E

WIPO

IPC/CE/30/5 Suppl.2
ORIGINAL: English
DATE: February 15, 2001

WORLD INTELLECTUAL PROPERTY ORGANIZATION
GENEVA

SPECIAL UNION FOR THE INTERNATIONAL PATENT CLASSIFICATION (IPC UNION)

COMMITTEE OF EXPERTS

Thirtieth Session
Geneva, February 19 to 23, 2001

COMMENTS ON DOCUMENT IPC/CE/30/5

Document prepared by the International Bureau

Annexes 1 to 4 to this document contain comments submitted by Japan, Sweden, the United Kingdom and the European Patent Office (EPO), concerning the most appropriate contents of the core level of the IPC.

[Annexes follow]

ANNEX 1/ANNEXE 1

Japan Patent Office

3-4-3 Kasumigaseki
Chiyoda-ku
Tokyo
Japan

February 15, 2001

Size of Core Level

I. Introduction

Discussions have been held on the core level size based on the case study by EPO.

II. Basic Concept

a) In IPC/REF/4, the limited size of structure and the stability are required in Core Level.
b) The number of entries in Core Level should be limited to the extent that small-scale patent offices adopting Core Level would not have any inconvenience in searching their domestic documents.
c) Also, it pointed out that JPO and EPO kept on studying automatic assignment of classification and that structural size of Core level should be set up in light of future automatic classification assignment.

III. Size of Core Level

Under the basic concept, the proper number of entries in Core Level should be decided based on the relation between the number of patent applications of small-scale Patent Offices (adopting Core Level) and that of large-scale Patent Offices (adopting Advanced Level).
In other words, it should be similar when comparing the ratios of a) and b).

a) "the number of applications of large-scale Offices" to "the number of entries in Advanced level"
b) "the number of application of small-scale Offices" to "the number of entries in Core level"

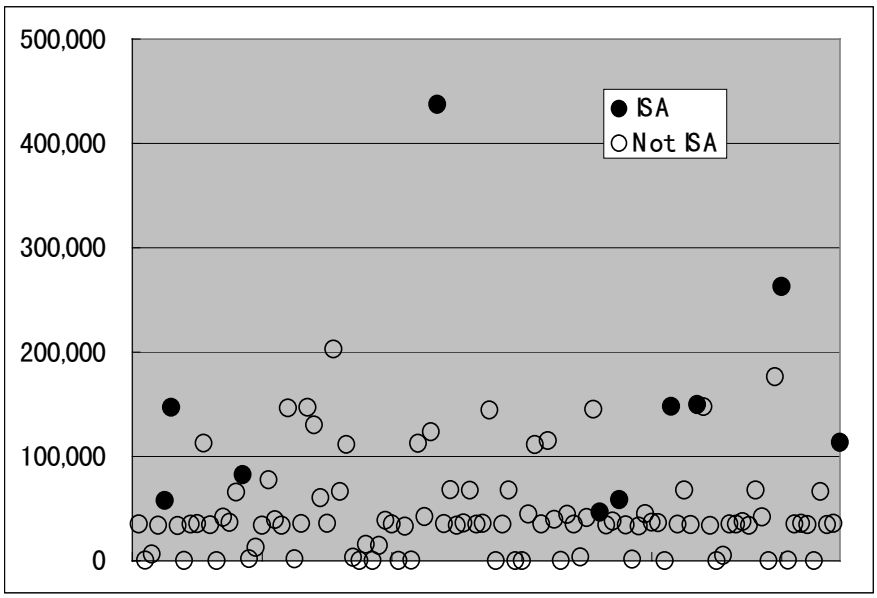
IV. Study conducted by JPO

a) As shown in a graph based on WIPO statistics, distribution of each country's number of applications is divided into two groups of 30,000-document level and 140,000- document level with a boundary of 100,000-document level.
b) Therefore, we can define 140,000-level group to be Offices that would adopt Advanced Level, as most ISA belong to 140,000-level group.
c) Similarly, we can define 30,000-level group to be Offices that would adopt Core Level.
d) Assuming that the number of entries in Advanced Level be 100% and Core Level x%, the following formula could be obtained.

$$\frac{\text{140,000 documents (Number of Applications of Large-scale Office)}}{\text{100 \% (Number of entries in Advanced level)}} = \frac{\text{30,000 documents (Number of Applications of Small-scale Office)}}{\text{x \% (Number of entries Core level)}}$$

$\therefore$ x = 20 %

e) Therefore, the JPO would like to propose the structural size (the number of entries) of Core Level should be around 20% of all the entries of IPC.



[Annex 2 follows/
L'annexe 2 suit]

ANNEX 2/ANNEXE 2

Swedish Patent and Registration Office

IPC Reform, Task 14 **February 2nd, 2001**

DETERMINATION OF THE MOST APPROPRIATE CONTENTS OF THE CORE LEVEL OF THE REFORMED IPC

COMMENTS

At IPC/REF/4, the EPO volunteered to submit the results of the further study on the split-up of the IPC in a core and an advanced level. IPC/REF/4/4, paragraph 47, invited comments on the study.

