**U.S. Patent (a.k.a., Patent Grant or Grant)**

**Cooperative Patent Classification Master Classification File (Patent-CPC-MCF)**

# Last updated August 2018

The U.S. Patent-CPC-MCF contains CPC classification information for all Utility patents granted/issued by the U.S. Patent and Trademark Office (USPTO). It is available monthly. It is available for download for no charge from USPTO website: <https://bulkdata.uspto.gov/data/patent/classification/cpc/>

It can also be searched for no charge on the USPTO website at: <http://www.uspto.gov/web/patents/classification/>

For more information about available USPTO electronic information products (EIP): <http://www.uspto.gov/learning-and-resources/electronic-bulk-data-products>

Each month there is one folder of XML data and one folder of text data created with the cut-off date of the last day of the month in the folder name. The records in XML format and text format are exactly the same other than the XML tagging.

**US\_Grant\_CPC\_MCF\_XML\_yyyy-mm-dd.zip (currently includes 1 folder of over 200 xml files)**

**US\_Grant\_CPC\_MCF\_Text\_yyyy-mm-dd.zip (currently includes 1 folder of over 200 text files)**

# There are currently over 200 xml files. The number of files keep growing as more U.S. patents are granted/issued. For early data, each file contains no more than 50,000 patent numbers excluding withdrawn patents.

US\_Grant\_CPC\_MCF\_00000001.xml contains patent numbers between 1 and 49999.

US\_Grant\_CPC\_MCF\_00050000.xml contains patent numbers between 50000 and 99999.

US\_Grant\_CPC\_MCF\_00100000.xml contains patent numbers between 100000 and 149999.

US\_Grant\_CPC\_MCF\_00150000.xml contains patent numbers between 150000 and 199999.

.

.

.

US\_Grant\_CPC\_MCF\_09950000.xml contains patent numbers between 9950000 and 9999999.

US\_Grant\_CPC\_MCF\_10000000.xml contains patent numbers between 1000000 and 10499999.

The records in each file are sorted first by patent number ascending and kind code. For the same patent number and kind code, the records are sorted by symbol position (F - First or L - Later); Among Later records, they were sorted by classification symbol and combination sets. The Later records not in combination set are sorted by classification value (I - Invention or A - Additional), then by classification symbol order ascending. In a combination set, the records sorted by group number and rank number.

## XML Schema and Data Sample

The schema files .xsd are included in **US\_Grant\_CPC\_MCF\_XML\_yyyy-mm-dd.zip**

|  |  |
| --- | --- |
|  | Main schema for CPC MCF |
|  | Common namespace components from WIPO ST.96 |
|  | Patent namespace components from WIPO ST.96 |

The root element of the XML schema for Grant\_XML files is uspat:CPCMasterClassificationFile with multiple uspat:CPCMasterClassificationRecord. Each record is for one patent number with one kind code. In the root element, two attributes @publicationStartNumber and @publicationEndNumber were populated with the actual start and end patent number in each file.

There is a possibility of one patent number with more than one kind code. The kind code value can be “B1”, “B2” or “A”. Each kind code will have its own record for the grant number. This situation is due to the reexamination certificates.

Here is an example of CPCMasterClassificationRecord of patent number 9000011 with kind code B2.

NOTE: Due to ongoing reclassification effort, the CPC symbols in the examples may not reflect the current CPC data.

