

Improving Intellectual Access to the U.S. Patent Classification

Bruce B. Cox

Office of Documentation Information, U.S. Patent and Trademark Office
Washington, DC 20231, USA

ABSTRACT

Understanding the U.S. Patent Classification (USPC) scheme is a significant challenge to the general public and the inventor. This is due in part to the highly discriminative and technical terminology used in the *Manual of Classification* which has developed in response to the needs of patent examiners. It is also due to the structure of the *Manual*, a complex hierarchical tree of 125,000 classifications¹ with as many as 15 levels from root to leaf. Each classification has associated with it a paragraph or more of text defining it and often providing references to other classifications. A prototype hypertext *Manual of Classification* incorporates a few small classes of this highly structured body of text with many non-intuitive cross-links. This prototype also includes a method for using terminology from various subject-specific controlled vocabularies (presumably more familiar to the general public).

BRIEF HISTORY OF THE USPC

Beginning in 1830, subject lists of U.S. patents appeared at irregular intervals. In 1872 a system of 145 classes superseded all prior classification schemes, and formed the framework upon which the present system is built. Starting with the 167 classes in the 1882 schedule, rudimentary cross-references were introduced, along with some explanatory notes, all of which were dropped from the 1897 schedule of 215 classes. Not until 1898 did Congress authorize employment of a special force "to revise and perfect the classification by subject matter, of all letters patent and publications ... which constitute the field of search" Since then, the U.S. Patent Classification has grown to over 400 classes divided into a total of more than 125,000 subclasses, and the "special force" now amounts to just over 100 full-time classifiers with additional supporting staff.²

STRUCTURE AND USE OF THE USPC

The USPC is not one classification scheme, but a collection of 407 schemes, each one reflecting the ideas and ability of the classifiers who created or modified it, the state of a particular field of technology, or *art*, at one or several points in time, and the pressures of time and other expediencies. Every subclass has a definition, a kind of elaborate scope note. The content of definitions ranges from fairly trivial to what may constitute a mini-tutorial on the state of one very narrow aspect of the art. A class definition elaborates the scope of the entire class and often includes a glossary of terms. Each class and subclass definition instructs the reader to see various other classifications for specific technologies not included under its rubric.

1. By convention, a class/subclass combination are referred to as a classification; e.g., subclass 35 of class 200 is expressed as the classification 200/35.
2. *Development and Use of Patent Classification Schemes*, January 1966, U.S. Patent Office. Distributed by GPO. LC Catalog No. 65-62235. Much of this section paraphrased from its historical introduction.

Search notes, or *search lines* as such instructions are called, document the classifier/examiner-negotiated borderlines which separate one classification from another. These conventions often appear to have been established without regard to logic or the USPC hierarchy. They are, nevertheless, of paramount importance for a successful patent search, since they cannot be arrived at by reason — they are strictly a matter of revelation.

Over the last 150 years different philosophies of classification have prevailed. In a "modern" class (post World War II), classifiers arrange patents primarily on the basis of the utility of the invention as described in the claims. Utility means function, effect, product, use, or application, depending on the nature of the technology. It is used in the very restrictive sense of direct, proximate, or necessary function rather than remote or accidental function. Applying utility as a basis for classification results in collecting together patents that claim similar processes or means that achieve similar results by the application of similar natural laws. Classes may also be arranged by structural features, effect or product, and industry or use.

The USPC supports primarily one aspect of the process of patent examination, the determination of the statutory requirement for novelty. Examiners may request reclassification because of an overly large subclass, because technologically related patents are scattered throughout many subclasses, or whenever they feel that the *status quo* slows down the process of locating patents which anticipate the claims of applications under examination.

Within a class, subclasses follow a conventional arrangement designed primarily to aid examiners. Inventions which represent the most complex organization of components provided for in a class appear higher in the schedule, followed by less complex organizations. A modern schedule will frequently have a "combined" subclass about mid-way which acts as a miscellaneous collector of inventions not provided for higher in the schedule. Following this combined subclass, subclasses for the basic elements of the technology, or components or parts thereof, appear near the bottom of the schedule. This will usually be followed by another miscellaneous catch-all for the basic elements not provided for elsewhere.

The task of assigning a patent or patent application to a classification begins at the top of the hierarchy. The first main-line (i.e., top-level) subclass which describes any of the claimed technology becomes the home for that patent, and any other patent which happens to claim that technology. This is true even if other claims in the patent are described by later subclasses. Only if there is no mention of the technology in a main-line subclass in a claim does the finger move to the next main-line subclass. Once an appropriate main-line subclass is found, the same procedure is followed at the next lower level of the hierarchy to that main-line.

For example, in the hypothetical class of junk shown in Figure I, a combined chain and bar will not be placed in subclasses 1, 2, or 3 since these subclasses call for a bar in combination with other elements. Since it is not provided for elsewhere, a combined chain and bar will be placed in subclass 5 or in one of the subclasses subordinate to it. A patent with a claim to a simple bar will be placed in subclass 6.

