

COOPERATIVE PATENT CLASSIFICATION



Prepared by USPTO and EPO

CPC Scheme Files Specification

Date	By	Version	Status	Modification
March 3, 2021	USPTO	0.1	Draft	Creation
May 28, 2021	USPTO	1.0	Revised	Clarified description

Contact: cpc@epo.org and/or at cpc@uspto.gov

Table of Contents

CPC SCHEME FILES SPECIFICATION.....	1
1. INTRODUCTION	3
2. BACKGROUND.....	3
3. CONTENT OF THE XML FILES	3
3.1 Information hierarchy	3
3.2 CPC class scheme object	4
3.3 Root element	4
3.4 CPC classification-item entry properties	4
3.5 Descriptive part of the CPC entry	7
3.6 Common elements	10
3.7 Use of special characters for the representation of chemical bonds	12
4. NOTABLE DIFFERENCES FROM IPC SCHEME XML.....	12
5. SAMPLES	12
5.1 Section entries with links to subclasses.....	12
5.2 Guidance headings and placeholder entries.....	13
5.3 Main trunk symbol entries	14
5.4 Indexing code entries	14
5.5 Sample reference to image.....	15

1. INTRODUCTION

The purpose of this document is to describe the structure and content of the Cooperative Patent Classification (CPC) scheme files corresponding to the following W3C XML Schema:

- `cpc-scheme-v1_7.xsd` (effective in the 2016.11 CPC release).

2. BACKGROUND

In January 2013, the CPC was launched as a joint patent classification system based on the European classification system (ECLA), also including practices from the United States Patent Classification (USPC) system. The CPC is based on the International Patent Classification (IPC) system, and shares the same general structure of the IPC Scheme and Definitions.

The CPC Scheme XML files serve as the authoritative data source of the CPC Scheme. The CPC Scheme XML is intended for upload into IT systems each time a new version of the CPC becomes available.

This specification describes the Scheme information exchanged as well as notable differences from the IPC XML structure.

3. CONTENT OF THE XML FILES

The XML files are encoded in Unicode UTF-8 character set. Special characters beyond the Basic Latin character set are encoded using numeric character references.

3.1 Information hierarchy

The CPC Scheme is subdivided into one section index file (`cpc-scheme.xml`), a set of files corresponding to each Section (e.g. `cpc-scheme-B.xml`, ...), and a set of files corresponding to each Subclass (e.g. `cpc-scheme-A61K.xml`, ...).

The CPC Scheme XML is arranged hierarchically (from top-down):

- (for section files): Section, Subsection (including placeholders), Class, Subclass
- (for subclass files): Subclass, Guidance heading (including placeholders), Main group, Subgroups (up to 12 dots)

Contrary to the IPC Scheme XML, the CPC Scheme XML consists only of "structured" entries. In the IPC Scheme XML, non-structured entries include: Subsections, Guidance headings, Notes, Subclass Indices (not adopted in CPC).

In order to maintain the hierarchical structure, placeholders are given for entries corresponding subsections and guidance headings in the CPC XML where no subsections or guidance headings occur within the scope of the Section or Subclass.

In addition, Notes (and Warnings) are always located with the CPC entry (symbol, or rarely, guidance headings).

In the CPC Scheme XML, properties are defined in attributes.

3.2 CPC class scheme object

The file contains a list of CPC class scheme objects, each including

- One CPC entry (`classification-item`)
 - Properties of the CPC entry (symbol, level, ...)
 - Descriptive part of the CPC entry (title, note, warning)

3.3 Root element

In the CPC Scheme, the root element name is `class-scheme`. The root element contains `classification-item` elements, which are hierarchically arranged.

3.3.1 CPC class-scheme element properties

Name: `publication-date`

Indicates the date of entry into force of the corresponding CPC version.

Name: `scheme-type`

Value: "cpc" corresponds to the classification scheme system.

Name: `publication-type`

Value: "official". Indicates the type of publication.

3.4 CPC classification-item entry properties

Within the `classification-item` element, the following properties exist:

3.4.1 `breakdown-code`

Name: `breakdown-code`

Value: Originally intended to distinguish breakdown indexing codes from orthogonal indexing codes. Breakdown indexing codes are 2000-series (additional-only) subgroups of main groups used to represent invention information. Orthogonal indexing codes correspond to those 2000-series main groups and their subgroups, which are used to identify additional-only information. This attribute is currently used to indicate any indexing code.