<uspat:CPCMasterClassificationRecord>

 <pat:ApplicationIdentification>

 <com:IPOfficeCode>US</com:IPOfficeCode>

 <com:ApplicationNumber>

 <com:ApplicationNumberText>13445338</com:ApplicationNumberText>

 </com:ApplicationNumber>

 </pat:ApplicationIdentification>

 <pat:PatentGrantIdentification>

 <com:IPOfficeCode>US</com:IPOfficeCode>

 <pat:PatentNumber>9000011</pat:PatentNumber>

 <com:PatentDocumentKindCode>B2</com:PatentDocumentKindCode>

 <pat:GrantDate>2015-04-07</pat:GrantDate>

 </pat:PatentGrantIdentification>

 <pat:CPCClassificationBag>

 <pat:MainCPC>

 <pat:CPCClassification>

 <pat:ClassificationVersionDate>2013-01-01</pat:ClassificationVersionDate>

 <pat:CPCSection>G</pat:CPCSection>

 <pat:Class>01</pat:Class>

 <pat:Subclass>N</pat:Subclass>

 <pat:MainGroup>33</pat:MainGroup>

 <pat:Subgroup>5094</pat:Subgroup>

 <com:SymbolPositionCode>F</com:SymbolPositionCode>

 <pat:CPCClassificationValueCode>I</pat:CPCClassificationValueCode>

 </pat:CPCClassification>

 </pat:MainCPC>

 <pat:FurtherCPC>

 <pat:CPCClassification>

 <pat:ClassificationVersionDate>2013-01-01</pat:ClassificationVersionDate>

 <pat:CPCSection>A</pat:CPCSection>

 <pat:Class>61</pat:Class>

 <pat:Subclass>K</pat:Subclass>

 <pat:MainGroup>31</pat:MainGroup>

 <pat:Subgroup>45</pat:Subgroup>

 <com:SymbolPositionCode>L</com:SymbolPositionCode>

 <pat:CPCClassificationValueCode>I</pat:CPCClassificationValueCode>

 </pat:CPCClassification>

 <pat:CPCClassification>

 <pat:ClassificationVersionDate>2013-01-01</pat:ClassificationVersionDate>

 <pat:CPCSection>A</pat:CPCSection>

 <pat:Class>61</pat:Class>

 <pat:Subclass>K</pat:Subclass>

 <pat:MainGroup>45</pat:MainGroup>

 <pat:Subgroup>06</pat:Subgroup>

 <com:SymbolPositionCode>L</com:SymbolPositionCode>

 <pat:CPCClassificationValueCode>I</pat:CPCClassificationValueCode>

 </pat:CPCClassification>

 <pat:CPCClassification>

 <pat:ClassificationVersionDate>2013-01-01</pat:ClassificationVersionDate>

 <pat:CPCSection>C</pat:CPCSection>

 <pat:Class>07</pat:Class>

 <pat:Subclass>D</pat:Subclass>

 <pat:MainGroup>211</pat:MainGroup>

 <pat:Subgroup>46</pat:Subgroup>

 <com:SymbolPositionCode>L</com:SymbolPositionCode>

 <pat:CPCClassificationValueCode>I</pat:CPCClassificationValueCode>

 </pat:CPCClassification>

 <pat:CPCClassification>

 <pat:ClassificationVersionDate>2013-01-01</pat:ClassificationVersionDate>

 <pat:CPCSection>G</pat:CPCSection>

 <pat:Class>01</pat:Class>

 <pat:Subclass>N</pat:Subclass>

 <pat:MainGroup>2333</pat:MainGroup>

 <pat:Subgroup>924</pat:Subgroup>

 <com:SymbolPositionCode>L</com:SymbolPositionCode>

 <pat:CPCClassificationValueCode>A</pat:CPCClassificationValueCode>

 </pat:CPCClassification>

 <pat:CPCClassification>

 <pat:ClassificationVersionDate>2013-01-01</pat:ClassificationVersionDate>

 <pat:CPCSection>G</pat:CPCSection>

 <pat:Class>01</pat:Class>

 <pat:Subclass>N</pat:Subclass>

 <pat:MainGroup>2333</pat:MainGroup>

 <pat:Subgroup>94</pat:Subgroup>

 <com:SymbolPositionCode>L</com:SymbolPositionCode>

 <pat:CPCClassificationValueCode>A</pat:CPCClassificationValueCode>

 </pat:CPCClassification>

 <pat:CPCClassification>

 <pat:ClassificationVersionDate>2013-01-01</pat:ClassificationVersionDate>

 <pat:CPCSection>G</pat:CPCSection>

 <pat:Class>01</pat:Class>

 <pat:Subclass>N</pat:Subclass>

 <pat:MainGroup>2800</pat:MainGroup>

 <pat:Subgroup>04</pat:Subgroup>

 <com:SymbolPositionCode>L</com:SymbolPositionCode>

 <pat:CPCClassificationValueCode>A</pat:CPCClassificationValueCode>

 </pat:CPCClassification>

 <pat:CPCClassification>

 <pat:ClassificationVersionDate>2013-01-01</pat:ClassificationVersionDate>

 <pat:CPCSection>G</pat:CPCSection>

 <pat:Class>01</pat:Class>

 <pat:Subclass>N</pat:Subclass>

 <pat:MainGroup>2800</pat:MainGroup>

 <pat:Subgroup>52</pat:Subgroup>

 <com:SymbolPositionCode>L</com:SymbolPositionCode>

 <pat:CPCClassificationValueCode>A</pat:CPCClassificationValueCode>

 </pat:CPCClassification>

 <pat:CPCCombinationSet>

 <pat:CPCGroupNumber>1</pat:CPCGroupNumber>

 <pat:CPCCombinationRank>

 <pat:CPCRankNumber>1</pat:CPCRankNumber>

 <pat:CPCClassification>

<pat:ClassificationVersionDate>2013-01-01</pat:ClassificationVersionDate>

 <pat:CPCSection>A</pat:CPCSection>

 <pat:Class>61</pat:Class>

 <pat:Subclass>K</pat:Subclass>

 <pat:MainGroup>31</pat:MainGroup>

 <pat:Subgroup>445</pat:Subgroup>

 <com:SymbolPositionCode>L</com:SymbolPositionCode> <pat:CPCClassificationValueCode>I</pat:CPCClassificationValueCode>

