E

WIPO

WORLD INTELLECTUAL PROPERTY ORGANIZATION

GENEVA

ADVANCED SEMINAR
ON THE INTERNATIONAL PATENT CLASSIFICATION
(IPC)

Newport, United Kingdom, December 7 to 11, 1998

DEVELOPMENT OF THE IPC AS A SEARCH TOOL

Discussion paper presented by Mr. Anders Bruun, Senior Examiner,
Swedish Patent and Registration Office, Stockholm

Background

Development of what was to become the IPC started in the middle of the 1950s. The first edition of the IPC was introduced in 1970. A long time has passed since then, and the requirements of an international classification system have changed a lot. However, apart from the much debated and not so successful introduction of indexing, the basic structure and philosophy of the IPC has remained unchanged through the years. There are four main background factors that have to be considered when future development of the IPC is discussed:

1. The growth of the documentation
2. Electronic storage of documents
3. Alternative search tools
4. Technical development in the fields searched

When the IPC was introduced, the term "PCT minimum documentation" did not exist. However, the corresponding documentation amounted in 1955 to around 3.5 million documents and in 1970 to around 7 million documents. IPC1 contained around 47000 groups. This would have resulted in an average group size of around 150 documents in 1970. The present size of the PCT minimum documentation is probably around 25 million documents, and it grows by several hundred thousand documents per year. IPC6 contains around 66000 groups. The average size of an IPC group is now almost 400 documents. It is obvious that search efficiency, when using the IPC as the only search tool, has decreased considerably. In reality, the picture is of course more complex than so, but major patent offices, such as the European and the Japanese, have long ago introduced elaborate additional classification tools in order to compensate for this development. Several IPC groups now contain more than ten thousand documents, making the IPC completely unpractical as only search tool.

The IPC was designed to be used with paper search files, as a sorting tool for creating conveniently sized, essentially static piles of documents that could be gone through manually, one by one. Fundamental features of the IPC are influenced by this, for example the general philosophy that one invention should be indicated by only one classification symbol. The possibility to make cross searches, using a combination of several search terms, is of little interest in paper search files, since there is no way of easily identifying and accessing a document. Several other features of the IPC, such as precedence references and last place rules, are intended to limit the number of classification symbols allotted to documents, thereby minimising the growth of the search files. A computer, on the other hand, can automatically identify documents using complex search questions and can quickly display a selected document. The number of classifications assigned to a document that is electronically stored is irrelevant. When IPC8 enters into force, presumably in the year 2005, most patent offices will have started getting rid of the paper search files.

When the IPC was introduced, classification was basically the only search tool available. Nowadays, a number of more or less advanced computerised search tools are available. These range from simple word searches in abstracts or full texts to complex deep indexing schemes or chemical structure search engines. These alternative search tools are provided both by commercial database companies and by patent offices. In some technical fields alternative search tools have more or less replaced classification as the main tool, but in most fields they are used as a complement to the classification. However, word searches are still completely inadequate in many fields, for example because no standardised terminology exists or because

the inventions can be described in an almost infinite number of alternative ways, such as is the case when they relate to shapes and spatial relationships of components. To some extent the existence of commercially available databases have been taken into account when revising the IPC. Since some years there is an instruction to be restrictive in revising technical fields where readily available commercial databases can meet the search needs.

Thirty or forty years is a long time in the development of technology. Terminology and subdivisions have become outdated in some of the original parts of the IPC. The IPC has seldom been developed at the same time as an emerging technology, but has more often reacted with a delay of five or ten years. During the years, a considerable experience in the construction and presentation of IPC schemes have been accumulated. This is reflected in the parts of the IPC that have been revised, but some old areas that have remained untouched by revision are full of non-standard wording and solutions. Revision work has mainly been directed towards subdivision of overpopulated areas, rather than towards updating old parts of the scheme.

The future

Given this background, which direction should new developments in the IPC take? Here is a list of ideas for future developments:

The present philosophy of "one invention – one classification" is a thing of the past. It served its purpose well in paper search files, but does not enable the full use of electronic search and storage of documents. A future IPC must enable giving information by the combination of two of more classification symbols. This not only gives a lot of search possibilities using Boolean expressions. It is also efficient in terms of revision work, since a relatively small number of "overlapping" entries gives a higher number of search possibilities than an extensive traditional subdivision.