Background: Search needs

The main purpose of the IPC is to serve as a tool for searching patent applications. This applies both for the advanced level and the core level. The core level must have search credibility. We do want the Master Classification Database to be populated with core level documents - and not only with new documents, but with all the world's documents. If no office sees any search use of the core level, then nobody will reclassify backfiles into it. No office can afford reclassification for its own sake. If the core level does not become good enough, offices will prefer the present chaos of different old IPC versions and national classifications.

Criteria for selection of the core level

All main groups must be part of the core level. If not, many documents would become homeless, or we would need to declare that a subclass symbol on its own could be a complete classification symbol. This would destroy the symmetry of the system.

We see no reason for deleting passive small groups from the core level by applying a minimum file size figure. One of the aims of the core level is that it should be stable, in order to minimise reclassification. The passive groups are the most stable of them all, and we can not see any reason for investing money in moving them. Furthermore, deletion of passive groups would force documents from passive groups into more active groups, leading to an unnecessary reduction of search efficiency.

We see no reason for including or excluding certain dot levels in the IPC in its entirety or in subclasses in their entirety. Different criteria should apply in different parts, since the documentation growth has been extremely different in different fields.

If we accept that single groups in an array of groups of the same indentation can be moved, then it means that a uniform first or last place rule is impossible. Such position rules mean that only entire dot levels can be moved - if not, there can be no compatibility between the two levels.

We would propose this approach:
The size of the core level should be decided in order to serve the search needs. We prefer an approach that focuses on documentation size, rather than focusing on whether, for example, all one-dot groups should go to the core level or a certain percentage of present IPC should be excluded.

Let us look at each main group. If the PCT-minimum documentation size is bigger than a fixed maximum figure, then all the one-dot groups go into the core level. Then look at each one-dot group. If the documentation is bigger than the maximum figure, then all two-dot groups indented under the one-dot group must go into the core level. This procedure is then repeated at each dot level. This is a simple and logic method, and since it deals with entire dot levels it is applicable also in the last place rule areas. It can also more or less automatically be used in the future in order to decide when advanced level groups need to be moved to the core level.

What should this maximum figure be?
The maximum figure must be chosen so that it results in realistic file sizes in the collections that can only be searched in the core level. When we revise the IPC, an average file size of 100-150 PCT-minimum documents is considered reasonable. It would be reasonable to apply similar quantitative criteria for file size in the collections that must be searched in the core level. If we can find out the proportion between the size of the PCT-minimum documentation and the size of the biggest potential core level collections, then we can just multiply the figure of 100-150 with that figure and get a maximum figure for the corresponding core level groups.

How big are the collections that have to be searched using the core level? We guess that no office will be searching a collection of more than one million unique documents. The PCT-minimum documentation contains between 20 and 25 million documents, that is between 20 and 25 times the hypothetical collection of one million documents. This means that a PCT-minimum number of documents in a group of 2000-4000 would correspond to 100-150 core level documents.

This appears to be a realistic figure. Consultation of examiners at the Swedish office and investigation of actual search files has led us to the conclusion that a maximum PCT-minimum figure of around 2000-2500 would allow a core level with realistic, but certainly not impressive, searchability for our non-PCT-minimum collections. A maximum figure of 5000 documents would make the search sets too big for reasonably efficient search. It has to be remembered that in most cases the core level collections will have to be searched in addition to the advanced level collections. It also has to be assumed that little in the way of full-text searching and other tools will be available for the non-PCT-minimum files, since many of them will be old and/or published in small languages.

As background information, the Swedish office is obliged to search the PCT-minimum documentation plus the documentation of Sweden, Norway, Denmark and Finland. The total number of these non-PCT-minimum documents is around 800,000. Of them, we have family information about corresponding PCT-minimum documents for around 200,000, which means advanced level classification data will be available. The remaining 600,000 documents will have to be searched, in one way or another. At the moment they are searched using a mixture

of different IPC versions and our old national classification system. We would only be interested in reclassifying backfiles into the core level if it becomes feasible for search. If it is not, then we would prefer our present "chaos", and then there would be a gap in the Master Classification Database.

The EPO investigations

Two alternative models have been investigated by the EPO and results provided on CD. We thank the EPO for this impressive work - it certainly has made it easier to envisage the consequences of what we are now trying to do.

We think the smaller of the two alternatives presented by the EPO is clearly too restricted for realistic searching, even in small collections. The bigger alternative would probably give realistic searches, but we are worried about the consequences of the model in last place rule areas, and we think that in some areas keeping all one-dot groups would lead to a too high number of core level groups. We have no doubt that the size of the core level can be reduced considerably in relation to the present IPC. However, we think that the reduction of the core level to 30% of the present IPC is too radical and would lead to unacceptable search efficiency.