 </pat:CPCClassification>

 </pat:CPCCombinationRank>

 <pat:CPCCombinationRank>

 <pat:CPCRankNumber>2</pat:CPCRankNumber>

 <pat:CPCClassification>

<pat:ClassificationVersionDate>2013-01-01</pat:ClassificationVersionDate>

 <pat:CPCSection>A</pat:CPCSection>

 <pat:Class>61</pat:Class>

 <pat:Subclass>K</pat:Subclass>

 <pat:MainGroup>2300</pat:MainGroup>

 <pat:Subgroup>00</pat:Subgroup>

 <com:SymbolPositionCode>L</com:SymbolPositionCode>

 <pat:CPCClassificationValueCode>A</pat:CPCClassificationValueCode>

 </pat:CPCClassification>

 </pat:CPCCombinationRank>

 </pat:CPCCombinationSet>

 </pat:FurtherCPC>

 </pat:CPCClassificationBag>

</uspat:CPCMasterClassificationRecord>

## Table 1. CPCClassification element content model

|  |  |
| --- | --- |
| **<CPCClassification>** | The <CPCClassification> element defines one complete CPC Classification symbol. |
| **<ClassificationVersionDate>** | The <ClassificationVersionDate> element will occur one time within the <CPCClassification>element and contain an 8-position numeric date in the format YYYY-MM-DD identifying the classification publication date and terminated by a </ClassificationVersionDate> end tag. |
| **</ClassificationVersionDate>** | End tag of ClassificationVersionDate |
| **<CPCSection>** | The <CPCSection> element will occur one time within the <CPCClassification>element and contain a 1-position alphabetic (uppercase) – possible value can be “A through H” and terminated by a </CPCSection> end tag.The section is the highest hierarchical level within the classification scheme and as such it represents the whole body of knowledge which may be regarded as proper to the field of Classification. |
| **</CPCSection>** | End tag of CPCSection |
| **<Class>** | The <Class> element will occur one time within the < CPCClassification >element and contain a -2-position numeric class-type attribute and terminated by a </Class> end tag.The code denotes the second level subdivision of the classification scheme and as such it is a further breakdown of the section's broad technical fields into high level subject matter. |
| **</Class>** | End tag of CPC Class |
| **<Subclass>** | The <Subclass> element will occur one time within < CPCClassification > element and contain a1-position alphabetic (uppercase) – possible value can be “A through Z” and terminated by a </Subclass> end tag.The code denotes the third level subdivision of the classification scheme and as such it is a further breakdown of subject matter into more novel subject matter. |
| **</Subclass>** | End tag of CPC Subclass |
| **<MainGroup>** | The <MainGroup> element will occur one time within the < CPCClassification >element and contain a 1 to 4positions numeric and terminated by a </MainGroup> end tag.The code denotes the fourth level subdivision of the classification scheme and as such is a further breakdown of the novel subject matter. |
| **</MainGroup>** | End tag of CPC MainGroup |
| **<Subgroup>** | The <Subgroup> element will occur one time within the < CPCClassification >element and contain a 2 to 6 positions numeric and terminated by a </Subgroup> end tag.The code denotes the fifth level subdivision of the classification scheme and as such is a further breakdown of the novel subject matter. |
| **</Subgroup>** | End tag of CPC Subgroup |
