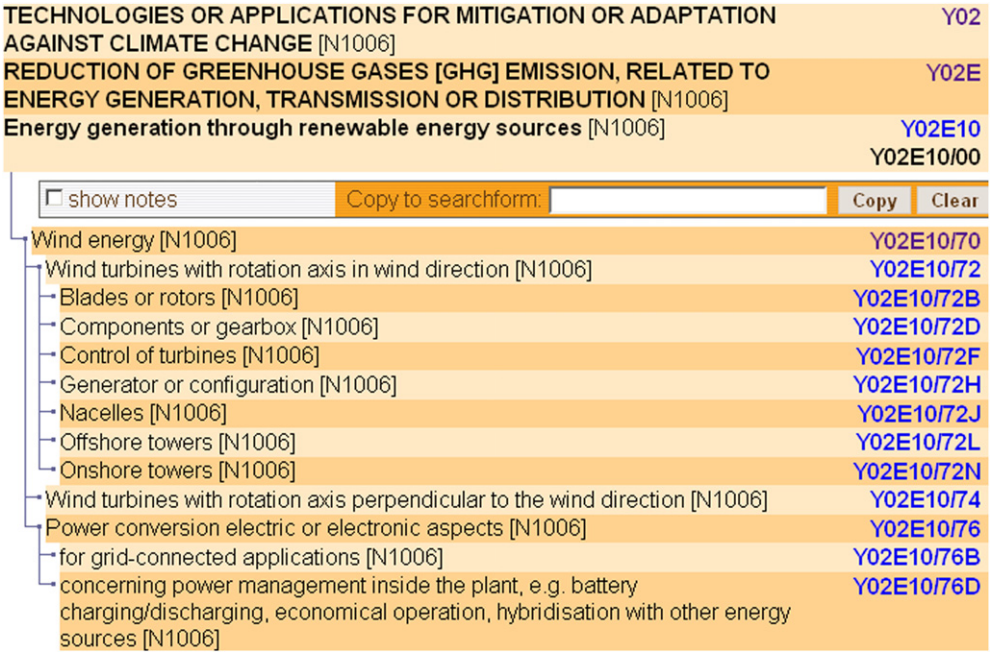

Fig. 4. Tagging Code Y02E10/70 from the EspaceNet classification website.

technologies that do not fit into single categories of the traditional scheme. The first one was Y01N dealing with nanotechnology [16]. It tries to collect applications from many fields that have a nanotechnology aspect (for the definition see the notes section of the respective classes). This code is then tagged on to the application without influencing the rest of its classification - hence the name.

This first attempt hinged upon the collection of carefully selected classes from all parts of the ECLA that had a connection to nanotechnology and automatically updating the collection in Y01N. The latest ones (Y02) announced at the Lausanne conference in 2010 are for "Technologies or applications for mitigation or adaptation against climate change" and are further subdivided into "Capture, storage, sequestration or disposal of greenhouse gases [GHG]" (Y02C) and "Reduction of greenhouse gases [GHG] emission, related to energy generation, transmission or distribution" (Y02E). This time experienced examiners have set up a mixture of classes and keywords to adequately describe the subject matter. Updates will be done automatically as soon as new applications are added.

The tagging codes are intended as an intermediate venture since it was announced in the meantime that with the addition of the changes in B82B and the newly introduced B82Y the nanotech inventions in Y01N will be migrated there - this means that Y01N will disappear in the immediate future.

In the daily life of a patent searcher tagging codes may not play a big role but they come in very handy for the subdivision of big projects or as welcome additional material for searches that may have otherwise exhausted the use of the other classification systems. Here is the place which fits the rest of this article's examples-see Fig. 4.

2.3. DeKla

Yet another classification system derived from IPC is the German DeKla classification. It is similar to ECLA in that it was meant only for the examiner's "private" use (hence it DPMA's internal name of "Prüfstoff-IPC" which might be translated as "examination material IPC"). It is different from ECLA in that there are no rules for its application, neither is there any kind of systematic supervision over its content or application. I couldn't find any numbers for the subgroups added to the IPC. The starting point is always the latest edition of the IPC and the symbols added consist of letters and numbers added to the sub-groups.

As you can see from this example (Fig. 5) the German examiner thought it necessary to add to the group definition directly ("controlling wind motors") as opposed to ECLA where the additions were mostly to subgroup 02 ("wind motors having rotation axis substantially in wind direction"). If you check both the "DE" and "EN" boxes and "DEKLA groups" in the IPC index on the German office's classification website [6] you will see immediately where sub-groups were added because the English IPC on the right side will remain empty. At the moment DEKLA descriptions are only available in German but a translation into English is in the works.