1	COMBINED BAR, LINK, AND BALL
2	COMBINED BAR AND LINK
3	COMBINED BAR AND BALL
4	COMBINED LINK AND BALL
5	CHAIN
21	. With end fastener
22	. With flaccid cover
23	.. Removable
6	BAR
7	LINK
8	BALL
9	. Hollow
10	.. Perforated
11	.. Grooved
12	. Perforated
13	. Grooved
14	. Mineral
15	.. Metallic
16	... Aluminum
17	... Zinc
18	. Rubber
19	. Ivory
20	MISCELLANEOUS

Figure I
Hypothetical Class for Junk Yard Scrap

This process has important implications for patent searching which are rarely appreciated by the general public. A thorough search for patents concerned with bars need not be concerned with any subclass following subclass 6, but it must include all the subclasses prior to that point in the schedule, whether they mention bar explicitly or not. The prototype described further on incorporates design features which are intended to unobtrusively guide use of the schedules. Unlike the example in Figure I, which is typical of the published *Manual of Classification*, a screen in the prototype (Screen I) initially displays only the top level, thereby preventing users from considering deeper subclasses until they have selected a classification already on the screen.

CONTROLLED VOCABULARIES

Approximately 35 subject-specific databases available on DIALOG include U.S. patents in their coverage. Someone at each of the organizations responsible for producing those databases has taken the effort to assign descriptors, often from a controlled vocabulary, to the patents they index. Since current classifications for patents are readily available, it is possible to link descriptors to these classifications by way of patent numbers. This concept was tested on an experimental basis, and the result installed as part of the prototype.

32 PROCEEDINGS OF THE 1ST ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

For the prototype, *Energy Research Abstracts (DOE Energy)*, files 103 and 104, on DIALOG), a product of the Office of Scientific and Technical Information, U.S. Department of Energy (DOE), was chosen because of its relevance to other initiatives at DOE and the National Institute of Standards and Technology (NIST), formerly the National Bureau of Standards, because of its well developed and maintained controlled vocabulary, because it indexes a substantial number of patents, and because there were no intellectual property rights at stake.

The descriptors and patent numbers from a sample of solar energy related patent records were downloaded from DIALOG file 104. Each patent number was paired with every descriptor assigned to it. In another list, each patent number was paired with all of its USPC classifications, derived from CASSIS/CLSF³ Using a database management system, a relational join of the two lists resulted in the pairing of DOE descriptors with USPC classifications. The number of times a descriptor was paired with a classification was used to rank the classifications associated with each descriptor.

PROTOTYPE SYSTEM

A prototype hypertext *Manual of Classification*, using the KnowledgePro expert system shell, has been constructed. It is designed to assist inventors and the general public using the 68 Patent Depository Libraries (PDL) in determining the "field of search" for an invention (determining which subclasses have associated patents that might be relevant to the invention in hand). Working from this combined list of patents (patent numbers) in the field of search, the searcher looks at *Official Gazette* entries, ordered by patent number, eliminating patents of no interest from the list. Finally, the searcher obtains copies of the remaining patents, usually from microfilm at a PDL.

The opening screen of the prototype presents the table of contents of the printed *Manual of Classification* with some additional items not in the printed tool. We will focus on two of these items, one for searching the USPC hierarchy and the other for searching the Energy Database Subject Thesaurus.

Screen I shows the first view of a class, class 2, and its main-line subclasses, forming classifications 2/1, 2/2, 2/46, and so on. Following the principles used by examiners, a searcher will select the first main-line subclass which describes one of the claims for the invention under study. If the inventor claims a protective function, classification 2/2 (class 2, subclass 2) will accept the patent. Note that even though the invention protects elbows, the patent is assigned to GUARDS AND PROTECTORS, not the un-numbered subclass or HAND OR ARM COVERINGS.

3. One of three CD-ROM patent data files available to the public from the Patent and Trademark Office on a subscription basis.

2/	APPAREL
2/1	MISCELLANEOUS
&2/2	GUARDS AND PROTECTORS
	&BODY BRACES AND SUPPORTS
&2/46	GARMENT PROTECTORS
2/64	BURIAL GARMENTS
&2/65	FUR GARMENTS
&2/67	BATHING GARMENTS
&2/69	BODY GARMENTS
	&HAND OR ARM COVERINGS
&2/171	HEAD COVERINGS
	&NETHER GARMENTS
	&GARMENT SUPPORTERS AND RETAINERS
&2/243 R	GENERAL STRUCTURE AND METHODS
2/279	ADORNMENT HANDKERCHIEFS, GARMENT WORN

	DIGESTS

#2/DIG 1	Ventilated garment

F1 Help Space Cont.	F3 Select F4 View	F5 Evaluate F6 Display KB	F7 Edit F8 DOS	Pg 1 of 2 F10 Quit
------------------------	----------------------	------------------------------	-------------------	-----------------------

Screen I
The Main-Line Subclasses of Class 2

Selecting classification 2/2 reveals those subclasses subordinate to it (the & indicates there are such subclasses). These subclasses further divide GUARDS AND PROTECTORS. Indeed, as shown in Screen II, there is an explicit subclass for inventions which protect arms and hands. A search for apparel for arms and hands must include the main-line subclass HAND OR ARM COVERINGS and all subclasses superior to it in the class.

