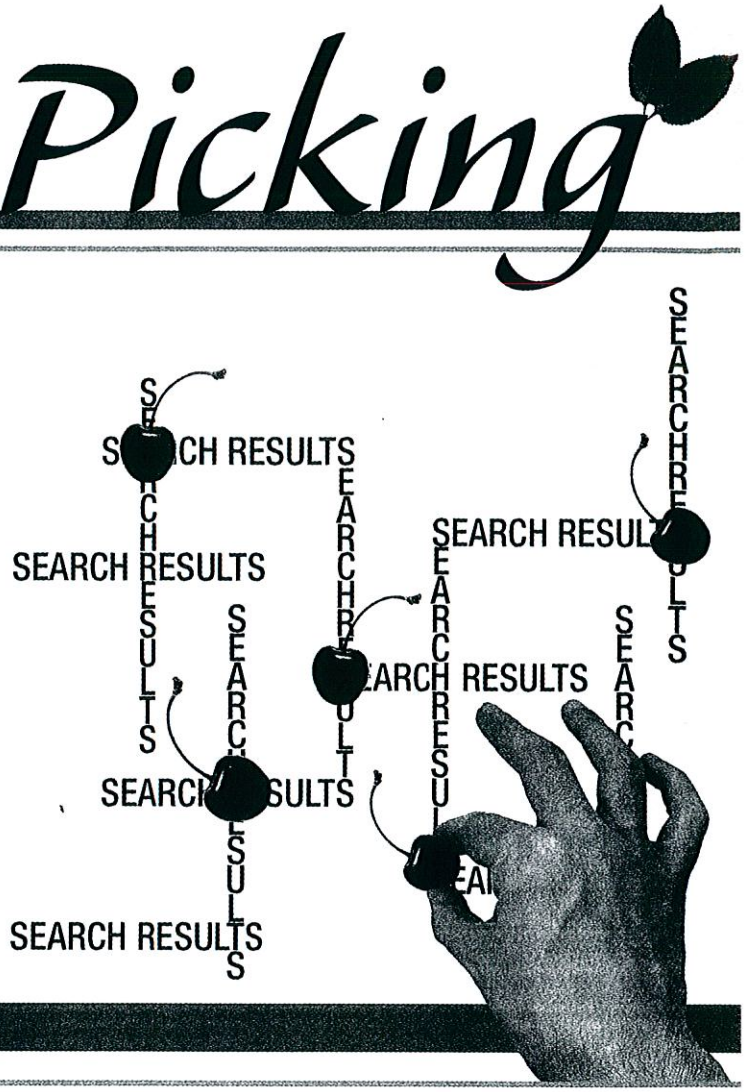
Cherry Picking FOR Patent Codes



by François Libmann

Classification codes can be hugely helpful when disambiguating a fuzzy topic or a poorly defined subject area. When it comes to patents, the U.S. Patent Classification System (www.uspto.gov/go/classification) is widely used when searching commercial patent databases to hone in on relevant patents. This is particularly critical for patent searches because the language in the body of the patent is frequently opaque and meant to reveal as little as possible to potential competitors.

Patent specialist and attorney Ron Kaminecki writes a monthly column, A Proximal and a Distal Tip, for Dialog's *Chronolog* newsletter. It is devoted to search tips for Dialog's intellectual property databases. In November 2009, he presented a methodology he called "cherry picking" (http://support.dialog.com/enewsletters/chronolog/200911). The strategy uses titles to identify the best hits. Briefly, you KEEP only the best four or five hits from a general search and then RANK the results by class code (CL). Essentially, you use the KEEP command to optimize usage of the RANK command. This reveals the codes actually used by patent examiners, not just those that seem to apply when you look at the system.

The KEEP command on Dialog creates a set called Set 0, to which you transfer selected items from any other set(s). Thus the command K3/2,4,6 will transfer items 2, 4, and 6 from set 3 into Set 0. The RANK command performs a statistical analysis of search results.

My colleagues and I found the methodology very intriguing. It could help clarify an ill-defined topic and identify a smaller entity, difficult to isolate from the larger picture. We decided to take Kaminecki's example and see how (or if) his method could be applied using Questel or STN.

THE STRATEGY ON DIALOG

We searched for U.S. patents dealing with contact lens cases, the problem being that "cases" has several meanings. The aim of the strategy was to identify the most relevant code in the U.S. Patent Classification system so it could be used in other patent databases.