7/00	Controlling wind motors	3H078
7/02	. the wind motors having rotation axis substantially in wind direction	3H078
7/04	. . Regulation, i.e. controlling automatically	3H078
	A Control of the number of revolutions	3H078
	B . Rotor's rotary surface deflected	3H078
	C . . Deflection to a side	3H078
	D . . . using the side vane	3H078
	E . Variable pitch control	3H078
	F . . utilising the centrifugal force that acts on the governor	3H078
	G . . utilising the centrifugal force that acts on the blade	3H078
	H . . Electrical	3H078
	J . using a spoiler	3H078
	K Azimuth control	3H078
	L . using an airflow direction detection device	3H078
	M . utilising the resistance force that acts on the windmill	3H078
	Z Others	3H078
7/06	. the wind motors having rotation axis substantially at right angle to wind direction	3H078
	A Savonius windmill	3H078
	B Paddle type windmill	3H078
	C Darrieus windmill	3H078
	Z Others	3H078

Fig. 6. F03D7/00 in the FI listing from the IPDL (note the links to the F-term schedule on the right hand side).

Although there is little indication as to what exactly is classified (both in countries and publication period) it is always worth a try to find patents that might not otherwise have been subjected to "deep indexing" - especially in art areas where German industry is traditionally strong. A second weakness is that applications classified with DeKla are at the moment only available either in the database accessible through the DPMA's website or through Minesoft's PatBase. Nevertheless you might see it as an access directly to the examiner's desk - you will know what the examiner thinks are the most pertinent prior art in their respective art units.

2.4. File Index

We now come to the first of two classification schemes added to IPC by the Japanese Patent Office (JPO) [19]. File Index (FI) is based on IPC4. Its structure is the same as IPC as can be seen in the next screen-shot (Fig. 6). In this case obviously there was a need felt to

F03D 7/00	**Steuern oder Regeln von Windkraftmaschinen**	**Controlling wind motors**
F03D 7/00 A02	. Bauart und Verstellung der Windfahne	
F03D 7/00 A04	. Windrosen/Hilfswindräder zur Regelung	
F03D 7/00 A06	. Abfahren, Bremsen (Bremsen an sich F03D 11/00)	
F03D 7/00 A08	. von Windparks [2009.01]	
F03D 7/02	. bei im wesentlichen in Windrichtung liegender Drehachse	. the wind motors having rotation axis substantially in wind direction
F03D 7/02 A02	. . Schwenken und Kippen des ganzen Windrades	
F03D 7/04	. . mit automatischer Regelung	. . Regulation, i.e. controlling automatically
F03D 7/06	. bei im wesentlichen rechtwinklig zur Windrichtung liegender Drehachse	. the wind motors having rotation axis substantially at right angle to wind direction

Fig. 5. F03D7/00 in DeKla from the DPMA's classification website.

add more detail to subgroup 04 ("Regulation, i.e. controlling automatically"). Although the Japanese office is also bound by the Strasbourg agreement to publish documents classified into the most current IPC, the File Index is its major tool for internally subdividing its collection. FI adds another 100,000 sub-groups to IPC. These take the form of either single letters or three digits subgroups. You best consult your database provider's manuals to find out how to correctly enter the classes when searching.

The JPO's website provides a good starting point for searching or browsing the scheme: choose "Patent Map Guidance" in the Industrial Property Digital Library (IPDL) [7] where you can enter a known IPC symbol into the FI search box to see what additions there are. Sometimes you may even find that the examiners have introduced groups for information contained in the application's figures that could otherwise only very difficultly be described with words: for an example search H01R23/68/310/E – and following the link there you will see which kind of connector you can find using the code.

FI is only attached to either original Japanese filings or to PCTs entering the national phase in Japan. Although this coverage is limited, it gives the searcher a unique tool to systematically search a huge amount of data which is otherwise only available to speakers of Japanese. Being based on IPC4 FI is very stable in its structure and most of the refining happens only at subgroup level. Unfortunately this also means that changes through revisions of the IPC-like B82 are there only down to group level and others like H04W or B60W ("… Control systems specially adapted for hybrid vehicles …") have not yet arrived at all in FI. The latter case is especially curious as hybrid cars are a strong asset of the Japanese automobile industry - actually these will still have to be searched in e.g. B60K6/00 ("Arrangement or mounting of plural diverse prime-movers for mutual or common propulsion, e.g. hybrid propulsion systems comprising electric motors and internal combustion engines").

2.4.1. F-term

Closely related to FI but totally different in layout and about double its scope are F-terms (for File Forming Terms). They are related in that there is an intimate connection between FI (in its IPC-like structure) and F-terms: for every F-term there is a corresponding entry in the FI tree. This is the most convenient way of accessing them: look up the FI symbol you already know (from IPC) and in the IPDL (see Fig. 6 above) you'll have on the right hand side a link into the F-term matrix. F-terms consist of a theme code (digit, letter, three digits) and are then subdivided into view points which then again are made up of the F-terms proper.

