

Ricerca per Range

Il concetto di ricerca per range di termini nasce per consentire di esprimere intervalli di chiavi da selezionare da un vocabolario da utilizzare tutte insieme come se si trattasse di un'unica chiave.

Per tale ragione si è concepita una sintassi che consentisse di esprimere queste condizioni. La sintassi, in seguito, è gradualmente evoluta comprendendo anche modalità che con gli intervalli di termini avevano ben poco a che fare ma che si prestavano per esprimere il tutto come se fosse un'unica chiave.

Al fine di meglio comprendere la finalità di questa forma d'estensione pensiamo all'adiacenza tra questo intervallo di termini ed un altro termine. Solo trattandolo come se fosse un termine singolo l'adiacenza ha ragionevole speranza di essere portata a termine e di riuscire in questo compito in tempi ragionevoli.

Nota: In tutti gli esempi successivi, quando si parla di estremo o termine, ci si riferisce ad una chiave utilizzata per compiere la ricerca. La risoluzione del range assumerà che la prima chiave abbia inizio subito dopo il carattere **{** e fine prima dell'identificatore di range e così per la seconda chiave, dal primo carattere che segue l'identificatore di range sino al carattere **}**. Ogni chiave può essere racchiusa tra doppi apici e deve essere racchiusa tra doppi apici quando il suo contenuto possa risultare ambiguo nei confronti di un identificatore di range. Non vale il ragionamento secondo il quale il primo identificatore incontrato viene preso in esame e gli altri trattati come semplici caratteri.

Sintassi di base

La sintassi di base è estremamente semplice. Il termine viene rappresentato da qualcosa che va racchiuso tra parentesi graffe **{ }** all'interno delle quali l'espressione cambia a seconda dei simboli e dei modificatori utilizzati.

Avremo quindi una condizione del tipo:

[campo]={espressione}

Le sintassi classiche

Quando parliamo di sintassi classica ci riferiamo al fatto che la sintassi rappresenti realmente un intervallo di termini.

L'intervallo può essere chiuso o aperto.

L'intervallo si dice chiuso quando entrambe gli estremi sono stati esplicitati e si dice invece aperto quando solo uno dei due estremi viene esplicitato.

La sintassi che ne deriva, per il range chiuso, è la seguente

[campo]={estremo inferiore|identificatore tipo range|estremo superiore}

Analogamente avremo due possibili ranges aperti

[campo]={estremo inferiore|identificatore tipo range}

[campo]={identificatore tipo range|estremo superiore}

Come detto gli estremi non sono entrambe obbligatori¹⁾ ed il significato del range cambia in virtù del suo identificatore.

Vediamo come

Identificatore tipo Range	Descrizione
	Range inclusivo. Entrambe gli estremi indicati sono inclusi nei termini da ricercare. Si può esprimere un range aperto ed in tal caso il vocabolario viene investigato in tale direzione in modo completo.
^	Range esclusivo. Entrambe gli estremi indicati sono esclusi dai termini da ricercare. Esso si estende quindi dal primo termine superiore all'estremo inferiore sino all'ultimo termine inferiore l'estremo superiore. Anche in questo caso il range può essere aperto e palesemente l'estensione nella direzione dell'estremo non espresso ha luogo comunque in modo completo.
!	Range negativo. I termini espressi come estremi vengono esclusi dalla selezione così come tutti i termini che ricadono tra di essi. In sostanza è una sorta di negazione del range inclusivo. Usarlo senza esprimere entrambe gli estremi non ha molto senso ²⁾ ed anche se si intende escludere un solo termine, il range deve indicare tale termine tanto come estremo inferiore quanto come estremo superiore.
<	Range esclusivo a sinistra. Da utilizzarsi, ovviamente, indicando entrambe gli estremi ³⁾ . In questo caso l'estremo di sinistra ovvero l'estremo inferiore viene escluso dal range mentre quello di destra, ovvero quello superiore, viene incluso nel range.
>	Range esclusivo a destra. E' l'esatto opposto del caso precedente. Da utilizzarsi, ovviamente, indicando entrambe gli estremi ⁴⁾ . In questo caso l'estremo di sinistra ovvero l'estremo inferiore viene incluso nel range mentre quello di destra, ovvero quello superiore, viene escluso dal range.

Vana è ogni ricerca in cui gli estremi siano mal espressi.

Esempi Il più classico degli esempi si ha con ranges inclusivi ed esclusivi, aperti e chiusi.

La ricerca...

[Numero]={10|100}

...prevede la selezione di tutti documenti in cui il canale Numero assume un valore compreso tra 10 e 100. Tale ricerca corrisponde formalmente a...



```
[Numero]>=10 AND [Numero]<=100
```

Sull'onda di questo esempio avremo...

```
[Numero]={|10}
```

...per cercare documenti con un qualsiasi numero minore o uguale a 10, identico quindi all'espressione...

```
[Numero]<=10
```

Identico ragionamento va fatto per la ricerca per Maggiore o Uguale.

