# CHAPTER 5: CÁC KĨ THUẬT TÌM KIẾM NÂNG CAO

Như đã đề cập,đã có nhiều ứng dụng cài đặt chức năng tìm kiếm dựa vào thư viện Lucene API.Nhiều ứng dụng đã xây dựng được những bộ tìm kiếm hoàn thiện với nhiều chức năng-trong đó phải kể đến Nutch của cùng tác giả,và để có thể làm được điều này,ta cần tìm hiểu sâu hơn những khả năng khác đã được xây dựng trong Lucene

## Sắp xếp kết quả trả về

Từ Lucene 1.4 các kết quả trả về được sắp xếp theo thứ tự giảm của độ ưu tiên (Score).nghĩa là các tài liệu nào có kết quả chính xác hơn sẽ xuất hiện ở đầu.Tuy nhiên với những phương thức chồng (overloaded) của sẻarch ta hoàn toàn có thể sắp xếp theo nhiều cách khác ,bao gồm cả việc sắp xếp các trường (Field) tùy tham số truyền vào.

Trong những trường hợp sau: ta đều sử dụng phương thức search và trả ra kết quả là collection Hits-chứa kết quả đã được sắp xếp .Ở đây ta chỉ cần quan tâm tới đối tượng Sort truyền vào ở dạng nào mà thôi.



### Sắp xếp theo độ chính xác

Đây là cách sắp xếp thường thấy và là mặc định trong phương thức search (Query) của đối tượng IndexSearcher.để sắp xếp theo kiểu này tham số truyền vào của Sort có thể là

* Search(query,null)
* Search(query,Sort.RELEVANCE)
* Search(query,new Sort())

### Sắp xếp theo độ thứ tự được đánh chỉ mục(index order)

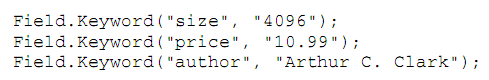
Thứ tự chỉ mục gọi là Id,tức là vị trí của tài liệu được lưu trong chỉ mục tùy vào loại truy vấn Query.chú ý rằng thứ tự của tài liệu trong chỉ mục không phải là độ ưu tiên-Score.Sắp xếp kiểu này Id sẽ được sắp tăng dần với tham số:

* Search(query,Sort.INDEXORDER)

### Sắp xếp theo Trường (Field)

Để săp xếp theo Field,ta cần chú ý một nguyên tắc là:

* Field đó phải được indexed nhưng không được tokened,chẳng hạn như là Field.Keyword
* Giá trị của Field đó phải được chuyển thành kiểu Integer,Floats hoặc là Strings,ví dụ như:



Khi đó,cần sắp xếp theo Field nào ta chỉ cần khởi tạo đối tượng Sort theo Field đã index là được,như ví dụ trên ta muốn sort theo size ta cần:

* Search(query,new Sort(“size”))

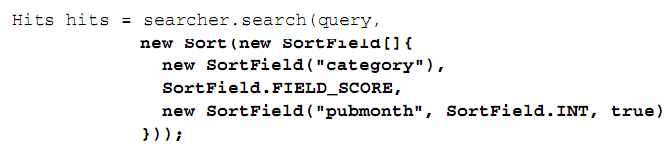
### Đảo ngược thứ tự đánh chỉ mục

Mặc nhiên,thứ tự đánh chỉ mục là giảm dần theo độ chính xác, và tăng dần các trường khác.Tuy nhiên ta hoàn toàn có thể đảo ngược thứ tự này theo từng trường ,chẳng hạn ví dụ sau sắp theo độ giảm của trường tên là “size”-tức là tài liệu nào có size lớn hơn sẽ xuất hiện ở đầu,bằng cách sau:

* Search(query,new Sort(“size”,true))

### Sắp xếp theo nhiều trường khác nhau(multiple fields)

Để thực hiện việc sắp xếp theo nhiều trường khác nhau theo tiêu chí nào đó,ta sử dụng một mảng SortFields và add các Field cần sắp xếp vào trong mảng đó,



Trong ví dụ trên,ta sắp xếp theo độ giảm độ ưu tiên của field “category”,đồng thời nếu các cuốn sách trong cùng “category” được sắp giảm theo năm xuât bản.

SortField có thể nhận các hằng số :

|  |  |
| --- | --- |
| SortField | Ý nghĩa |
| FIELD\_SCORE | kiểu sắp xếp theo độ chính xác |
| FIELD\_DOC | Sắp xếp theo ID của tài liệu |
| FIELD\_AUTO | Sắp xếp theo thứ tự Alphabate của tên Field |
| FIELD\_STRING , FIELD\_INT, FIELD\_FLOAT | Tùy vào từng field nhận kiểu dữ liệu nào mà ta sắp xếp theo kiểu tương ứng |