| **<SymbolPositionCode>** | The <SymbolPositionCode> element will occur one time within the <CPCClassification >element and contain 1-position alphabetic (uppercase) – “F” defining “first” for the sole or first “invention information” CPC, or “L” defining “later” for any second and succeeding “invention information” CPC and for any “non-invention information” CPC. And, terminated by a </SymbolPositionCode> end tag.The code that specifies the position of the classification symbol. |
| **</SymbolPositionCode>** | End tag of SymbolPositionCode |
| **<ClassificationValueCode>** | The <ClassificationValueCode> element will occur one time within the < CPCClassification>element and contain a1-position alphabetic (uppercase) – “I” defining “invention information” or “A” defining “Additional information”. And, terminated by a </ClassificationValueCode> end tag.The code that distinguishes between invention information (invention) and additional information (non-invention/additional), when describing a classification symbol on a document. |
| **</ClassificationValueCode>** | End tag of ClassificationValueCode |
| **<ActionDate/>** | ActionDate element is not used in US CPC. |
| **<GeneratingOfficeCode/>** | GeneratingOfficeCode is not used in US CPC. |
| **</CPCClassification>** | End tag of CPCClassification |

## Table 2. CPCClassificaitonBag content model

|  |  |
| --- | --- |
| **<CPCClassificationBag>** | The <CPCClassificationBag>element occurs one time within the <MasterClassificationRecord>element and contains the MainCPC (CPCClassification), a FurtherCPC element with one or more CPCClassification element and/or CPCCombinationSet. It is terminated by the </CPCClassificationBag>end tag. |
| **<MainCPC>** | The <MainCPC>element is mandatory and will occur one time within the <CPCClassificationBag>element and contain the Main CPC Classification and terminated by the </MainCPC>end tag. |
| **<CPCClassification/>** | One CPCClassification element under MainCPC. See Table 1 above. |
| **</MainCPC>** | End tag of MainCPC |
| **<FurtherCPC>** | The <FurtherCPC>element is optional and will one or more CPCClassification element and an optional CPCCombinationSet with one or more CPCClassification element. And terminated by the </FurtherCPC>end tag. |
| **<CPCClassification/>** | See Table 1 above. |
| **<CPCCombinationSet>** | A combination set is a group of CPC symbols that have one base class and one or more subsequent ranked symbols that are linked together to convey special classification information. |
| **<CPCGroupNumber>** | The <CPCGroupMumber> element will occur one time within each <CPCCombinationSet>element and contain a numeric value that is used to identify a group of symbols, when allocating a combination set of symbols to a patent document. And, terminated by a </CPCGroupNumber> end tag. |
| **</CPCGroupNumber>** | End CPCGroupNumber tag |
| **<CPCCombinationRank>** | The <CPCCombinationRank> element will occur one time within each <CPCCombinationSet>element and contain a sequential number that is used to identify the rank of a symbol within a combination set. (Order of the symbols is important). And, terminated by a </CPCCombinationRank> end tag. |
| **<CPCRankNumber>** | The <CPCRankNumber>element will occur one time within each <CPCCombinationRank>element and contain a numeric value and terminated by a </CPCRankNumber> end tag. |
| **</CPCRankNumber>** | End tag of CPCRankNumber |
| **<CPCClassification/>** | See Table 1 above. |
| **</CPCCombinationRank>** | End CPCCombinationRank tag |
| **</CPCCombinationSet>** | End CPCCombination tag |
| **</FurtherCPC>** | End FurtherCPC tag |
| **</CPCClassificationBag>** | End CPCClassificaitonBag tag |