Se, per altro, la ricerca viene espressa come...

```
[Numero]={20^}
```

...la ricerca richiede ogni valore superiore a 20, quindi 20 escluso, mentre non avendo indicato l'estremo superiore non si esclude alcun valore in tale direzione. La ricerca è quindi equivalente a...

```
[Numero]>20
```

Poiché gli estremi vengono esclusi se si fa uso dell'identificatore di range ^, un campo contenente solo numero interi non darebbe alcun esito per la ricerca...

```
[Numero]={14^15}
```

...in quanto gli estremi verrebbero entrambe esclusi.

Allo stesso modo la ricerca...

```
[Numero]={10|10}
```

...è un modo assolutamente arzigogolato per dire...

```
[Numero]=10
```

Il comportamento dei range nei quali l'esclusione dei termini riguarda solo uno dei due estremi è facilmente derivabile dagli esempi indicati.

Vediamo ora la ricerca...

```
[Anno]={1990!1999}
```

...essa troverà i documenti nei quali appare un anno che non rientra tra il 1990 ed il 1999 entrambe compresi. Si noti quindi che la forma...

```
[Anno]={1990!1990}
```

...che indica che non si vogliono i documenti dove appare un valore diverso da 1990 corrisponde alla forma...

```
[Anno]<>1990
```

...e rappresenta il modo in cui il server, dalla versione indicata, traduce questo tipo di espressione.

Attenzione: Apparentemente l'indicazione di '[Anno]<>1990' ovvero di '[Anno]={1990!1990}' non è del tutto equivalente all'espressione 'NOT ([Anno]=1990)'.

L'espressione positiva con negazione, infatti, non avvalendosi di un range, non fa uso dei termini presenti in un vocabolario e quindi selezionerà anche tutti quei documenti in cui non c'è alcun valore per il campo 'Anno'⁵⁾.

Facendo invece uso del range, ci si avvale dei soli termini presenti nel vocabolario quindi, salvo casi particolari, si selezioneranno solo ed esclusivamente documenti in cui il campo 'Anno' contiene un valore valido ma tale valore non è quello indicato.

Le sintassi speciali

Si intendono sintassi speciali quelle che sono state derivate dalla sintassi di base dei range per assolvere ad altre finalità pur avvalendosi della stessa logica e quindi producendo come risultato finale un'estensione di uno o più termini da usare dinamicamente con altre condizioni di ricerca come, ad esempio, un'adiacenza.

Ci sono solo due casi attualmente sviluppati: la ricerca per lista di termini e la ricerca per proiezione.

Lista di termini

La ricerca per **lista di termini** prevede una sintassi in cui l'identificatore di tipo di range è la virgola (,). Essa viene usata per separare termini da porre tutti in OR tra di loro. Avremo quindi una forma come la seguente

```
[campo]={term1,term2,term3,term4,...termN}
```

Nota: In corrispondenza dell'emissione del server 22.1.3.5 tale sintassi è stata ampliata consentendo anche che l'elenco dei termini sia composto, di fatto, da un termine solo come nell'esempio che segue

`[campo]={term1}`

Tale sintassi non ha molto senso⁶⁾ ed equivale pressoché in tutto alla sintassi

`[campo]=term1`

Proiezione (o Join)

La ricerca per **proiezione** è, in sostanza, quella che nell'ambiente Highway/eXt raWay è sempre stata denominata direttamente *proiezione*.

Evocare una *proiezione* è un'attività non particolarmente semplice e sicuramente non si riesce ad esprimerla in una semplice selezione.

Per ovviare a questa carenza è stata sviluppata una modalità di ricerca che consente di esprimere questa richiesta in modo semplice e comprensibile.

La sintassi adottata è una derivazione dei range per cui avremo

`[campo]={percorso dal quale estrarre i dati;nome del file di selezione}`

Il percorso dal quale estrarre i dati dev'essere chiave e può essere espresso anche sotto forma di *Search Alias*. Esso può corrispondere anche ad una chiave virtuale frutto di concatenazione o altro

Vediamo un esempio:

In un applicazione ho documenti e fascicoli. Data una selezione di documenti voglio identificare tutti i fascicoli cui questi documenti appartengono.

La selezione di nostro interesse avrà quindi la seguente forma

`[/fascicolo/@numero]={XML,/doc/rif_interni/rif/@cod_fasc;3se...}`

Questo ci dice che data la selezione *3se...* desideriamo estrarre dal canale noto come *XML,/doc/rif_interni/rif/@cod_fasc* i valori dei documenti selezionati ed utilizzarli per compiere la ricerca nel canale */fascicolo/@numero*.

Esempi La ricerca per lista di valori...

```
[CognomeENome]=" Tirabassi Roberto" OR " Bazzigotti Franco"
```

...è analoga all'espressione...

```
[CognomeENome]={ " Tirabassi Roberto", " Bazzigotti Franco"}
```

... conducendo allo stesso risultato.

Altrettanto dicasi per una ricerca del tipo...

```
[Numero]={1,3,5,7,8,2,4}
```

...nella quale si evidenzia che l'ordine dei componenti l'elenco non è significativo per un buon esito della ricerca.