- `true` = Any indexing code (includes **both** breakdown codes and orthogonal codes).
- `false` (default) = A entry which does not correspond to an indexing code.

3.4.2 `not-allocatable`

Name: `not-allocatable`

Value: Used to indicate that documents cannot be allocated to this symbol for classification. A symbol is allocatable only when it corresponds to a group or subgroup.

- `true` = documents cannot be allocated to this entry
- `false` (Default) = documents can be allocated to the symbol of this entry

3.4.3 level

Name: level

Value: Indicates the hierarchical indent level of the entry. Note that Guidance and Subsection headings are considered part of the hierarchy in the CPC (contrary to the IPC).

- 2 = Section
- 3 = Subsection, or a subsection placeholder in order to maintain the hierarchical structure)
- 4 = Class
- 5 = Subclass
- 6 = Guidance heading, or a guidance heading placeholder in order to maintain the hierarchical structure
- 7 = Main group
- 8-19 = Subgroups from 1 dot to 12 dots (respectively)

3.4.4 additional-only

Name: additional-only

Value: Used to indicate symbols which can only have a classification value of "additional information." All 2000-series symbols (indexing codes) and Y-series symbols are given a value of "true". All other symbols should be given a value of "false". Applies only when attribute `not-allocatable` indicates that documents may be allocated.

- true = the corresponding group can only be allocated as "additional information"
- false = the corresponding group can be allocated as either of "additional information" or "invention information"

3.4.5 link-file

OPTIONAL

Name: link-file

Value: name of the corresponding subclass XML file, when the `classification-item` element appears in a section XML file. Used to link section and subclass XML together. For example, Section B would contain a link to the file "cpc-scheme-B01K.xml" corresponding to the subclass entry.

3.4.6 sort-key

Name: sort-key

Value: formatted symbol for sorting purposes. "Main trunk" symbols have sort-keys which are identical to the symbol. Indexing codes are assigned sort-keys equivalent to the group part minus 2000.

Scheme symbols appear in alphanumerical order according to their sort-keys.

For example, group A61K 2039/505 has a sort-key of 'A61K 39/505', and appears in the CPC Scheme after group A61K 39/44. Breakdown indexing codes (e.g. A61K 2039/505) have sort-keys which place them subordinate to a main trunk symbol, while orthogonal indexing codes (e.g. C07K 2317/00) have sort-keys which place them toward the end of the subclass Scheme.

3.4.7 definition-exists

OPTIONAL

Name: definition-exists

Value:

- true = A Definition exists for this symbol. In CPC, Definitions are only available at the Subclass or Group level.
- false = No Definition exists for this symbol.

3.4.8 ipc-concordant

OPTIONAL

Name: ipc-concordant

Value: Indicates the symbol in the IPC Scheme where the subject-matter corresponds. Where there is no IPC Scheme equivalent, CPCONLY is given (e.g. Y-series symbols and CPC-only orthogonal indexing codes). The full collection of IPC-concordant values comprises the master CPC-to-IPC Concordance List (CICL). IPC-concordant values are provided for all allocatable CPC symbols (groups and subgroups). Since CPC is in most of the cases a refinement of the current IPC, this attribute typically reflects this extension of IPC. In most cases, the value of this attribute corresponds to the closest IPC parent group from which CPC extends.

- CPCONLY, or
- an existing IPC symbol (corresponding to IPC numbering. For instance, if an indexing code corresponds to an IPC indexing group, the CPC classification symbol will have CPC's 2000-series numbering convention, but the IPC-concordant value will correspond to the symbol number as it appears in IPC).

The IPC-concordant values are updated when the IPC scheme is updated and/or through revision of the CPC. The IPC-concordant value is provided through intellectual feedback, and cannot be used to infer the hierarchical relationship between the CPC and IPC schemes.