As stated above, we would prefer a model were only the main groups are guaranteed a place in the core level, but where the maximum number of documents per group is somewhat smaller and only entire dot levels are moved. In our opinion, the last criterion is of course necessary in order to make the two-level IPC compatible with future general first or last place rules.

However, this problem also has to be seen in connection with a future emerging expanded advanced level, which will presumably contain new groups, not present in the existing IPC, at all dot levels. As stated before, we really do not believe that the two-level internally compatible IPC is consistent with first or last place rules, but that is, of course, another question.

Anders Bruun

[Annex 3 follows/
L’annexe 3 suit]

ANNEX 3/ANNEXE 3

GB Comments 08/02/01

Establishment of the restricted core level

We have read the EP paper of 23/12/00 on this subject with great interest and would like to thank them for the work done on behalf of the IPC community.

The precise nature of the restriction of the core level is essentially a compromise between the needs of offices having smaller, particularly non-PCT minimum collections and the needs of the Trilateral offices in being able to change the core level quickly. Even if it means a little more work now, we should make sure that the decision upon how much the core level is restricted can be fully justified in terms of retaining a sufficient level of searchability.

With that in mind, we have the following specific comments on the above-mentioned EP paper, related to the numbered paragraphs of that document:-

6. Although the restriction to 30% was mentioned at the 4th meeting of the Reform Group, as is acknowledged in the paper, this was very much a tentative figure. We think that restriction to main group level would be far too drastic in many subject matter areas to retain viability as a search tool.

9-11. We are uneasy about using averaged approximations as a basis for the restriction.

13. We think the idea of only using percentages rather than file size is doing things the wrong way round. We think that it will produce similar results in some areas of the IPC to that produced by the algorithm, but in others there may be distortions. We think that with either percentage or algorithm the last place rule will mean that main groups using the last place rule will have to be treated separately.

15. We favour the idea of a combination method roughly as outlined in para. 15, so that the algorithm approach, which can be amply justified, is used wherever possible. We are just a little concerned for subclasses where the last place rule is only used in one or two isolated places, e.g. F16H. For such areas we would propose to use the algorithm approach but not in respect of those one or two main groups where last place rule is used.

Conclusion

We think that the reasoning behind using the algorithm approach is largely a good one and should be persisted with. The subclasses using last place rule extensively are indeed invalid for use with the algorithm as compatibility between core and advanced level would not be maintained, a basic principle of the core/advanced philosophy being lost.

We have come a long way since the first ideas of the restriction being to IPC7 without indexing. On the other hand we think a restriction to main groups, one dot groups or the like, was certainly too crude and not logically justifiable. The advantage of the algorithm approach is to allow the restriction to be tailored to the activity of the subject matter area and we still think that basic principle is a good one.

Jim Calvert
UK Patent Office

[Annex 4 follows/
L'annexe 4 suit]

ANNEX 4/ANNEXE 4

Europäisches Patentamt GD1 | **European Patent Office** DG1 | **Office européen des brevets** DG1

Principal Directorate Documentation

HP/01.030/hp

SIZE OF THE CORE LEVEL

Determination and use of it

I. Introduction

The EPO used an algorithm to test with different parameters of groups size the impact on the size of the core level. The determination of the size of the core level can only be done in a correct way after considering the use of the core level and the criteria linked to it. It is felt that the use of the core level is still not always understood in an appropriate way.

II. Use of the core level

The IPC is developed for classifying and retrieving the classified information. As soon as a two-tiered IPC is available both activities can be applied for core and advanced level. However are these activities to be applied in a systematic way for one of these levels by each Office?

In the past it was already mentioned that the advanced level can be used for retrieval by all Offices interested in it. Offices can apply the core level for classification and the advanced level for retrieval but it is also possible to use the core level for classification of a part of the documents published and the advanced level for the rest of the documents published. In some countries a specific part of technology is highly developed and the advanced level is more appropriate for this technology. The mixed use of core and advanced level for classification is not a problem as with the two-tiered IPC it becomes necessary to indicate always per classification symbol the level used for it. The same level information is also taken over in the MCF which also allows the retrieval per case and per selected level.

For the sake of completeness it is mentioned that the database maintenance must assure that in case of advanced level classification the core level is generated automatically to make the search files complete for the core level.

III. Obligation linked to the core level

With the introduction of the two-tiered IPC Offices are obliged to use one of these levels as mentioned above. It means that smaller Offices must be able to apply the core level at least in the totality for all technical fields. This condition of applying at least the core level is important to guarantee the completeness for search purposes.

IV. Criteria for determining the size of the core level

In the past it was mentioned that the stability of the core level and the size of the national or regional collections are the basis for determining the size of the core level. The obligation to apply at minimum the core level is a further criterium which should fully be taken into account for determining the size of the core level. The last criterium seems to be the decisive one as if not fulfilled the core level is incomplete.

[End of Annex 4 and of document/
Fin de l'annexe 4 et du document]