

## Manual del uso de CRAT

### **Lenguaje de Búsqueda**

En la siguiente sección se va a describir el lenguaje que se utiliza para llevar a cabo los búsquedas dentro de la herramienta. Para distinguir más claramente cual es el formato que se ha de introducir, se van a poner los ejemplos en negrita con el texto exacto que debería de utilizarse.

### **Buscando palabras**

Para realizar la búsqueda de una palabra dentro de los documentos/corpus que se tienen seleccionados se ha de introducir la siguiente secuencia **[palabra]**

En resumen, hay que poner la palabra entre corchetes y esta será la palabra a buscar, concretamente el ejemplo anterior buscaría en el corpus seleccionado, las frases que contienen la palabra *palabra*.

### **Buscado palabras con wildcard**

El uso de wildcard es muy común en los buscadores más potentes pues habilitan un sin fin de combinaciones de búsqueda. Un wildcard es indicar “comodines” o carencia de información dentro de la palabra que se va a buscar.

Los wildcard que permite la aplicación son los símbolos: ? \*

El símbolo \* hace referente a 0,1 o n letras mientras que el símbolo ? hace referencia a 0 o 1 letra.

Para aclarar su significado veamos un ejemplo: El siguiente ejemplo buscaría las frases que tuvieran al menos una palabra que empezará por “go- “ **[go\*]** el resultado sería por ejemplo *goes, go, ....*

Sin embargo la siguiente búsqueda **[go?]** nos proporcionaría también el resultado de *go* e incluso el de *god* pero nunca el de *goes* por que? Por que en esta palabra después de go- hay dos letras y no solo 0 o 1 que es lo que indica esta búsqueda. Sin embargo también se podría escribir la siguiente búsqueda **[go??]**

Estos wildcard se pueden usar combinados, por ejemplo **[\*aripo?a\*]** sería una expresión permitida

## Relaciones simples

Los operadores booleanos nos permite relacionar varios términos de búsqueda mediante una relación lógica. Dentro de la aplicación se permiten los operadores booleanos: *y, o, no*

Por ejemplo la siguiente búsqueda produciría como resultado las frases que contienen o bien la palabra *go*, o bien la palabra *went*. **[go]{OR}[went]** (Recordemos que {OR} no hace falta escribirlo, solo hay que seleccionar en el cuadro de dialogo correspondiente)

En lógica clásica, se suele decir que el {OR} es una suma, por que si nos fijamos, esta búsqueda es la suma de los resultados producidos por [go] y los producidos por [went]



Merece la pena comentar en este punto, la a veces falta utilización del operador lógico *y*. Comencemos diciendo lo que NO es este operador con un ejemplo: **[go]{AND}[went]** Esta búsqueda NO significa que se obtengan las frases CON GO Y CON WENT y TAMPOCO significa que se van a obtener las frases que AL MISMO TIEMPO CONTENGAN LAS PALABRAS GO Y WENT (EN LA MISMA FRASE).

El significado real sería el de obtener las frases en las que una de sus palabras contenga a la misma vez la forma *go* y la forma *went* y esto es IMPOSIBLE (una palabra hasta ahora solo tiene una forma de escribirse). Por tanto esta búsqueda NO devolvería ningún resultado.

Y entonces si es así, por que se incluye? Muy sencillo si que sería interesante la siguiente búsqueda: **[go\*]{AND}[\*ing]** Esta búsqueda devolvería las frases que contienen una palabra que a la vez (la misma palabra) comienza por *go-* y acaba en *-ing*.

Por último, y operado lógico NO significa “todo lo que no sea”. Por ejemplo la búsqueda **{NO}[go]** devolvería todas las frases del corpus excepto las que contiene la palabra *go* en su interior.

## Relaciones complejas

En las últimas versiones, se ha introducido dentro de la aplicación la aceptación de un operador mucho más potente que los operadores lógicos, este es el operador de sucesión. El operador de sucesión viene a aportar el significado “seguido de...”. y se representa por +.

El siguiente ejemplo devolvería las frases que contienen la palabra *go* seguida inmediatamente de la palabra *to*. **[go]+[to]**

## Parentesis

Dentro del lenguaje se pueden utilizar paréntesis para agrupar términos y para cambiar las prioridades de los operadores lógicos. El orden de prioridad de estos operadores es el siguiente: no > y > o.