3.4.9 c-set-base-allowed

OPTIONAL

Name: c-set-base-allowed

Value: Indicates if a symbol is allowed for allocation as the first symbol in any combination set (rank 1). The symbol should only be used as a base symbol in accordance with the particular C-Set rule governing the area. This information corresponds to the "List of technical areas where Combination Sets are authorised for classification" available on cpcinfo.org.

- true = the symbol may be used as a base symbol in a C-Set, in accordance with the relevant C-Set rules
- false = the symbol cannot be used as a base symbol in any C-Set

3.4.10 c-set-subsequent-allowed

OPTIONAL

Name: c-set-subsequent-allowed

Value: Indicates if a symbol is allowed for allocation as a subsequent symbol in any combination set (rank 2 and further). The symbol should only be used as a subsequent symbol in accordance with the particular C-Set rule governing the area.

Due to current limitations with the data available in the "List of technical areas where Combination Sets are authorised for classification", this attribute is currently set to "true" for all groups which are not in Section Y.

- true = the symbol may be used as a subsequent symbol in a C-Set, in accordance with the relevant C-Set rules
- false = the symbol cannot be used as a subsequent symbol in any C-Set

3.4.11 date-revised

Name: date-revised

Value: Indicates the CPC publication version where an update to the symbol, indent, title, or notes/warnings may have been made. This is not an authoritative indication of revision.

3.4.12 status

Name: status

Value: Indicates the status of the symbol as published or frozen.

- published (Default) = Symbol is active in the published scheme.
- frozen = Indicates that the group will be deleted after reclassification of documents is completed in a process known as "finalisation". Documents currently allocated to a frozen group will be reclassified according to the Revision Concordance List. When a symbol is frozen, new documents cannot be allocated.

3.5 Descriptive part of the CPC entry

CPC descriptive elements are content-oriented, rather than data-oriented. Therefore, components of a title, such as title parts and reference parts (as governed by the "Guide to the IPC"), are not directly reflected in the title components.

For example, the title for A23F 5/243 is "{Liquid, semi-liquid or non-dried semi-solid coffee extract preparations; Coffee gels; Liquid coffee in solid capsules (A23F 5/246 takes precedence)}"

This group has three title parts, and one reference part associated with the last title part. This title is represented in XML in the following manner, with a single `title-part` element:

```
<class-title date-revised="2013-01-01">
  <title-part>
    <CPC-specific-text>
      <text>Liquid, semi-liquid or non-dried semi-solid coffee
extract preparations; Coffee gels; Liquid coffee in solid capsules
</text>
      <reference>
        <text><class-ref scheme="cpc">A23F5/246</class-ref> takes
precedence</text>
      </reference>
    </CPC-specific-text>
  </title-part>
</class-title>
```

3.5.1 classification-symbol

The CPC symbol of the entry. For guidance or subsection heading entries, or their respective placeholder entries, the element `classification-symbol` corresponds to the first main group (or first class symbol for subsection headings). The `classification-symbol` is not displayed for heading entries or their placeholders (corresponding to `classification-item` attribute `level` of '3' or '6') in the presentation of the CPC Scheme.

3.5.2 class-title

Text content for every CPC symbol title and guidance or subsection heading starts with a `class-title`.

3.5.2.1 title-part

One or more title parts start with a `title-part`.

3.5.2.2 CPC-specific-text

Contains `text` or `reference` elements. Represents text which should be specific only to CPC (and is not present in IPC). This element is used to indicate text which extends from the existing IPC title text. In addition, CPC-only groups which are breakouts (i.e. subgroups) of IPC groups would have the entire title text captured in a `CPC-specific-text` element. Due to legacy data quality issues, the presence or absence of `CPC-specific-text` cannot be interpreted as an authoritative indicator of differences between CPC and IPC. For presentation purposes, curly brackets (braces) surround text in this element.

3.5.2.3 reference

Contains `text` or `CPC-specific-text` elements. For presentation purposes, parentheses surround text in this element. Reference parts are separated by a semicolon in the `text` element.

3.5.2.4 text

mixed content model:

- UTF-8 encoded Unicode characters
- reference symbol element `class-ref`
- rich text elements: (underline, (_sub)script, (^sup)erscript
- external image references: `media`

Note: the use of inline underline formatting is generally reserved to Latin phrases and certain standard expressions in the CPC Scheme.