Estensioni della ricerca per proiezione



Not implemented/release yet

La ricerca per proiezione è stato ampliato nelle versioni dalla 24.5.1 in avanti.

Per essi si prevede che la sintassi discussa precedentemente venga così ampliata:

`[campo]={vocabolario dal estrarre i termini di ricerca;nome del file di selezione;frequenza minima;frequenza massima}`

Tutti i parametri, ad eccezione del **primo**, sono opzionali. Avremo quindi la possibilità di compiere una ricerca effettuando un'analisi spettrale di un vocabolario, basata o meno su una precedente selezione, selezionando da essa solo i termini che rientrano in uno specifico intervallo di frequenze.

Qualora non si indichi la selezione l'analisi spettrale viene compiuta solo sulle frequenze. Per esse, omettere la frequenza minima comporta impostarla automaticamente al valore 1 mentre omettere quella massima non ci causa alcun limite.

La sintassi è stata concepita per poter evolvere ulteriormente secondo un criterio posizionale.

Vediamo alcuni esempi chiarificatori.

Esempi Rimaniamo nell'esempio di cui sopra, vale a dire un archivio che preveda fascicoli e documenti.

Selezioniamo tutti i fascicoli nei quali figurino almeno un documento il cui mittente è Roberto Tirabassi.

```
[/doc/minuta/mittente/@nome_persona]="Roberto Tirabassi"
```

```
<?THEN?>
```

```
[/fascicolo/@numero]={/doc/rif_interni/fascicoli_collegati/fasc/@cod;<?SEL0?>}
```



La prima ricerca seleziona dei documenti. La seconda, che da essa deriva, compie un'analisi spettrale sul vocabolario dei fascicoli collegati così come citati nei documenti così da estrarre i codici(Numeri) dei soli fascicoli cui appartengono questi documenti.

Selezioniamo tutti i fascicoli nei quali figurino non meno di 3 documenti.

```
[/fascicolo/@numero]={/doc/rif_interni/fascicoli_collegati/fasc/@cod;;3}
```

L'analisi spettrale viene compiuta senza l'ausilio di una precedente selezione ma limitandosi ad indicare che il codice del fascicolo presente nei vocabolari dei documenti deve apparire come minimo in 3 documenti diversi.

Per analogia avremo

```
[/doc/minuta/mittente/@nome_persona]="Roberto Tirabassi"  
<?THEN?>  
[/fascicolo/@numero]={/doc/rif_interni/fascicoli_collegati/fasc/@cod;<?SEL0?>;3}
```

In questo modo selezioniamo i fascicoli nei quali appaiono almeno 3 documenti i cui mittente è Roberto Tirabassi. Ed ancora:

```
[/doc/minuta/mittente/@nome_persona]="Roberto Tirabassi"  
<?THEN?>  
[/doc/rif_interni/fascicoli_collegati/fasc/@cod]={/doc/rif_interni/fascicoli_collegati/fasc/@cod;<?SEL0?>}  
<?THEN?>  
[/fascicolo/@numero]={/doc/rif_interni/fascicoli_collegati/fasc/@cod;<?SEL1?>;3}
```

Quest'espressione è più complessa. Prima selezioniamo tutti i documenti il cui mittente è Roberto Tirabassi, poi selezioniamo, indistintamente, tutti i documenti che appartengono agli stessi fascicoli. In fine selezioniamo solo quei fascicoli in cui appaia almeno un record il cui mittente è Roberto Tirabassi e che contengano non meno di 3 documenti (anche se di mittenti diversi).

Selezioniamo tutti i documenti che sono in un fascicolo e che sono i soli documenti in tale fascicolo.

```
[/doc/rif_interni/fascicoli_collegati/fasc/@cod]={/doc/rif_interni/fascicoli_collegati/fasc/@cod;;1;1}
```

Questa ricerca agisce su una sola tipologia di unità informativa (ed un solo vocabolario) e quindi applica direttamente su di essa l'analisi spettrale. Si naviga il vocabolario che ci interessa e si estraggono da esso solo i termini aventi frequenza pari ad 1 (Limite inferiore 1, limite superiore 1). Non si applica alcuna restrizione riferita ad una precedente selezione ma solo quella sulla frequenza. E' facile notare come da questa si possa derivare, teoricamente, la ricerca

```
[/doc/rif_interni/fascicoli_collegati/fasc/@cod]={/doc/rif_interni/fascicoli_collegati/fasc/@cod}
```

Questa non applica alcuna restrizione e, di fatto, equivale alla ricerca

```
[/doc/rif_interni/fascicoli_collegati/fasc/@cod]=*
```

Va da se che in questo caso è improprio sfruttare la sintassi della ricerca per proiezione.

1)
salvo qualche caso particolare

2)
La cosa può essere ottenuta in modo più efficiente usando con un range aperto uno dei due casi precedenti

3), 4)
Altrimenti tanto vale usare uno dei precedenti casi con un range aperto

5)
Ad esclusione del caso in cui il campo 'Anno' sia stato configurato come empty_key, ma trattandosi di un'eccezione non la si prende in esame in questa sede.



6)

Richiede un lavoro inutile