We conducted the strategy in File 340, CLAIMS/US Patents, which contains abstracts for more than 4.6 million U.S. patents issued by the U.S. Patent and Trademark Office (USPTO) since 1950. See the Dialog Bluesheet for a full description (http://library.dialog.com/bluesheets/html/bl0340.html).

The full strategy, with some explanations and annotations, is shown on page 39.

File 340:CLAIMS(R)/US Patent 1950-2010/Apr 06
(c) 2010 IFI/CLAIMS(R)

?_S CONTACT()LENS? (Selects patents in which contact and lens appear in strict adjacency; the ? is the truncation symbol)

806861 CONTACT
138727 LENS?
S1 5664 CONTACT()LENS?

?_S CASE OR STORAGE OR CONTAINER
206219 CASE
497618 STORAGE
232665 CONTAINER
S2 874065 CASE OR STORAGE OR CONTAINER

?_S S1(15N)S2
5664 S1
874065 S2
S3 559 S1(15N)S2 (15N means that at least 1 element of set 1 is within 15 words maximum from at least 1 element of set 2 in any order.)

?_T3/6/1-20 (Displays the titles of the first 20 documents)

3/6/1
12401079 2010-0072082
M/CONTACT LENS STORAGE CASE

3/6/2
12374570 2010-0045571
E/BEAM-SCAN DISPLAY APPARATUS, DISPLAY METHOD, AND VEHICLE

3/6/3
12368619 2010-0039620
M/CONTACT LENS AND METHODS OF MANUFACTURE AND FITTING SUCH LENSES AND
COMPUTER PROGRAM PRODUCT

(...)

3/6/19
12185118 2009-0185135
M/Real image forming eye examination lens utilizing two reflecting surfaces
providing upright image

3/6/20
12179323 2009-0179340 2009-0041328
C/Lens molds with coating; Biblio Data Only
?_K3/1,5,10,12,13,16,17 (7 relevant documents are copied to Set 0)
S0 7 3/1,5,10,12,13,16,17

?_RANK cl s0
(Identifies the classification codes most frequently used in the seven patent records.)

DIALOG RANK Results

RANK: S0/1-7 Field: CL= File(s): 340
(Rank fields found in 7 records — 6 unique terms) Page 1 of 1
RANK No. Items Term

1 6 206005100
2 1 053431000
3 1 206205000
4 1 220345300
5 1 422024000
6 1 514276000
Enter desired option(s) or enter RANK number(s) to save terms.
?_ 1
RANK numbers saved: 1

?_EXIT

Temp SearchSave "TD314" stored
Creating temporary SearchSave ... TD314
Enter EXS to execute the SearchSave
(For comparison purposes, we also RANKed the whole set 3)

?_RANK cl S3
Started processing RANK
Processing 100 of 559 records
(...)
Processing 500 of 559 records
Completed Ranking 559 records
DIALOG RANK Results

RANK: S3/1-559 Field: CL= File(s): 340
(Rank fields found in 559 records — 792 unique terms)
Page 1 of 99
RANK No. Items Term

1 240 206005100
2 123 134901000
3 45 D03264000
4 41 206005000
5 41 422300000
6 31 422301000
7 26 510112000
8 25 134137000
(The ratio between the number of documents obtained with the first two codes is 1-6 in the first case and 1-2 in the second one. Furthermore, the following codes are different, which we expected because, logically, the result is less relevant. On the other hand, the cost of the Rank operator in this database, depending on the number of answers taken in account [it is not the case for all databases], is largely cheaper to use it on a limited number of relevant answers.)
The next step is to move the search to File 124, CLAIMS/REFERENCE:

?_B124

File 124:CLAIMS/REFERENCE 2001/2007Q1 (c) 2007 IFI/CLAIMS(R) PATENT SERVICES

Set Items Description

?_EXS
Executing TD314
S1 1 CL="206005100"
?_T1/9

1/9/1
DIALOG(R)File 124:CLAIMS/REFERENCE 2001/2007Q1
(c) 2007 IFI/CLAIMS(R) PATENT SERVICES. All rts. reserv.