Another way to get access is starting in Patent Map Guidance by clicking on F-term and then choosing from the Theme Selection Groups. This makes only sense when you already know your way around the F-term groups but may be useful if you want to further explore the groups in the neighborhood of one you already are aware of.

To keep with our example look at the following excerpt from 3H078 "Wind Motors" corresponding to F03D in Fig. 7. You should be aware of the fact that 3H078 covers not only F03D7/00 but F03D1/00 through 11/04 (see "FI Cover Range" at the top). You can see that view point AA "Types of wind motors" consists of 25 F-terms with space left for further subdivisions should they become necessary in the future. You will find room for things like AA14 "Bicycles" and AA27 "Conversion to mechanical energy" that you will not find in any other classification. While you can of course search for bicycles in title or abstract, a concept like converting wind power into mechanical energy will be hard to paraphrase successfully in what little English text you normally have for Japanese only patents.

F-terms are applied generously (actually they are not confined to the claimed invention but also cover aspects of the description) - documents with more than 30 F-terms are no exception. Therefore it is a good idea to "AND" them for precision rather than using them "OR-ed" together for recall, e.g. in the example above, 3H078/AA14 and 3H078/AA27 together yield six hits, while combined with OR there are 617. The same restriction apply as to the FI with regard to coverage. Unfortunately there are areas where the F-term definitions have not yet been translated into English rendering them inaccessible for non-Japanese speakers. So e.g. "4E019" has four view points with 42 F-terms, alas only in Japanese.

3H078	Wind motors F03D1/00-11/04@Z									
Viewpoint	F-term									
AA	AA00	AA01	AA02	AA03		AA05	AA06	AA07	AA08	AA09	
	TYPES OF WIND MOTORS	. Wind motors with a rotary axis that faces into the wind	. . Rotor types	. . Multiple vanes		. Wind motors with a rotary axis at a right angle to the wind	. . Paddles	. . Savonius paddles	. . Darius paddles	. . Sailwing paddles	
		AA11	AA12	AA13	AA14	AA15		AA17			
		. Wind motors with characteristic installation areas	. . Vehicles	. . . Automobiles	. . . Bicycles	. . . Trains		. . Ships			
		AA21	AA22	AA23	AA24		AA26	AA27			
		. Types of energy conversion	. . Conversion to heat energy	. . . Agitation system	. . . Throttle system		. . Conversion to electric energy	. . Conversion to mechanical energy			
		AA31	AA32		AA34						AA40
		. Combination of multiple wind motors	. . Combination of different types of wind motors		. Combination of wind motors with other types of energy-producing devices						. Others

Fig. 7. Excerpt from F-term 3H078 corresponding to FI F03D1/00 through 11/04.

To complete the number of singular classifications of the JPO one must also mention the broad so-called facets. There are eleven general ones of them ranging from ZAA "Nanotechnology" to ZYY "Vehicle behavior control by unspecified or multiple methods". Some of them apply across all areas of technology, others like ZHV "Hybrid vehicles" only to a narrow field. Facets can also be found for some groups like those in B21B for metal milling (like "BBJ - thick plate rolling mills").

It is an amazing amount of work that goes into these classes (imagine a Japanese examiner classifying a new application first into FI, then finding the appropriate F-terms and then also adding IPC for publication). This makes sense economically only if there is a reward for all that effort - in return it will make the examiner's future work easier and more precise. Searchers outside of patent offices can only be glad that the EPO, JPO and DPAM share their classifications and their in-house patent collections (without which all of this would make little sense).

There are two major classification systems left that share in all their differences some common ground: they are not derived or otherwise closely related to IPC - this is no shame and might lead to interesting new ways of access to documents; classifiers and users are not necessarily the same; both can be considered to have suffered from some neglect by their respective managements; they both suffer from legacy problems.

2.5. Derwent Manual Codes

The Derwent Manual Codes scheme is the only patent classification that is not given by a patent office but by a private company. It follows a different philosophy from IPC both in its selection of patent documents for inclusion into the patents database as well as in the application of the classification. Derwent (now a subsidiary of ThomsonReuters) selects its content from the whole of patent literature and splits it up into chemical, mechanical and electric patents. All of them are assigned a so-called Derwent class (a rough one letter, two digit code). Different areas of technology are then further classified: while chemical, agrochemical and pharmaceutical patents were the founding base of the whole service, electrical patents were only included from the mid-70s but they both get deeper indexing with what is known as Manual Codes (MC). Mechanical and general engineering publications are included but were until recently not deep-indexed. Nowadays there are manual codes for general engineering and there has been talk about upcoming Manual Codes for mechanics as well. This description will concentrate on the Electrical Patents Index (EPI) as the author cannot say anything substantiated on the Chemical Patent Index (CPI).