3.5.3 notes-and-warnings

Text content for all Scheme Notes or Warnings start with a `notes-and-warnings` element.

3.5.3.1 note

Contains one or more Notes or Warnings (each immediate child `note-paragraph` element within the `note` element). Multiple Notes and Warnings are automatically numbered in presentational views. For presentation purposes, a heading (e.g. "Note" or "Warning") precedes the `note` element content.

type attribute:

Value: value of "note" is to indicate Note content or value of "warning" to indicate Warning content.

3.5.3.1.1 note-paragraph

mixed content model:

- UTF-8 encoded Unicode characters
- reference symbol element `class-ref`
- rich text elements: (underline, (_sub)script, (^sup)erscript
- line break element `br`
- external image references: `media`
- `CPC-specific-note` element
- subnote list-style content
- table content
- (`pre`)formatted content: deprecated element used to convey content which should be displayed as a table. This element is being progressively removed or replaced with the `table` element.
- `reclassification-date` element

`CPC-specific-note`: Represents Note or Warning text which should be specific only to CPC (and is not present in IPC). Analogous to `CPC-specific-text`, `CPC-only` groups which are breakouts (i.e. subgroups) of IPC groups would have the entire Note (or Warning) text captured in a `CPC-specific-note` element. Due to legacy data quality issues, the presence or absence of `CPC-specific-note` cannot be interpreted as an authoritative indicator of differences between CPC and IPC. For presentation purposes, curly brackets (braces) surround text in this element.

Mixed content model:

- UTF-8 encoded Unicode characters
- reference symbol element `class-ref`
- rich text elements: (underline, (_sub)script, (^sup)erscript
- external image references: `media`

- nested lists: `subnote`
- nested `table`
- `subnote`: nested list within a Note. Each list-item is contained in one or more `note-paragraph` elements. The list style is indicated with the `type` attribute. Possible `type` attribute values are "bullet", "Alpha" (uppercase), "alpha" (lowercase), "Roman" (uppercase), "roman" (lowercase), and "number".
- `table`: nested table content within a Note or Warning. Contains one or more `row` elements, which in turn contain table cell content in `entry` elements.
- `entry`: mixed content model:
 - UTF-8 encoded Unicode characters
 - reference symbol element `class-ref`
 - rich text elements: `(u)nderline`, `(sub)script`, `(sup)erscript`
 - external image references: `media`

Scheme Notes and Warnings share common content models, wherein the Warning content model is a subset of the content model used for Notes.

3.5.4 meta-data:

Not used.

3.6 Common elements

These elements are used within note and title elements described above.

- reference symbol element: `class-ref`
- rich text elements: `(u)nderline`, `(sub)script`, `(sup)erscript`
- external image references: `media`

3.6.1 class-ref

A reference to a CPC entry by its classification symbol.

3.6.1.1 CPC reference symbol properties

Classification scheme of origin

Name: scheme

Value: cpc, ipc, not-mapped.

Purpose: Indicates the classification system to which the referenced symbol belongs. Almost all reference symbols correspond to the CPC system.

In cases where CPC deviates from IPC, a Scheme Warning will provide details on this deviation from IPC. In this case, references to symbols in the IPC system which are not adopted in the CPC system have a value of 'ipc'.

References to symbols which do not correspond to valid CPC entries are set to 'not-mapped', and are being progressively removed or replaced with valid reference symbols.

3.6.2 media

A reference to an image

Image properties

3.6.2.1 alternative text

Name: alt

Value: Contains alternative text describing the image.

3.6.2.2 image file name

Name: file-name

Value: Contains the image file name (following the convention: "cpc-sch-<subclass>-<####>.ext", where ext refers to the extension corresponding to the image type).

3.6.2.3 image id

Name: id

Value: identifier associated with the image.

3.6.2.4 image file type

Name: type

Value: Indicates the image type: 'gif', 'jpeg', 'tif', 'bmp', 'png', or 'unknown'. Typical Scheme images correspond to gif type.