00156536
U.S. Patent Manual of Classification
Class Title: Contact lens
Level: 03 Class Code: 206005100
Hierarchy Level Class/Subclass Title
206000000 01 (IPC B65D) SPECIAL RECEPTACLE OR PACKAGE
206005000 02 FOR EYEGLASS OR SPECTACLE
206005100 03 Contact lens
(We go back to File 340)

?_B340

File 340:CLAIMS(R)/US Patent 1950-2010/Apr 06
(c) 2010 IFI/CLAIMS(R)

Set Items Description

?_EXS
Executing TD314
S1 455 CL="206005100"
(The first 10 results are relevant.)

?_T1/6/1-10

1/6/1
12401079 2010-0072082
M/CONTACT LENS STORAGE CASE

1/6/2
12373246 2010-0044247
M/SAFETY KIT FOR CONTACT LENSES

1/6/3
12367263 2010-0038264
M/PACKAGING FOR OPTHALMIC LENS

1/6/4
12335455 2010-0006455
M/CONTACT LENS STORAGE CASE

1/6/5
12242406 2009-0242431
M/SCREW CAP PACKAGE FOR CONTACT LENS

(...)

1/6/8
12211906 2009-0211925
M/CONTACT LENS STORAGE AND CLEANING CASE

1/6/9
12200164 2009-0200182
M/PRE-FILLED CONTACT LENS CONTAINER

1/6/10
12173627 2009-0173643
M/Packaging Solutions


```
=> D BROWSE
:TRIAL 1-20

L9 ANSWER 1 OF 35 USPAT2 on STN
AN  2009:93500 USPAT2
TI  Contact lens case with date storing feature
INCL INCLM: 116/308.000
  INCLS: 116/312.000; 206/005.100; 206/459.100
NCL NCLM: 116/308.000
  NCLS: 116/312.000; 206/005.100; 206/459.100
IC  IPCI G09F0009-00 [I,A]
  IPCI-2 A45C0011-04 [I,A]; G09F0011-04 [I,A]; G09F0011-00 [I,C*]
  IPCR G09F0009-00 [I,C]; G09F0009-00 [I,A]

L9 ANSWER 2 OF 35 USPAT2 on STN
AN  2008:217558 USPAT2
TI  Method of cleaning contact lenses via sonication
INCL INCLM: 134/001.000
  INCLS: 134/901.000
NCL NCLM: 134/001.000
  NCLS: 134/901.000
IC  IPCI B08B0003-12 [I,A]
  IPCI-2 B08B0003-12 [I,A]
  IPCR B08B0003-12 [I,C]; B08B0003-12 [I,A]
(...)
```

Using the TAG command we select the relevant ones:

```
:TAG 1 3 4 8 13 14 17 18 19
ANSWER 1 TAGGED
ANSWER 3 TAGGED
ANSWER 4 TAGGED
ANSWER 8 TAGGED
ANSWER 13 TAGGED
ANSWER 14 TAGGED
ANSWER 17 TAGGED
ANSWER 18 TAGGED
ANSWER 19 TAGGED
:END
```

We use the SORT command to create a step containing the tagged references. We can choose any sort field.

```
=> SORT L9
SORT ENTIRE ANSWER SET? (Y)/N:N
ENTER ANSWER NUMBERS OR TAGGED (?):TAGGED
ENTER SORT FIELDS AND SORT DIRECTION (?):AU
PROCESSING COMPLETED FOR L9
L10   9 SORT L9 TAGGED AU

=> ANALYZE L10 1-
ENTER DISPLAY CODE (TI) OR ?:INCLM
L11   ANALYZE L10 1- INCLM :  4 TERMS

=> D ENTIRE
L11   ANALYZE L10 1- INCLM :  4 TERMS

TERM # # OCC # DOC % DOC INCLM

  1  6  6 66.67 206005100
  2  1  1 11.11 053329000
  3  1  1 11.11 116308000
  4  1  1 11.11 351160000H
********* END OF L11***]
```

TWO APPROACHES FOR QUESTEL

To execute a similar search on Questel, the strategy differs depending on the service used: Qweb with or without the Imagination interface or the Orbit.com platform, which is the new name of QPAT. (The Orbit.com name is a nod to the people who remember Orbit, one of the oldest U.S. hosts, which was acquired by Questel in 1994.)

With Qweb, we chose the USPAT file, which offers the full text of U.S. patents with classification codes. File 340 on Dialog has only title abstracts and claims (at least for the more recent patents).

The ability to search the entire patent has no specific bearing in this case and may reduce the precision. Thus, we limited the search to titles, abstracts, and claims.