## Table 3. CPCMasterClassificationFile content model

|  |  |
| --- | --- |
| **<CPCMasterClassificationFile>** | <MasterClassificationFile> is a collection of <MasterClassificationRecord>. It is the root element of the document. |
| **@publicationStartNumber** | This attribute lists the start publication number of the CPCMasterClassificationFile |
| **@publicationEndNumber** | This attribute lists the end publication number of the CPCMasterClassificationFile |
| **<CPCMasterClassificationRecord>** | The <CPCMasterClassificationRecord>element is mandatory and will occur one time within the <CPCClassificationBag>element and contain the patent application identification, patent publication identification and the CPCClassificationBag. |
| **<ApplicationIdentification>** | The Application information of the Patent Application |
| **<IPOfficeCode>** | IP Office Code of US patent is always ‘US’  |
| **</IPOfficeCode>** | End tag of IPOfficeCode |
| **<ApplicationNumber>** | Numbers used by IPOs in order to identify each application received. It can contain ApplicationNumberText or WIPO ST.13ApplicationNumber |
| **<ApplicationNumberText>** | This element is the US patent application number. |
| **</ApplicationNumberText>** | End tag of ApplicationNumberText |
| **</ApplicationNumber>** | End tag of ApplicationNumber |
| **</ApplicationIdentification>** | End tag of ApplicationIdentification |
| **<PatentPublicationIdentification>** | The publication information of the patent application |
| **<IPOfficeCode>** | IP Office Code of US patent is always ‘US’  |
| **</IPOfficeCode>** | End tag of IPOfficeCode |
| **<PublicationNumber>** | US Pre-Grant patent application publication number |
| **</PublicationNumber>** | End tag of PublicationNumber |
| **<PatentDocumentKindCode>** | The document kind code of Pre-Grant patent application publication |
| **</PatentDocumentKindCode>** | End tag of PatentDocumentKindCode |
| **PublicationDate** | The Pre-Grant patent application publication date in yyyy-mm-dd date format. |
| **</PublicationDate>** | End tag of PublicationNumber |
| **</PatentPublicationIdentification>** | End tag of PatentPublicationIdenfication |
| **<CPCClassificationBag>** | One <MasterClassifictionRecord> contains one <CPCClassificationBag> element. See Table 2 above for details. |
| **</CPCClassificationBag>** | End tag of CPCClassificationBag |
| **</MasterClassificationRecord>** | End tag of MasterClassificationRecord |
| **</MasterClassificationFile>** | End tag of MasterClassificationFile |

NOTE:

1. To promote data exchange, the XML schema is implementing the World Intellectual Property Office (WIPO) ST. 96 standard (<http://www.wipo.int/standards/en/part_03_standards.html>). XML namespace prefix “uspat” is for US specific components. Prefix “pat” is for ST. 96 Patent namespace and prefix “com” is for ST. 96 Common namespace.

2. In each record, we identify the patent application number if it is available, the patent number, kind code, grant date and CPC Classification bag with Main CPC and Further CPCs. The XML content model schema of CPCMasterClassificationRecord is based on WIPO ST.96 standards with the exceptions noted below.

3. Main CPC was changed to allow empty content to accommodate the rare case of a missing first position allocation. The data would be corrected in the next CPC-MCF publication.