3.7 Use of special characters for the representation of chemical bonds

The following characters/images are derived from IPC convention, and describe the intended representation of chemical bonds when they appear inline in Scheme titles or Notes

— (em dash), represented as “—”	simple bond in inline chemical formulae
= (equal sign)	double bond in inline chemical formulae
≡ (identical to sign), represented as “≡”	triple bond in inline chemical formulae
⤴ (image only, corresponding to llinkthree.gif used in IPC)	left triple hydrogen bond in inline chemical formulae
⤵ (image only, corresponding to rlinkthree.gif used in IPC)	right triple hydrogen bond in inline chemical formulae
⤵ (image, corresponding to llinkt.gif used in IPC) or special character such as right-pointing angle bracket, represented as “〉”, or > (greater than), represented with reserved entity “>”	left double hydrogen bond in inline chemical formulae
⤵ (image, corresponding to rlinkt.gif used in IPC) or special character such as left-pointing angle bracket, represented as “〈”, or < (less than), represented with reserved entity “<”	right double hydrogen bond in inline chemical formulae

4. NOTABLE DIFFERENCES FROM IPC SCHEME XML

IPC Warnings are stored separately from the Scheme. In CPC, Warnings are considered to be part of the Scheme content, and therefore are included in the scheme data files.

IPC Notes and Guidance/Subsection headings are considered "non-structured" entries. As such, an IPC Note may encompass a range of symbols. Where such Notes also adopted in the CPC Scheme, they have been relocated to a single symbol (or guidance heading) entry, as all entries in CPC are "structured" and anchored in the Scheme hierarchy.

The IPC Scheme includes indexing codes (`ipcEntry` with `kind` attribute value "I"). Where these codes are adopted into CPC, the symbol number corresponds to a 2000-series group number, and corresponding attributes `additional-only` and `breakdown-code` are set to "true".

5. SAMPLES

5.1 Section entries with links to subclasses

```
<classification-item breakdown-code="false" not-allocatable="true" level="4"
  additional-only="false" sort-key="B04" date-revised="2019-05-01"
  status="published">
  <classification-symbol>B04</classification-symbol>
  <class-title date-revised="2019-05-01">
    <title-part>
      <text>CENTRIFUGAL APPARATUS OR MACHINES FOR CARRYING-OUT
      PHYSICAL OR CHEMICAL PROCESSES</text>
    </title-part>
  </class-title>
  <classification-item breakdown-code="false" not-allocatable="true"
  level="5" additional-only="false" link-file="cpc-scheme-B04B.xml"
  sort-key="B04B" definition-exists="true" date-revised="2018-08-01"
  status="published">
```

```

    <classification-symbol>B04B</classification-symbol>
    <class-title date-revised="2013-01-01">
      <title-part>
        <text>CENTRIFUGES </text>
        <reference>
          <text>high-speed drum mills <class-ref
scheme="cpc">B02C19/11</class-ref>; domestic spin driers <class-ref
scheme="cpc">D06F</class-ref>; analysing, measuring or monitoring
physical or chemical properties of samples during centrifuging,
<u>see</u> the relevant subclasses for these procedures, e.g. <class-ref
scheme="cpc">G01N</class-ref></text>
          </reference>
        </title-part>
      </class-title>
      <notes-and-warnings date-revised="2013-01-01">
        <note type="note">
          <note-paragraph>This subclass <u>covers</u> machines or
apparatus for separating, mixing, drying, extracting, purifying, or like
treating in which centrifugal effects are generated by rotary bowls or
other rotors. Where such machines or apparatus involve pumping effects,
such effects must be incidental or subsidiary to the treating. </note-
paragraph>
        </note>
        <note type="warning">
          <note-paragraph>In this subclass non-limiting references
(in the sense of paragraph 39 of the Guide to the IPC) may still be
displayed in the scheme.</note-paragraph>
        </note>
      </notes-and-warnings>
    </classification-item>
...
</classification-item>

