Bulletin Search Interface

The **bulletin search interface** for the Meteorological Service of Canada http data server is available at http://dd.weather.gc.ca/cgi-bin/bulletin_search.pl.

The bulletin searches are conducted via URL query strings, i.e.

http://dd.weather.gc.ca/cgi-bin/bulletin_search.pl?
variable1=value1&variable2=value2

It can take a long time, many minutes, before the search script gives a result.

The search allows using regular expressions as well as what has been dubbed *the intuitive interface*. The pros, cons, and differences between the two systems will be discussed below followed by examples of how the intuitive interface can be employed via example query strings. For more information regarding perl regular expressions go to:

- http://perldoc.perl.org/perlre.html
- http://www.english.uga.edu/humcomp/perl/regex2a.html.

NOTE: There is one specific limitation of the search script regarding date searches. On occasion, bulletins will not be placed in the proper directory which can lead to them being excluded from the search. For example, A bulletin issued late on one day might be placed into the next day's directory which precludes it from being counted in the search. Similar situations can occur based on the hour of the bulletin which can also differ from the hour directory into which it is placed. Broadening your search to include adjacent days or hours can aid you in these special cases.

Contents

1 About

This document is version 1.2 on date 2014-09-24

The latest version can be downloaded from:
http://dd.weather.gc.ca/bulletins/doc/

For more information about the bulletins, see:
http://dd.weather.gc.ca/bulletins/doc/README_bulletins.txt

2 Example Query Strings

Below are some example query strings to demonstrate the flexibility and possibilities of the system. These examples are for use with the intuitive interface (i.e. the value of the parameter 'search' is by default set to 'intuit')

- Find all bulletins for Feb. 26, 2006 from product FP and issuing office CWAO:

http://dd.weather.gc.ca/cgi-bin/bulletin_search.pl?year=2006&month=2&day=26&product=fp&issuer=cwao

- Find all bulletins for today from product SA with the default issuing office with location CN:

http://dd.weather.gc.ca/cgi-bin/bulletin_search.pl?product=sa&location=cn

- Find all bulletins for today from products SA, FP, and WW:

http://dd.weather.gc.ca/cgi-bin/bulletin_search.pl?product=sa,fp,ww

- Find all bulletins for today not from products SA, FP, and WW

http://dd.weather.gc.ca/cgi-bin/bulletin_search.pl?product=!sa,fp,ww

- Find all bulletins for today from all issuing offices which are not CWAO or DRRN with a station code of CYYZ:

http://dd.weather.gc.ca/cgi-bin/bulletin_search.pl?issuer=!cwao,drrn&station=cyyz

- Find all bulletins for today from product SA before noon with issuing office EGGY:

http://dd.weather.gc.ca/cgi-bin/bulletin_search.pl?product=sa&issuer=eggy&hour=<11

- Find all bulletins for today from any product with a correction code

http://dd.weather.gc.ca/cgi-bin/bulletin_search.pl?correction=!none

- Find all bulletins for today from any product with a correction code but no station code

http://dd.weather.gc.ca/cgi-bin/bulletin_search.pl?correction=!none&station=none

- Find all bulletins for today from any product with issuing office CWAO:

http://dd.weather.gc.ca/cgi-bin/bulletin_search.pl?issuer=cwao

- Find all bulletins for today from product product SA with the default issuing office and a header code greater than 50:

http://dd.weather.gc.ca/cgi-bin/bulletin_search.pl?product=sa&header=>51

- Find all bulletins for today from products which end with either A or Z which have a station code which does not begin with C and a header code between 35 and 72 and have occurred after noon:

http://dd.weather.gc.ca/cgi-bin/bulletin_search.pl?product=*a,*z&station=!c*&header=

- Find all bulletins for today with stations which start with C and end with R and start with V

http://dd.weather.gc.ca/cgi-bin/bulletin_search.pl?station=C*R,V*

- Find all bulletins for today with issuing which start with C and end with L and do not end with Z

http://dd.weather.gc.ca/cgi-bin/bulletin_search.pl?issuer=!C*L,*Z

3 Different Searchable Properties of Bulletins

There are many searchable fields available through the script with various different characteristics. Some are numeric, some are alphanumeric, some are mandatory in a bulletin and some are optional. Optional fields can have existence and nonexistence checks, alphanumeric fields can have substring checks, numeric fields can have numeric range checks and all fields can negate and group searches together with an exclamation mark and comma separated values. Below is a listing of the fields (with the variable name to be used in the query string in parentheses) including some examples of their values as well as their default values.

3.1 Universal alphanumeric fields

Variable	Description	Default Value
product	Product Code. Some example codes include SA, AA, etc.	Any valid product
location	Location Code. Some example codes include CN, TS, etc.	Any valid location
issuer	Issuing Office Code. Some examples include CWAO, CYPL, etc.	If set:The default issuing office for the product If not set: Any valid issuing office

3.2 Universal numeric fields

Variable	Description	Default Value
year	The year the bulletin was made	Current year
month	The month the bulletin was made	Current month
day	The day the bulletin was made	Current day
hour	The hour the bulletin was made	Any valid hour
minute	The minute the bulletin was made	Any valid minute

3.3 Optional alphanumeric fields

Variable	Description	Default Value
station	Station code. Some examples include CYYG,	Any valid station (including no station)

	WZB, CWKW, etc.	
correction	Correction code. Some examples include CCA, AAA, etc.	Any valid correction code (including no correction code)

3.4 Optional numeric fields

Variable	Description	Default Value
header	Header code. Some example codes include 15, 31, etc.	Any valid header value (including no header code)

4 Additional Query String Variables

There are additional query string variables which allow you to select from the various outputting methods available within the script.