As already said, patents are selected for inclusion into the Derwent World Patent Index (DWPI) by their subject matter. Documents are selected either by their being already classified into certain IPC classes or "according to their electrical content" - a category not further explained on the ThomsonReuters' website. The EPI Manual Codes (about 9000 in all according to the same website) are then applied by indexers independent of original IPC.

Manual codes consist of a letter (A-N for chemical patents, P and Q for general and mechanical engineering and S-X for electrical engineering). They are further subdivided with two digits (these are the broad Derwent classes (DC) of the DWPI). ThomsonReuters' description of this scheme can e.g. be found on their website [8]. The chemical and electrical engineering (and to some extent mechanical engineering now also) sections are then submitted to finer classification with alternating letters and digits.

When this author first started patent searching many years ago the Derwent MC were (along with the rewritten titles and abstracts) the epitome of added value to the sparse official publication. The EPI was in several ways superior to the IPC: it was more clearly worded (they called a mobile phone a mobile phone and a laser a laser), it was much more up to date with respect to the latest technology (e.g. TCP/IP routing or many fields to do with automotive electrics), it was much more detailed in many areas (e.g. magnetic resonance imaging). Despite its limited scope (only covering 40 authorities in the 1990s) and the considerable rates for database use it was the first choice for all kinds of patent searches because of being an alternative to IPC and in combination with much more telling titles, abstracts and normalized vocabulary allowing very accurate searches.

Much has changed since: all of the above described classifications by the different patent offices have since become available, Derwent's policy of offering substantial added value seems to have been to some extent superseded by the desire to offer as many patents from as many national offices as possible with less care for the quality of the individual data item. Due to the rising costs for abstracting and indexing ensuing from the skyrocketing numbers of patents, rewriting, indexing and quality control are now being done in different parts of the world. Users experience lower quality both in abstracting and the application of the coding. Although the manual codes are revised almost every year now, this seems to happen only from customer's input and not by an internal review board which results in development of the scheme only in certain areas. And most importantly: the revised coding is not applied to the back-file, leaving you with several entries for some technologies with different starting dates which forces upon the searcher that once much dreaded following classification backwards in time as in the IPC days of old. If you compare Derwent code X15-B05 "Wind power: control, monitoring and testing" (Fig. 8) to the more detailed groups above you'll see how little detail there is even compared to IPC. "[2006]" in the listing indicates the time the code was introduced – what of patents published before that time? There is no indication where to search those with Manual Codes.

Notwithstanding the perceived loss of quality for the sake of adding more sources WPI remains a valuable database that - besides value added by rewriting titles and abstracts, offering the so-called standardised title terms and streamlining applicants

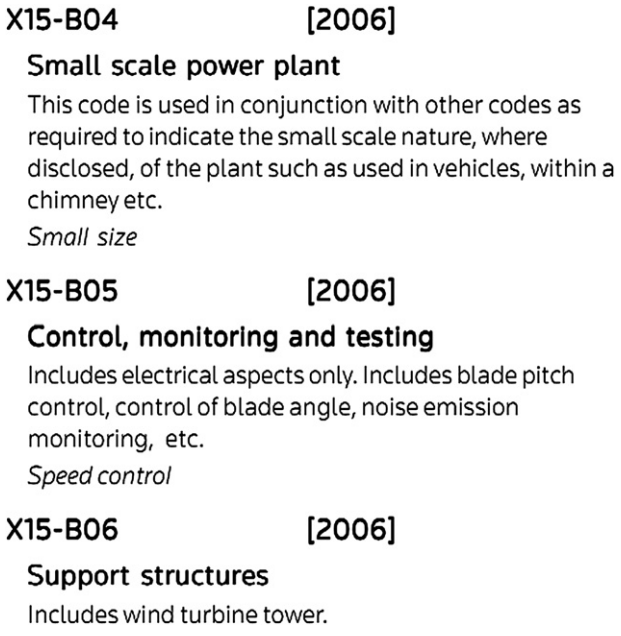

Fig. 8. Controlling wind power installations in Derwent Manual Codes.

names - offers a different access point to content by its own classification system (see also section “Comparing results” below).

2.6. US classification

The US classification is the oldest patent classification scheme still in use. It was for a long time also the only one with regular (sometimes bi-monthly) revisions and even more importantly those revisions were also applied retrospectively to the back-file (even beyond the beginnings of the classification (1836) back to the very first US patent in 1790). Searching for US X1 on the website of the United States Patent and Trademark Office (USPTO) you will find it having a currently valid US Patent Class (USPC). The classification covers only US documents (both published 18 months after application and granted). The division of subject matter into classes and subclasses (up to three and six digits respectively) is similar to the other systems we have seen up to now. But there most similarities stop: for searchers accustomed to the subject treatment of IPC-like systems it is often confusing to try and find one’s way around the USPC. You might be used to look for blood pumps under medical devices (A61), for fuel pumps under vehicles (B60) or combustion engines (F02) and for pumps “as such” under positive displacement machines (F04) - that is you go for the application you have in mind. Not so with USPC: here you have to consider the utility of the invention, the function of the procedure, operation or machinery, its proximate effect or the product produced or its structure. Others have explained this much better [9].