```

5.2 Guidance headings and placeholder entries

Machines for harvesting root crops

```

<classification-item breakdown-code="false" not-allocatable="true" level="6"
additional-only="false" sort-key="A01D13/00" date-revised="2013-01-01"
status="published">
  <classification-symbol>A01D13/00</classification-symbol>
  <class-title date-revised="2013-01-01">
    <title-part>
      <text>Machines for harvesting root crops</text>
    </title-part>
  </class-title>
  <classification-item breakdown-code="false" not-allocatable="false"
level="7" additional-only="false" sort-key="A01D13/00"
definition-exists="true" ipc-concordant="A01D13/00"
c-set-base-allowed="false" c-set-subsequent-allowed="true"
date-revised="2013-01-01" status="published">
    <classification-symbol>A01D13/00</classification-symbol>
    <class-title date-revised="2013-01-01">
      <title-part>
        <text>Diggers, e.g. potato ploughs</text>
      </title-part>
    </class-title>
  </classification-item>
...
</classification-item>

```

Placeholder (indicated by horizontal bar when preceded by a guidance heading)

```
<classification-item breakdown-code="false" not-allocatable="true" level="6"
  additional-only="false" sort-key="F02M21/00" date-revised="2019-02-01"
  status="published">
  <classification-symbol>F02M21/00</classification-symbol>
  <classification-item breakdown-code="false" not-allocatable="false"
  level="7" additional-only="false" sort-key="F02M21/00"
  definition-exists="true" ipc-concordant="F02M21/00"
  c-set-base-allowed="false" c-set-subsequent-allowed="true"
  date-revised="2021-01-01" status="published">
    <classification-symbol>F02M21/00</classification-symbol>
    <class-title date-revised="2013-01-01">
      <title-part>
        <text>Apparatus for supplying engines with non-liquid
        fuels, e.g. gaseous fuels stored in liquid form</text>
      </title-part>
    </class-title>
    ...
  </classification-item>
</classification-item>
```

5.3 Main trunk symbol entries

```
<classification-item breakdown-code="false" not-allocatable="false"
  level="10" additional-only="false" sort-key="C01F17/241"
  definition-exists="false" ipc-concordant="C01F17/241"
  c-set-base-allowed="false" c-set-subsequent-allowed="true"
  date-revised="2020-01-01" status="published">
  <classification-symbol>C01F17/241</classification-symbol>
  <class-title date-revised="2020-01-01">
    <title-part>
      <text>containing two or more rare earth metals, e.g.
      NdPrO<sub>3</sub> or LaNdPrO<sub>3</sub></text>
    </title-part>
  </class-title>
  <notes-and-warnings>
    <note type="warning">
      <note-paragraph warning-type="reclass-destination">Group <class-
      ref scheme="cpc">C01F17/241</class-ref> is incomplete pending
      reclassification of documents from group <class-ref
      scheme="cpc">C01F17/206</class-ref>. <br/>Groups <class-ref
      scheme="cpc">C01F17/206</class-ref> and <class-ref
      scheme="cpc">C01F17/241</class-ref> should be considered in order to perform
      a complete search.</note-paragraph>
    </note>
  </notes-and-warnings>
</classification-item>
```

5.4 Indexing code entries

Breakdown indexing code

```
<classification-item breakdown-code="true" not-allocatable="false"
  level="10" additional-only="true" sort-key="G01K3/145"
  definition-exists="false" ipc-concordant="G01K3/14"
  c-set-base-allowed="false" c-set-subsequent-allowed="true"
  date-revised="2013-01-01" status="published">
  <classification-symbol>G01K2003/145</classification-symbol>
  <class-title date-revised="2013-01-01">
    <title-part>
      <CPC-specific-text>
        <text>Hotspot localization</text>
      </CPC-specific-text>
    </title-part>
  </class-title>
</classification-item>
```

```
    </title-part>
  </class-title>
</classification-item>
```

Orthogonal indexing code

```
<classification-item breakdown-code="true" not-allocatable="false" level="8"
  additional-only="true" sort-key="B60L200/16" definition-exists="false"
  ipc-concordant="CPCONLY" c-set-base-allowed="false"
  c-set-subsequent-allowed="true" date-revised="2013-01-01"
  status="published">
  <classification-symbol>B60L2200/16</classification-symbol>
  <class-title date-revised="2013-01-01">
    <title-part>
      <text>Single-axle vehicles</text>
    </title-part>
  </class-title>
</classification-item>
```

5.5 Sample reference to image

```
<media id="1029" file-name="cpc-sch-A61K-1029.gif" type="gif"/>
```

End of document