In USPAT, the first step is /TI/AB/CLMS CONTACT W LENS+, which gives 3,878 answers (W is the adjacency operator and + is the unlimited truncation).

The step /TI,AB,CLMS Case OR storage OR container gives 569,144 documents.

Proximity operators on Questel are limited to a maximum of nine words between the search terms, in any order. On the other hand, S (Sentence) and P (Paragraph) operators are available.

In this example, we chose the D operator (the equivalent of Dialog's NEAR command, but limited to nine words). The command 1 D 2 gives 356 answers in step 3. That's fewer than in Dialog, which is logical because we tightened the distance between the two concepts.

To continue the comparison, the Dialog KEEP command has no strict equivalent in Questel, but to obtain a similar result, you can use the command FOCUS (FO), which allows you to see, one by one, the results of the previous step. In this database, the format is "Title, inventor, patentee, context around the search terms." For each document, it is possible to keep it or not. The command can be stopped at any time, and the kept documents are to be found in step 202.

The command MEMS/PCLO extracts the original U.S. classification code for the 11 documents we kept, and we issued the PRT MEMS command to see their frequency:

We delighted in discovering that cherry picking could be used to optimize classification selection. We think selecting codes (or descriptors) and discerning the most relevant should spark more interest among information professionals in the RANK command.

206005100 : 6
D0326400 : 4
351160000H : 1

Doing the same with the 356 documents of the step 3 we obtain the following:

206005100 : 19
D03264000 : 3
250461100 : 2

In this case, the results for the strategy are less spectacular, probably due to the proximity word count difference.

On the Orbit.com platform, we achieve a "nonexpress" search typing the strategy in the upper box: (contact W lens+) D (case OR storage OR container) limiting to the basic index and the claims. We then limited to U.S. patents.

Looking at the first 20 documents, we selected the relevant ones. With a click we send them to "my list." Then, we select them all and choose in the drop-down menu to analyze them by U.S. class codes.

The result is very clear. The code 206005100 is present in 83.33% of the patents and the code 116308000 in 16.66%.

A FORGOTTEN COMMAND IN STN

For STN, we first thought that such a search was not possible. However, looking more closely, the STN experts proposed using the command D BROWSE and its subcommands.

We searched the USPAT2 database, which is similar to USPAT from Questel, but with a shorter backfile (but there are other databases on STN with U.S. patents).

As on Questel, we limit the search to titles, abstracts, and claims.

The first step is S (contact () lens?)/TI,AB,CLM, where ( ) is a strict adjacency and ? is an unlimited truncation, just as you find with Dialog search syntax.

The result (L1) is 776 documents.

We then write S (case OR storage OR container)/TI, AB,CLM. Note that the command PLURAL (automatic plural) is set ON in our preferences.

The result (L2) is 156 273 documents.

S L1 (15A) L2 , (15A being the equivalent to 15N on Dialog) gives 35 documents. The fact that there are fewer documents is due to the smaller coverage of the USPAT2 database.

We then enter the command D BROWSE, which is a command slightly different from others since it opens a series of subcommands available only that way.

We visualize the first 20 more-recent documents with the command TRIAL 1-20 in the sidebar on page 40.

PHILOSOPHIZING ON SEARCH COMMANDS

We found this exercise interesting for several reasons.

First, we found it an intriguing learning experience to take the Dialog strategy of using the KEEP and RANK commands, ones that may not be all that well-known even to Dialog searchers, and transferring the cherry-picking technique to Questel and STN. We delighted in discovering that cherry picking could be used to optimize classification selection. We think selecting codes (or descriptors) and discerning the most relevant should spark more interest among information professionals in the RANK command.

Second, although we applied the cherry-picking method to patent databases, we believe it is not limited to this kind of database. If it can be used in any indexed database, it is then a general method.

We would like also to point out the fact that illustrating this methodology on three important classical hosts shows that for specific strategies, the tools can be different, but the end results are quite similar.

We were heartened that our experiment reaffirmed the importance of the unique advanced search features offered by these three hosts, competitive though they are.

All these possibilities available to info pros can also be considered as a way for them to show how their expertise with traditional search hosts add value to their research capabilities.

François Libmann *(flibmann@fla-consultants.com) is director, FLA Consultants. A version of this article, in French, appeared in the March 2010 issue of* BASES, *published by BASES PUBLICATIONS, a sister company of FLA Consultants, with the title "La cueillette des cerises."*

Comments? Email the editor (marydee@xmission.com).