For the unaccustomed (and this author considers himself after many years of searching to still be one of them) problems do not end with having found the right class for his or her subject. Subclasses (digits number four to six) as such give no hint about their hierarchical place in the system - by just looking at the symbols you cannot tell whether they are headings or subheadings: you should always refer to the USPTO website [10] to make sure you are aware of the indentation level of the subclasses. Further, digits seven to nine are so-called decimal places and may offer further subdivisions of the subject matter at hand. Normally classes and subclasses are divided by a forward slash and subclasses and decimal places by a dot: this does not mean that you may just transfer your findings into the database of your choice - nearly every provider has its own way of implementing the USPC schedule (with or without slashes or dots, as is, or padded with zeros). Make sure to consult their data-sheets. Besides the subclasses there are also Digests, Cross-reference Art Collections, Alpha subclasses, E-subclasses and FOR subclasses, for a short overview see e.g. the wiki entry at “The Intellogist” [11]. Also due to using a paper based procedure for a long time the numbering of the subclasses is discontinuous as in the following example (Fig. 9). This is not only confusing but makes range searching next to impossible (but for the USPTO’s East/West system or using STN’s classification thesaurus).

Many US practitioners also lament the perceived deterioration of the once praised system especially in the mechanical and engineering arts. The classification department once 60 strong has fallen to two (!) experts - this is supposed to have come as a reaction of the promise of imminent “intelligent” searching methods and the rise of chemists with the different experience of searching in the life-sciences to decision-making positions within the USPTO. This downfall of the USPTO classification work is voiced in many publications (e.g. see “My 2010 wishes for the U.S. Patent Examiner” by Ron D. Katznelson [12] or Cecilia R. Dickson’s statement “This lack of expertise, combined with an erosion in the completeness and proper use of the Patent Office classification system in recent years, has resulted in the routing of nanotech applications to many different areas of the Office with little consistency.” [13] and through personal communication.)

But of course with the USPTO being one of the most important patent offices in the world, their huge treasure of prior art cannot

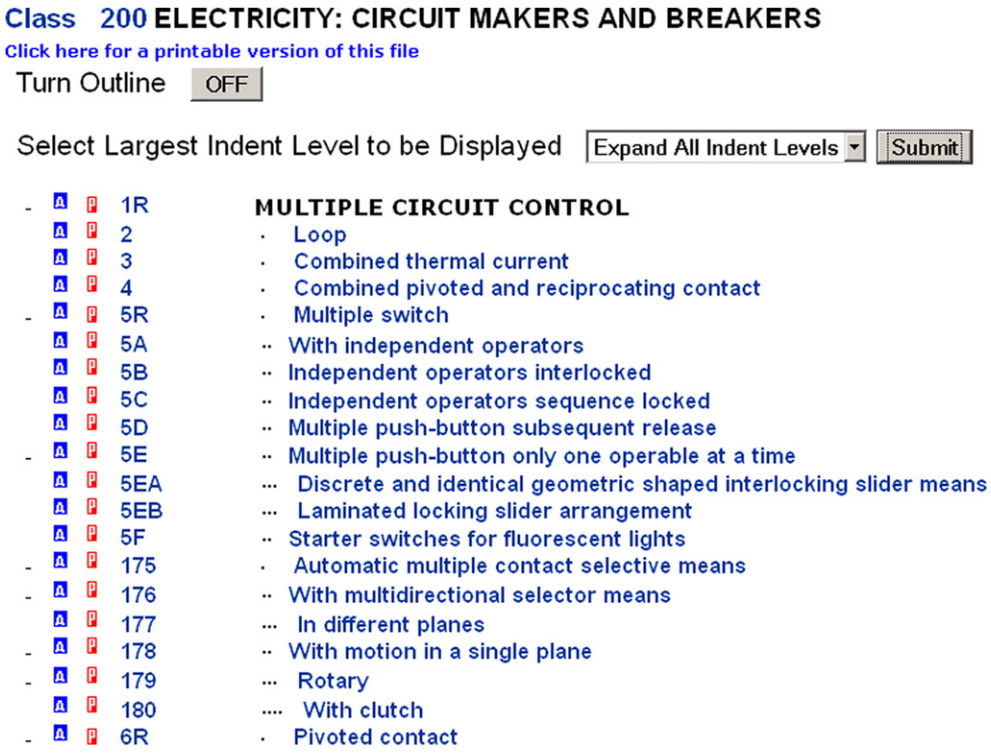

Fig. 9. Excerpt from USPC 200 showing discontinous numbering of headings.