Precedence references and last place rules
A start is to abolish precedence references and last place rules, and replace them with a standardised way of indicating that overlapping places exist. Precedence references and last place rules are features that were implemented because of the concerns about the growing paper documentation. They inherently lead to reduced search efficiency, since they direct documents away from places where their inventions are expressed. Overlaps actually give new search possibilities. Two mutually exclusive bits of information (A and B) only give two selection alternatives. Two overlapping bits of information gives five selection alternatives: A; B; A and B; A not B; B not A. A traditional classification scheme giving these five search possibilities would be complicated and difficult to use.

Indexing and multiple classification
The next logical step, taking the same idea further, is indexing. The present use of indexing in the IPC has admittedly to a large extent been a failure. The hybrid systems were controversial from the beginning. Indexing is basically a computer search tool, and it was probably introduced too early, before users were ready to invest work in future computer searches. Indexing is not obligatory, so a searcher can never trust that that all relevant documents have been indexed. The construction and use of indexing schemes has not been standardised and internally consistent in the same way as the classification schemes, leading to confusion. The

presentation of indexing codes is very similar to classification codes, leading to further confusion. Few of the commercial database providers give complete and correct indexing information, thus denying the computer searchers the use of a system which is mainly intended for them.

Nevertheless, over the years some experience of indexing schemes has been gained. Now is the time to use that experience in order to create better and more acceptable indexing schemes for a future IPC. There are several different ideas on indexing, and it would probably be best not to use more than one or two of these.

Global use of classification entries for indexing?

This is an approach that initially seems rather appealing – it would immediately give a large number of indexing entries that are familiar and well proven. There are several disadvantages, though. One is that since many classification entries would be used to indicate non-invention indexing information, these entries would be overloaded with documents, making them inefficient for searches. A solution to this problem would have to be found. The present way of indicating indexing use in double purpose parts of the IPC is to replace the oblique stroke by a colon. This has proved insufficient. The EPO's ICO indexing schemes use another approach, where a different series of first letters of the subclass symbol to indicate indexing use, for example R23N instead of F23N. This is an elegant solution which could solve the problem. However, one very important problem would remain, and that is that a schemes which are drawn up for classification probably do not work so well for indexing. They give a level of detail which might not be desirable and they often contain classification rules which are not suitable for indexing.

Special entries for global indexing?

It has been put forward that there might be some entries that could be useful for indicating additional information in the entire IPC, or at least large parts of it. During a recent IPC revision project it was put forward that the provision of lighting in different apparatus might be such a case. A brain-storm for further ideas brought forward the examples of "application of shape-memory metals", "use of compostable or biodegradable materials" and "special adaptation for recycling or re-manufacturing". Any terms for global indexing would have to be carefully selected and some system for avoiding redundant information would have to be implemented.

Deep indexing versus light indexing?

Deep indexing methods involve the identification of each different element of information that appears in a long list of entries. This is a time-consuming effort, which results in long lists of indexing terms, which might or might not be of interest for future searches. Although certainly useful in many fields, deep indexing is a method which is alien to the concept of classification, in which identification of relevant invention information is a fundamental part. What can be achieved by deep indexing is similar to what can be achieved be full-text search - full-text search can indeed be said to be the ultimate form of deep indexing. Light indexing, giving a limited number of relevant indexing entries that are applicable to a limited field, forms a more natural extension of classification. One of the reservations against indexing is that it leads to indication of "non-invention" information. The risk for this is smaller in a system with a few selected indexing terms.

Multiple classification?

The ideal system is probably one which is not so much "classification and indexing", but rather "multidimensional classification". The concept of multiple classification has been introduced in several places in the IPC, but only as a special solution for special reasons. The introduction of multiple classification as a standard solution when drawing up new subdivisions should be investigated. This would be a way to avoid several, if not all, of the precedence references in future revisions.