The relationship between superior and subordinate subclasses may be one of three types: genus-species, combination-subcombination, or a mixture of the two. Whenever a subclass number is displayed without the & prefix, selecting it will show the subclass definition. From this definition, cross-references may be followed to other, possibly more relevant classifications.

In contrast to the hierarchy of subclasses stands the hierarchy of terms in a thesaurus. The prototype contains a subset of the *Energy Database Subject Thesaurus* related to solar energy. Subject headings in this thesaurus may be browsed alphabetically, or string-searched for a word stem. Since these terms are short, the system restricts searching to a single-word stem.

2/	APPAREL
&2/2	GUARDS AND PROTECTORS
2/2	.Diving type
&2/2.1 R	.Penetration resistant
2/2.5	.For wearer's head
&2/410	.Hand or arm
&2/16	.Leg
&2/22	
&2/171	&HAND OR ARM COVERINGS
	HEAD COVERINGS
	&NETHER GARMENTS
	&GARMENT SUPPORTERS AND RETAINERS
&2/243 R	GENERAL STRUCTURE AND METHODS
2/279	ADORNMENT HANDKERCHIEFS, GARMENT WORN

	DIGESTS

#2/DIG 1	Ventilated garment

F1 Help	F3 Select	F5 Evaluate	F7 Edit	Pg 1 of 1
Space Cont.	F4 View	F6 Display KB	F8 DOS	F10 Quit

Screen II
Subclasses Subordinate to 2/2

Just as the classification schedules enforce a strict traversal of the USPC, the thesaurus subsystem requires that the user first view the thesaurus entry for a heading, exposing that heading's position in the hierarchy. From there, other thesaurus entries may be viewed, or a ranked list of the classes of the patents assigned that term by DOE indexers may be viewed. In Screen III a key-word search on *flow* produced an initial list of thesaurus entries. Following the narrow term — broad term relationships of the thesaurus led the searcher to the entry for HEAT TRANSFER FLUIDS. Satisfied that the proper term for the original *flow* concept has been located, a ranked list of the USPC classifications of the patents assigned that term by DOE indexers was displayed. From the list of USPC classifications one can view the corresponding definitions, and from the definitions, either obtain a list of patents related to the subject, or follow the search notes (cross-references) to other definitions.

SOLAR ENERGY THES	HEAT TRANSFER FLUIDS ->	HEAT TRANSFER FLUIDS	Rank	Classification
AIR FLOW ->	HEAT TRANSFER FLUIDS	DA April 1978	27	126/437
COMPRESSIBLE FLO	BT1 FLUIDS ->		25	126/450
COUNTERFLOW SYST	RT BLACK LIQUIDS		13	126/443
FLOW RATE ->	RT HEAT TRANSFER		12	126/422
FLOW REGULATORS	RT WORKING FLUID		12	126/435
FLUID FLOW ->			12	126/438
GAS FLOW ->	NT2 NATURAL GAS		10	126/433
LAMINAR FLOW ->	NT3 GAS CONDENS		10	126/449
LIQUID FLOW ->	NT3 LEASE CONDE		9	126/417
SOLAR COLLECTORS	NT3 LIQUEFIED P		8	126/426
SOLAR WATER HEAT	NT3 PLANT CONDE		8	126/434
	NT1 QUANTUM FLUID		7	126/419
	NT2 HELIUM II		7	126/440
	NT1 RESERVOIR FLU		7	126/442
	NT1 WORKING FLUID		7	126/444
	NT2 HYDRAULIC FL		6	126/425
	NT2 REFRIGERANTS		6	165/170
	RT FLUID FLOW ->		6	60/641.8

F1 Help Space Cont.	F3 Select F4 View	F5 Evaluate F6 Display KB	F7 Edit F8 DOS	Pg 1 of 5 F10 Quit
------------------------	----------------------	------------------------------	-------------------	-----------------------

Screen III

A Search in the DOE Energy Thesaurus

FUTURE

It is expected that four groups of expert and naive users will test the prototype: patent examiners, patent classifiers, the Energy Related Inventions Program staff at NIST, and end-users and professional staff at PDLs. This evaluation will be designed and conducted by students of the Worcester Polytechnic Institute as part of their degree requirements. The prototype will be judged a success provided it improves speed and accuracy of locating patents relevant to the subject of the search.

We plan to enhance the prototype by exploiting the expert system capabilities of KnowledgePro at two levels: one to aid in selection of a class, and the other to aid in selection of subclasses in accordance with a strict traversal of the USPC hierarchies as described earlier. The internal structure of each USPC class represents the state of classification practice at the time it was developed. Although the majority of classes are of the "modern" type described earlier, it may be necessary to have as many as five algorithms for subclass selection in order to accommodate classes arranged on the basis of structure, effect, or industry.