sc=F03D7	7541
ic=F03D7	7261
ec=F03D7	2843
gc=F03D7	443
jci=F03D7	1197
sc=R03D7	0
ec=F03D7 not (ic=F03D7 or gc=F03D7 or jci=F03D7)	231
gc=F03D7 not (ic=F03D7 or ec=F03D7 or jci=F03D7)	19
jci=F03D7 not (ic=F03D7 or ec=F03D7 or gc=F03D7)	27
ic=F03D7 not (ec=F03D7 or gc=F03D7 or jci=F03D7)	3570

Fig. 10. Comparing results for different IPC-based classification schemes.

be ignored. So you are either versed in the classification system or you have to find ways around your ignorance. Even the above mentioned primer at "The Intellogist" proposes to find some documents by text searching and then start analysing their USPC and expand from there. Using databases that offer more than one classification (as most nowadays do) you may take another lane: search for the class symbol that you already know from your favorite scheme and statistically analyse the resulting hit set for USPC. This will show that the top postings of patents corresponding to IPC F03D7/00 turn up in US classes 290 (Prime mover dynamo plants), 416 (Fluid reaction surfaces) and 415 (Rotary kinetic fluid motors or pumps) - none of them even mentioning the word "wind" on the main headings level.

Not to forget the just started harmonisation of classification, as part of the IP5 project for a Common Hybrid Classification (CHC), between the European and US patent offices by means of the Cooperative Patent Classification (CPC): too little is known at the time of writing (January 2011) to say something meaningful yet. Certainly both do have areas where one is stronger than the other (e.g. US class 705 dealing with "Data processing: financial, business practice, management, or cost/price determination" which is not admissible subject matter under the European Patent Convention, but nevertheless important for searchers in an international company). If the new system is not lived by both corps of examiners the output from it may not be an improvement but a system that will in places be corroding from the inside.

3. Comparison of results

After now having had a look at some of the characteristics of the publicly accessible patent classification systems let me also try to compare the results one might get using either one or the other of those systems. First let's look at the results from those schemes that are based on IPC (Fig. 10). The database used was PatBase and there just searching F03D7 implicitly also searches all sub-groups. The number of hits is largest when using the super-search field "SC" because it collects all families that have F03D7 in any of the classification fields. Second is IPC, then the European class, then DeKla and last Japanese FI classification. Differences of the different systems were done to get those hits that were special to one of the classifications.

It turns out that especially for the German and Japanese hits there were many utility models in the lists - documents that were mostly classified for their general layout and less for the controlling aspect and for the Japanese documents never classified into IPC or European class at all. With the European class only documents it was similar: in IPC they were either classified according to layout ("Wind motors with rotation axis substantially at right-angles to wind direction" F03D3) or for details or component parts (F03D11) where in contrast the European examiner thought it necessary to add the control aspect for completeness. As already mentioned parallel ICO R03D7 does not exist. Even putting a lot of trust into those classifications should never lead to neglecting the one most widely used system: there are well over 3500 hits that have none of the above ECLA, DeKla, ICO or FI classes but only IPC and must absolutely not be left out of consideration: many of them opening access to data of e.g. Korean or Chinese origin without any text in a Western language.

F-terms do have a close relationship to FI which is derived from IPC. By the sheer amount of subdivision they play in a league of their own and cannot be compared to the other systems in the given granularity. Starting from FI they should always be picked for detail and in most cases must be combined with others to result in a meaningful and workable hit set. As already mentioned they are

```
L4   ANSWER 3 OF 309  WPIDS COPYRIGHT 2010       THOMSON REUTERS on STN
TI   Apparatus for controlling e.g. wind-powered generator for household
     application, controls discharge circuit and power inverter according to
     updated reference values for controlling generation of main electric power
IPCI G05F-0001/10 [I,A]; G05F-0001/10 [I,C]; H02P-0009/00 [I,A]; H02P-0009/00
     [I,C]
MC   EPI: U24-D05; X13-F25G; X13-G02T5; X13-G25G; X15-B01B; X15-B05; X15-C01;
          X15-C03; X15-G

L4   ANSWER 4 OF 309  WPIDS COPYRIGHT 2010       THOMSON REUTERS on STN
TI   Wind turbine facility power supplying arrangement of wind turbine facility
     switchgear power supply control system, has one or more parallel
     electrical power supply paths to provide electrical power to switchgear
IPCI F03D-0009/00 [I,A]; F03D-0009/00 [I,C]; H02P-0009/00 [I,A]; H02P-0009/00
     [I,C]
MC   EPI: U24-H; X15-B01A; X15-B05

L4   ANSWER 5 OF 309  WPIDS COPYRIGHT 2010       THOMSON REUTERS on STN
TI   Wind turbine, has control system detecting rotational speed of blades
     about axis and selectively resisting movement of blades to different
     incline positions of blades based on rotational speed of blades
IPCI F03B-0003/00 [I,C]; F03B-0003/14 [I,A]
MC   EPI: X15-B01A; X15-B05
```