It is extremely important to find a system which can be accepted by a majority of offices and made obligatory. In any case, if the concept of "multidimensional classification" is accepted, the borderline between classification and indexing would disappear and everything that is used to record invention information could be called classification. Indexing schemes that are intended for recording invention information could simply be called classification schemes. This would avoid modification of the Strasbourg Agreement in order to make indexing obligatory.

A revision policy for the computer era

As stated earlier, use of classification as the only search tool is no longer relevant in many technical fields. The IPC should therefore be adapted to coexist with the other search tools. In technical fields where classification is the main search tool, it must still on its own make it possible to break down the documentation into manageable parts. In other fields, the IPC will serve as a first subdivision tool combined with text searches or other alternative tools. In any case, what will always be needed is a high quality up-to-date basic classification framework providing clear subdivisions which can be used to accurately delimit search areas. Clear, unambiguous subdivisions will be even more important in the future, since the documentation growth will not stop and the importance of accurately limiting searches will be bigger. Furthermore, in a computer environment it is often, at least with present technology, more time-consuming and less user-friendly to quickly manually scan documents than in a paper search file.

The degree of subdivision can be dependent on the alternative search tools available, as long as the complete tool kit of the searcher makes it possible to do efficient high-quality searches. This will put a different emphasis on the quantitative criteria used for adopting new revision projects. The current file growth criteria are very strict, probably too strict to allow efficient subdivision in fields where classification is the only tool for limiting a search. On the other hand, in fields where classification is a secondary tool or used in connection with text searches, revision should perhaps mainly be directed towards clarification and technical updating of the existing scheme.

As stated above, an important requirement of a classification scheme is that it is clear and unambiguous, and gives clearly defined entries for search. The classification scheme should be built up in an internally consistent way, so that users can find the same familiar solutions applied in different areas of the scheme. The present scheme has been developed over a period of more than forty years. In parts, it is old-fashioned in wording and subdivisions and inconsistent in wording and practice, for example regarding placement and use of references. It would be desirable to go through the entire IPC and update and harmonise all parts of it. Problem areas should be identified and given special attention, perhaps leading to complete revision of some fields.

The presentation of the IPC

If the computer versions of the IPC, on Internet, CD-ROM or DVD, are made the main versions, the layout and presentation and perhaps even some of the basic features should be changed considerably in order to give quick and easy access to all the relevant information. The IPC:CLASS CD-ROM has indicated a couple of problems with the presentation of notes, guide headings and subsections in the hierarchical mode, which have to be addressed and which perhaps point to the unsuitability of these features for computerisation. A computer version would also give the possibility to hide non-essential information, such as informative references, and to link additional information, such as statistical data or example documents, to the different entries. If the revision period is shortened it will be extremely important to enable automatic linking between different versions of the IPC.

Conclusion

The present position of the IPC is a dilemma. Its efficiency has dropped to a level where several users have found it necessary to introduce additional classification tools. Some big information providers, even among the biggest patent offices of the world, have not committed themselves to use the IPC in a serious way. Still, the IPC is a major international standard – it is used by almost 100 industrial property offices around the world and relied on by thousands of searchers. It is also the only truly international classification system. This is easy to forget in a world where patent activities are more and more concentrated to a few big organisations.

A minimum ambition is that the IPC must be good enough to be used as a basis for the internal classification tools developed by the big patent offices. If it is not, the result will be a fragmentation of classification into several different systems again, exactly the situation that the IPC was originally intended to stop. We have already seen the start of this development.

This would argue for quality of the scheme, rather than quantity of documents, as the main criteria for revision. Up to this point, the IPC has been developed step by step. A new computerised IPC and the new computerised environment at patent offices will require a re-engineering of the system in order to enable efficient use of digital search files and computer tools. This will require considerable resources, but it will be necessary if we want to maintain the IPC as a modern search tool.

However, the IPC is also used for documentation which does not form part of the PCT minimum documentation, and is thus not interesting for the big offices. This documentation must still be accessible for efficient search. Much of it will be published in "small" languages, so it will not be well served by text searches or text analysis tools. As a consequence, the quantitative aspects can not be disregarded in future development of the IPC, especially not in technical fields where alternative search tools are not available or suitable.

The future of the IPC as an international search tool will probably be decided during the next revision period. If it is not rejuvenated, it will be marginalised.

[End of document]