4. In CPCClassification element content model, GrantDate was modified to allow a null value due to missing data.

5. In Further CPC, the Classifications were sorted by CPCClassification then CPCCombinationSet (with group number and rank number), CPCClassifications was sorted by classification values (I - Invention or A - Additional), then sorted by the symbol sorting order. In combination set, they were sorted by group number and rank number.

**US\_Grant\_CPC\_MCF\_Text\_yyyy-mm-dd.zip (currently includes 1 folder of over 200 text files)**

# Corresponding to each US-Grant\_CPC\_MCF\_XML file, there is a file for the Text version of the US Grant CPC Master Classification File.

# There are currently over 200 text files. The number of files will grow as more U.S. Patents are granted/issued. For early data, each file contains no more than 50,000 patent numbers excluding withdrawn patents.

US\_Grant\_CPC\_MCF\_00000001.txt contains patent numbers between 1 and 99999.

US\_Grant\_CPC\_MCF\_00100000.txt contains patent numbers between 100000 and 199999.

US\_Grant\_CPC\_MCF\_00200000.txt contains patent numbers between 200000 and 299999.

.

.

.

US\_Grant\_CPC\_MCF\_09850000.txt contains patent numbers between 9850000 and 9899999.

US\_Grant\_CPC\_MCF\_09900000.txt contains patent numbers between 8900000 and 9949999. US\_Grant\_CPC\_MCF\_09950000.txt contains patent numbers between 9950000 and 9999999. US\_Grant\_CPC\_MCF\_10000000.txt contains patent numbers between 10000000 and 10499999.

The records are sorted exactly the same way as in XML format. They are sorted first by patent number ascending and kind code. For the same patent number and kind code, the records are sorted by symbol position (F - First or L - Later); Among Later records, they were sorted by classification symbol and combination sets. The Later records not in combination set are sorted by classification value (I - Invention or A -Additional), then by classification symbol order ascending. In a combination set, the records are sorted by group number and rank number.

There is a possibility of one patent number with more than one kind code. The kind code value can be “B1”, “B2” or “A”. Each kind code will have its own record for the grant number. This situation is due to the reexamination certificates.

## Text File Example:

*A 1B61C 11/04 20130101FI 0 0*

*A 1Y10T 16/3819 20150115LA 0 0*

*A 2C02F 1/00 20130101FI 0 0*

*A 3G01B 7/107 20130101FI 0 0*

*B215117487 9900000H03K 17/14 20130101FI 0 0*

*B215117487 9900000H03K 17/165 20130101LI 0 0*

*B215117487 9900000H03K 17/168 20130101LI 0 0*

*B215117487 9900000H03K 17/28 20130101LI 0 0*

*B21464371910000000G01S 7/4863 20130101FI 0 0*

*B21464371910000000G01S 7/4865 20130101LI 0 0*

*B21464371910000000G01S 7/4914 20130101LI 0 0*

*B21464371910000000G01S 7/4917 20130101LI 0 0*

*B21464371910000000G01S 13/89 20130101LI 0 0*

## Line interpretation

***B213445338 9000011A61K2300/00 20130101LA 1 2***

***123456789012345678901234567890123456789012345678***

Position [01-02] US Patent Document Kind Code, they can be “B1”, “B2” or “A“

Position [03-10] 8-digit US Patent Application Number, empty space if not available

Position [11-18] Up to 8 digits US Patent Grant Number (right align)

Position [19] CPC Section

Position [20-21] CPC Class

Position [22] CPC Subclass

Position [23-26] Up to 4 digits CPC Main Group (right align)

Position [27] “/” Separator

Position [28-33] Up to 6 digits CPC Subgroup (left align)

Position [34-41] CPC Classification Version Date in YYYYMMDD format

Position [42] CPC Symbol Position (“F” or “L”)

Position [43] CPC Classification Value Code (“I” or “A”)

Position [44-46] Up to 3 digits CPC Classification Combination Set Group Number (right align)

Position [47-48] Up to 2 digits CPC Classification Combination Set Rank Number (right align)