Fig. 11. Examples of hits in WPI with Manual Code X15-B05 and not IPC F03D7 (with permission of STN International).

```
L6    ANSWER 3 OF 590     INPADOCDB COPYRIGHT 2010 EPO/FIZ KA on STN
TI    TETHERED GLIDER SYSTEM FOR POWER GENERATION.

IPC   CODE             VERSION  POS INV LEVEL    CC ASSIGNMENT DATE     STAT
----  ---------------- -------- --- --- -------- -- ---------- -------- ----
IPCI  F03D-0005-06     (200601)  F   I  Advanced US Human      20100916  O
      B64C-0031-06     (200601)  L   I  Advanced US Human      20100916  O
      F03D-0007-00     (200601)  L   I  Advanced US Human      20100916  O <--
      F03D-0009-00     (200601)  L   I  Advanced US Human      20100916  O
      H02P-0009-04     (200601)  L   I  Advanced US Human      20100916  O
      F03D-0005-00     (2010)    F   I  Core*    RC Machine    20100916  O
      B64C-0031-00     (2010)    L   I  Core*    RC Machine    20100916  O
      F03D-0007-00     (2010)    L   I  Core*    RC Machine    20100916  O <--
      F03D-0009-00     (2010)    L   I  Core*    RC Machine    20100916  O
      H02P-0009-04     (2010)    L   I  Core*    RC Machine    20100916  O
NCL   NCLM 290/044.000
      NCLS 290/055.000
```

Fig. 12. Classification origin display in STN's InpadocDB (with permission of STN International).

added in abundance to documents for every detail disclosed in the application. Searching for View point "Types of wind motors" 3H078/AA alone gets you over 4800 hits in the IPDL.

The comparison with the remaining two classification schemes is more complex because there is no one-to-one relationship between them and there is to my knowledge no single database offering them all. Looking at Derwent's X15-B05 and analysing it for IPC, ECLA and FI we see that indeed F03D7 and its sub-groups come out at top in all of these but that there are also a number of partially divergent classes like H02J (Circuit arrangement) and H02P (Control or regulation of electric motors) among the top 20. When looking at those hits that have only X15-B05 and not F03D7 we find that these - according to their Derwent title - actually do have a control aspect which is not mirrored in their IPC classification (Fig. 11).

For the USPC results there are two different comparisons: on the one hand you can search for the US classes identified as outlined above but on the other you may also compare the original USPTO-applied IPC to the results that contain also non-US family members (and accordingly have been classified by other offices as well). Firstly searching the most frequent subclasses 290, 415 and 416 as identified from our previous IPC result sets will quickly show you that there is no single USPC for wind driven generators. Using these classes you will always end up with an assortment of generators with different driving fluids (wind, gases, water). You will not be able to reproduce a classification only search with a similarly clean result as the one you can reach with either of the other classifications - you will have to resort to text searching as well.

The only database provider to have fully integrated WIPO's Standard ST.8 for IPC classification [14] notation is STN. There you not only can search IPC symbols but also restrict your searches to other criteria like "Generating office", "Source of classification" or whether it is original or reclassified data. When analysing the results of those hits that were originally classified into F03D7/00 by the USPTO (a mere 600 of over 10,000 documents) you'll find a number of them whose inclusion might be debatable (see Fig. 12 which the EPO classified into F03D5 "Other wind motors" and R05B240/921 "kept aloft due to aerodynamic effects"). This may or may not point to the known practise of USPTO examiners accepting computer generated concordance suggestions for IPC classes.

Of course to be able to take more than a cursory glance at results obtained by searching with different classification systems further in-depth comparison of result sets with thorough analysis of single documents would be necessary - which the present article cannot provide.

4. Conclusion

We have now seen some of the differences of the existing schemes to classify patent documents. They differ in a number of ways: some cover publications from numerous sources worldwide, while others are content to only cover their own national office's documents. They all try by different means to stay abreast of technological developments and nearly all of them apply the changes and additions in the classification also to the their back-file. Some are centrally administered and controlled, others leave it to the examiner/classifier to do the right thing in their own interest. Now what is the lesson for patent searchers?

There certainly is a downside to those differing systems: you have to know and keep up with several systems, how they are structured and have an idea how they are implemented by the respective database producers. You are at the provider's mercy when it comes to implementing and updating their stores. You have to collect, update and browse documentation that is sometimes hard to come by and maintain current. Some of the systems (especially the "free" sources) only cover a limited amount of documents and at most support two or three of the schemes outlined. A comprehensive search is only possible if you have access to several databases and the time to prepare your classification search is even more drawn out if you have to resort to the different free sources of national offices. In one word: there will be additional work preparing your search. All of this makes even seasoned searchers shy away from using several of the systems and keep to the one(s) they feel most comfortable with.