4.1 Output Format (*format*)

Description: Raw Text outputs a newline separated text file of bulletin content or file locations depending on the value of the 'output' parameter. HTML output will provide an unordered list of file locations or bulletin content depending on the value of the 'output' parameter.
Available options: Raw Text (*raw*) and HTML List (*html*).
Default value: Raw Text (*raw*).

4.2 Output Type (*output*)

Description: Provides the fully qualified location of the bulletin when set to 'name' or the actual contents of the bulletin when set to *content*.
Available options: Bulletin Location (*name*) and Bulletin Content (*content*).
Default value: Bulletin Content (*content*).

4.3 Search Technique (*search*)

Description: When set to *regex* no guessing is made on your behalf under the assumption you have inputted a regular expression for yourself. *intuit* prepares your search parameters for you as described below.
Available options: Intuitive Interface (*intuit*) and Regular Expressions (*regex*).
Default value: Intuitive Interface (*intuit*).

5 Intuitive vs. Regular Expression Interface

The syntax of the intuitive interface is an effort to simplify regular expressions for use by people unfamiliar with the intricacies of regular expressions. This creates a more intuitive interpretation of the search fields by the user but can confuse those experienced with regular expressions. For example, while in regex to denote zero or more of any characters you must use ".*" with the decimal symbolizing any character and the asterisk denoting zero or more of the character preceding it (in this instance a special character meaning any character), the default interface uses "*" as meaning zero or

more of the *appropriate characters* for the field.

That is, a field which accepts only the letters of the alphabet will be automatically restricted to only looking for letters of the alphabet in the default interface; it will be similarly limited by the acceptable length of the field. If, for instance, you wanted to find any bulletins with a product code which begins with the letter "*S*" you would set the query string variable as "*product=S**" and the default interface would convert this into a safe and characterbounded regular expression which can be used in the search. Because of this simplification searches looking for repeated characters are not available in the default interface; if such functionality (where "*S**" means zero or more consecutive S's) is required use regular expressions.

The intuitive interface is enabled by default however it can be overwritten and perl regular expressions can be used if you are more familiar with that syntax. More information about Perl regular expressions can be found at both http://perldoc.perl.org/perlre.html and http://www.english.uga.edu/humcomp/perl/regex2a.html. Perl regular expressions are activated by adding a query string variable "*search*" with a value of "*regex*". That is, "*search=regex*". to explicitly use the intuitive interface you may set the search variable to "*intuit*" (eg "*search=intuit*").

Substring (including prefix and suffix) searches are handled by three special characters borrowed from regular expressions though their meanings have been slightly adjusted. As evidenced above, the asterisk has incorporated a character set upon which to match; the same is true of the decimal and the question mark. The decimal has retained it's regex meaning as any single valid character allowing finer control over the length of the matched values. If you wanted to ensure that the issuing code of your bulletins was five characters long (rather than the typical four character length) you could set the query string variable as "*issuer=.....*" making each matched issuing office five characters long but still bound by the field's possible values (in this instance both letters and numbers).

One method of search where the intuitive interface exceeds regex is numeric comparison. Regular expressions are characterbased searches and so numeric comparison is prohibitively complicated; the intuitive interface allows you to search numeric fields within (and without) ranges as well as greater than and less than comparisons. One limitation is that it only supports the values from 0 to 99; however, given that these are the only possible values the numeric fields that can be searched in bulletins, it is an acceptable compromise.

Additionally, you can search for the opposite of any of these in the intuitive interface simply by placing an exclamation mark "!" which negates the current search. Any of the previously discussed methods can be negated through the exclamation mark. Negation of phrases (rather than single characters) in regular expressions is not as simple and its syntax can be problematic.

Below are some examples showing the distinct advantages to the intuitive interface.

5.1 Numeric fields do not require zeropadding

5.1.1 Intuitive Interface

```
?hour=2
```

5.1.2 Regular Expressions

```
?hour=02
```

5.2 Easier to do "or" searches

5.2.1 Intuitive Interface
```
?hour=0,6,12,18
```

5.2.2 Regular Expressions
```
?hour=(00|06|12|18)
```

5.3 Easier searching between ranges

5.3.1 Intuitive Interface
```
?header=1352
```

5.3.2 Regular Expressions
```
?header=(1[39]|[24][09]|5[02])
```

5.4 Easier to use "greater than/less than" operator

5.4.1 Intuitive Interface
```
?hour=>13
```

5.4.2 Regular Expressions
```
?hour=(1[39]|2[03])
```

5.5 Easier to do negation

5.5.1 Intuitive Interface
```
?issuer=!CYUL
```

5.5.2 Regular Expressions
```
?issuer=(?!CYUL).*
```

5.6 Combining negation with ranges is much easier

5.6.1 Intuitive Interface
```
?header=!1520
```

5.6.2 Regular Expressions

```
?header=(?!1[59]|20).*
```

5.7 Combining negation with multiple values is easier

5.7.1 Intuitive Interface

```
?product=!SA,AA,WW,WF
```

5.7.2 Regular Expressions

```
?product=(?!SA|AA|WW|WF).*
```