Es decir que la siguiente búsqueda: `{NOT}{go}{AND}[ t=v+++++]{OR}[ t=vi+++++]` podría no quedar nada clara para una persona no experimentada, sin embargo se podría escribir de la siguiente forma:

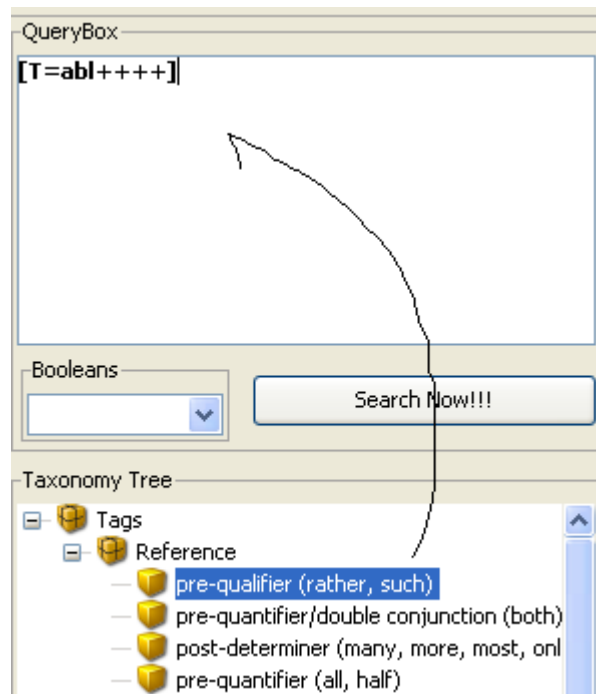
`{NOT} ( [go] {AND} ( [t=v+++++]{OR}[t=vi+++++] ) ) )`

Con los paréntesis tenemos que se podría leer la búsqueda como las frases que no tengan una palabra go estando esta etiquetada o bien con la etiqueta v+++++ o bien con la etiqueta vi+++++

Si no hubiéramos puesto los paréntesis no hubiera sido el mismo pues el AND se hubiera ejecutado antes del OR por prioridad y entonces habría resultado las frases que o bien no tienen la palabra go con la etiqueta v+++++ o bien no tienen la etiqueta vi+++++

## Busqueda de tag

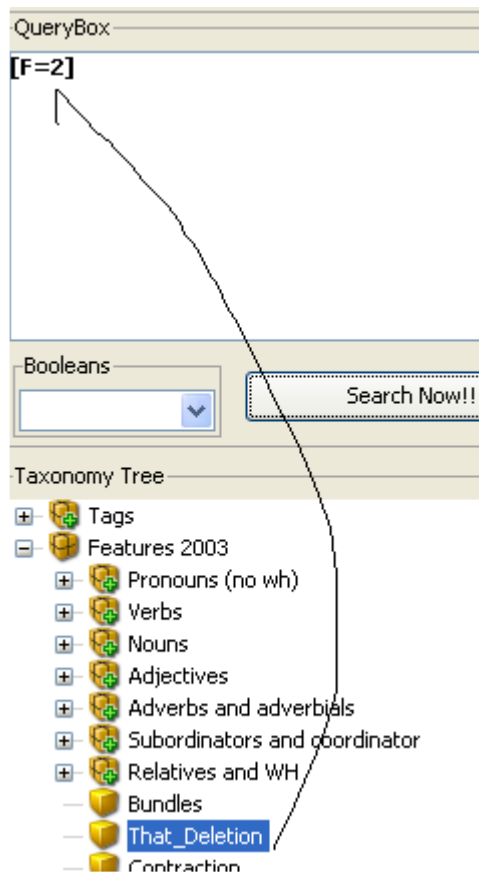
Para realizar la búsqueda de tag simplemente hay que arrastrarla dentro de la caja de búsqueda.



En este caso se buscarían las frases que contiene la etiqueta asociada a “pre-qualifier”. Sin embargo se podrían buscar las frases que tiene un prequalificador seguido de una palabra que empiece por a- en cuyo caso la búsqueda sería: `[T=abl+++++][a*]`

### Búsqueda de Features

Análogamente, para buscar una feature solo hay que arrastrarla a la caja. (ver imagen)



Esta búsqueda buscaría las frases que tiene la característica de *That Deletion* pero aun mejor, estas se pueden cambiar con cualquier otra produciendo cosas muy curiosas. Por ejemplo: **[F=2]+[you]** produciría las frases que tiene la característica de *That Deletion* seguida de la palabra you. Por ejemplo:

	Main Word	
m ... it's good ... I think so I	mean	you have free time at nights 1
es 1 ... and tell me what you	think	you see it 's kind of a story in p
... ah 2 ... and yes I do n't	know	you have also there the Turkist
2 ... in four day in two days I	think	you 've seen all ... 1 ... mhm 2 t
m 2 ... but you need n't ha I	mean	you need n't have very good m
out er ... good I'm going to	show	you some ... some photos not pl
ah 1 ... so what what do you	think	you 've learnt ... what do you thi
u 've learnt ... what do you	think	you 've learnt from 2 ... I learnt
ood okay em I'm going to	show	you 1 ... some pictures now ... Pili
... now Alberto I'm going to	show	you ... four pictures ... and ... the
u see ... what story do you	think	you see here ... so you 've got t
od Alberto I'm going to ...	show	you this piece of paper ... and w
panish 1 ... yeah ... do you	think	you would like to go back ... to
1 ... 1 yeah 2 oh yeah 2 I	understand	you ... 1 ... yeah ... I'm from Cali
od okay now I'm going to	show	you ... a series of pictures so th
uld you go again ... do you	think	you will go again or 2 2 huy it
:3 1 uhuh uhuh.. wow , I	think	you made the point ... good.. ok
l... okay , now I'm going to	show	you here is a picture... so it say:
... if you say you are this I	believe	you but... you know that 's not tr
ly that ... 1 Uhuh , do you	think	you might study that ? 2 No ... I
js see ... now I'm going to	show	you a series of pictures ... and t
Uhuh uhuh , and do you	think	you 'd like to do theatre in Murc
... Right , now I'm going to	show	you a series of pictures , so the
t was the atmosphere you	know	you know people was er very ni
id okay Erika I'm going to	show	you ... last thing this piece of pa

### ***Búsqueda de Dimensiones***

Esta búsqueda se realiza de la misma forma que con las *features* sin embargo ahora el significado sería que se obtienen las frases que contienen alguna de las palabras que están asociadas a la dimensión que se desea buscar.

### ***Todo a la vez***

Decir que se pueden hacer tantas mezclas de operadores y búsquedas como se puedan imaginar sin ningún tipo de limitación.

Por ejemplo se podría hacer la siguiente pedazo de búsqueda:

```
[D88=1]{OR}[GO]{OR}[F=11]({we*}{AND}{*nt}){OR}[T=abl++++]
```

### ***Lo que no se puede hacer***

La siguiente búsqueda NO esta permitida: **[go][went]** por que? Por que no se ha especificada que relación hay entre *go* y *went*, van seguidos? Se dan a la vez? Son las frases de *go* y las de *went*?

## Fichero de Configuración

La configuración de la aplicación se realiza mediante un fichero de texto. Este fichero se llama *config.xml* y debe de estar en la carpeta que contiene el programa. Por tanto para cambiar la configuración de la aplicación solo hay que editar este fichero con la configuración apropiada.

Dentro del fichero se pueden encontrar lineas con el siguiente patrón:

```
<entry key="NUM_PHRASES_IN_TABLE">50</entry>
```

Esto significa que el parámetro que se puede configurar se llama “NUM\_PHRASES\_IN\_TABLE” y que el valor para este parámetro es 50.

Veamos ahora que podemos configurar:

- El numero de frases que van a aparecer en la tabla de resultados es configurable mediante la propiedad *NUM\_PHRASES\_IN\_TABLE*. Su valor es un numero entre 1 e 99.999.999
- Se puede configurar el abanico que se muestra de una frase cuando no se puede identificar donde comienza/termina. Se realiza mediante la propiedad *NUM\_WORDS\_IN\_PHRASE\_CONTEXT* que toma un numero entre 1 e 99.999.999 y por ejemplo si toma el valor de 20 significa 20 palabras a la derecha y 20 a la izquierda.
- Se puede configurar si la aplicación va a realizar la identificación del inicio/final de las frases cuando sea posible (este proceso puede causar una bajada de rendimiento en las búsquedas). Se realiza mediante la propiedad *ALLOW\_INTELLIGENT\_CUT\_WHEN\_CAN\_BE* que puede tomar los valores ON y OFF (ojo que es sensible a mayusculas).
- Se puede configurar si se desean reemplazar cierto elementos que se encuentran el corpus de Biber por otro más estético o no. (en caso de realizarse este cambio se verán en los resultado símbolos de “ ... “). Este proceso se configura mediante *REPLACE\_RENDERER* que puede tomar los valores ON y OFF.
- Se puede configurar si el usuario desea que se centren las frases respecto a la palabra que ha causado la busqueda o no (en caso de hace el centrado el rendimiento decaerá un poco pero no tanto como el corte inteligente de frases). Se realiza mediante la propiedad *ALLOW\_CENTER\_PHRASES* que puede tomar los valores ON y OFF.
- También se puede configurar si se desea información de debug en la pantalla de consola o no, esta información principalmente mide los tiempos en ms que tardan las búsquedas en realizarse. Esto se puede configurar con la propiedad *DEBUG* que puede valer ON y OFF.
- Las propiedades que comienzan por *DB\_....* y la propiedad *MySQL\_SERVICE\_NAME* son propiedades para configurar la base de datos y solo se necesitarían cambiarse en el caso en el que se quisiera utilizar otra base de datos que se tenga instalada en el equipo para hacer correr la aplicación.