Notwithstanding all of these difficulties I think it worth your while for certain types of searches (especially opposition and freedom-to-operate) to try and include as many classification systems as feasible within your allotted time: it is the only way to cross language barriers (especially the Japanese one); it identifies additional subject matter that otherwise might not be included as searchable text in the database you use (e.g. German utility models); it might even identify non-textual matter (as in the FI example above); it will give you a view from different angles as examiners from different offices might very well take a different approach regarding the subject matter (also because of national differences in allowable subject matter). But most of all you may put most trust in those classification systems where classifier and user are the same. There the incentive to do an exceptionally good job is greatest, not because it is a duty, but because the examiner will see an immediate return for the effort which has to be put into finding the right slot(s) for the document at hand. And it gives you

a glimpse of the examiner's desk and also into her or his reasoning. I think it will be quite obvious that people put more effort into classifying patents if the quality of their and their peer's work will in the long run hinge on the initial effort. This in turn gives the searcher unique opportunities to enhance their searches with insights additional to IPC and keyword searching. In other words: it takes all kinds to make a world.

Acknowlegdement

This article has been developed from the author's presentation at the European Patent Office Patent Information Conference in Lausannne, Switzerland in October 2010.

I am very much indebted to the following people for generously sharing their knowledge in preparing this article: Dominic DeMarco (DemarcoIP), Bernhard Geyer (German Patent and Trademark Office), Tatsuya Tada (Japanese Patent Office), Victor Veefkind (European Patent Office) and Heiko Wongel (European Patent Office). And my colleague Chris Richardson for proofreading and general good advice.

Last but not least I'd like to stress that the opinions expressed in the present article are strictly my own and not necessarily those of my employer.

References

[1] http://www.wipo.int/classifications/ipc/en/.
[2] http://www.stn%96%3binternational.de/fileadmin/be_user/STN/pdf/presentations/0910ClassOnClass.pdf.
[3] http://www.espacenet.com/getstarted/classification_search_en.pdf.
[4] http://www.wipo.int/treaties/en/ShowResults.jsp?lang=en&;treaty_id=11.
[5] http://worldwide.espacenet.com/help?topic=classesqh&locale=en_EP&method=handleHelpTopic.
[6] http://depatisnet.dpma.de/ipc/language.do?lang=EN.
[7] http://www.ipdl.inpit.go.jp/homepg_e.ipdl.
[8] http://science.thomsonreuters.com/support/patents/dwpiref/reftools/classification/#eng.
[9] Falasco L. United states patent classification: system organization. World Pat Inform 2002;24(2):111–7.
[10] http://www.uspto.gov/web/patents/classification.
[11] http://www.intellogist.com/wiki/US_Patent_Classification_System.
[12] http://works.bepress.com/cgi/viewcontent.cgi?article=1059&context=rkatznelson.
[13] Dickson CR. Creating and protecting Intellectual Property rights for nanotechnology. BNA World Intellect Property Rep 2010;12:31.
[14] http://www.wipo.int/wipostad/en/standards/en/st8/1-0.
[15] Cullen SE. A practical approach to the reformed IPC. World Pat Inform 2009; 31(3):193–8.
[16] Scheu M, Veefkind V, Verbandt Y, Galan EM, Absalom R, Forster W. Mapping nanotechnology patents: the EPO approach. World Pat Inform 2006;28(3): 204–11.
[17] Makarov M. The eighth edition of the IPC. World Pat Inform 2006;28(2): 122–6.
[18] Wongel H. The reform of the IPC-consequences for the users. World Pat Inform 2005;27(3):227–31.
[19] Schellner I. Japanese File Index classification and F-terms. World Pat Inform 2002;24(3):197–201.
[20] Hull D, Ait-Mokhtar S, Chuat M, Eisele A, Gaussier E, Grefenstette G, et al. Language technologies and patent search and classification. World Pat Inform 2001;23(3):265–8.
[21] Adams S. Comparing the IPC and the US classification systems for the patent searcher. World Pat Inform 2001;23(1):15–21.
[22] Goebel M. Mirror, mirror on the wall, squishy and soggy, 2 Nanos Tall: Strategies, methods and tools for searching homogeneous catalysts – An EPO perspective (Part 1. Introduction and patents). World Pat Inform 2010;32(1): 39–52.
[23] Adams S. English-language support tools for the use of Japanese F-term patent subject searching on line. World Pat Inform 2008;30(1):5–20.
[24] Goebel M, Hintermaier F. Searcher's little helper – ICO index terms. World Pat Inform 2011;33(3):260–8.