

УДК 026.06

**К. О. Сбойчаков**

*ГПНТБ России*

## **Построение и использование фасетов при поиске в САБ ИРБИС64**

Изложены результаты работы по развитию оригинальной методики построения и использования фасетов в Системе автоматизации библиотек ИРБИС64.

**Ключевые слова:** Система автоматизации библиотек ИРБИС64, фасеты, построение, использование.

---

UDC 026.06

**Konstantin Sboichakov**

*Russian National Public Library for Science  
and Technology, Moscow, Russia*

## **Designing and using facets in IRBIS64 LAS retrieval**

The results of developing original methods of building and using facets in IRBIS64 Library Automation System are discussed.

**Keywords:** Library Automation System, facets, design, use.

---

Фасетом будем называть поисковый инверсный термин – один из верхних элементов списка, отсортированного по числу вхождений данного типа терминов в результат поиска. Тип фасета, таким образом, совпадает с типом инверсного термина и задаётся в ИРБИС64 префиксом, например для авторов А=. Каждому результату поиска соответствует свой набор фасетов, указывающих на наиболее часто встречающиеся термины (в соответствии с их типом) в найденных документах.

Таким образом, фасеты помогают пользователю автоматизированной поисковой системы сделать предварительную оценку результатов поиска по различным возможным смысловым срезам и особенно эффективны при значительной выдаче.

Фасеты являются инструментом сжатия найденной информации по нескольким типам поисковых терминов. Как форма сжатия фасеты, конечно, несовершенны; они максимально полезны при поиске в больших текстовых массивах, когда поисковая задача нечётко определена. В этом случае тип фасета – это слова из текстов.

При поиске по библиографии, которая хорошо структурирована заранее, фасеты могут дать пользователю подсказки, в неявном виде содержащиеся в ЭК. В этом случае фасеты служат инструментом извлечения знаний. В связи с тем, что библиографические записи в ЭК слабо связаны друг с другом, такая информация полезна, скорее, исследователю, чем простому пользователю.

Построение инверсных файлов, необходимых для получения фасетов по результатам поиска, происходит при загрузке словаря БД ЭК.

Параметр `CREATE_CELLS_ON_IF_REORGANIZATION=1` (умолчание 0) в ини-файле `APM` Администратор ИРБИС64 (`irbisa.ini`) позволяет построить фасеты в режиме перезагрузки словаря БД ЭК.

Параметр `MAX_POSTINGS_NUMBER_IN_CELL` (умолчание 5) устанавливает минимально необходимое количество ссылок для термина, чтобы он вошёл в список фасетов.

Параметр `MAX_LENGTH_CELL_TERM` (умолчание 2) устанавливает минимальную длину фасета (термина словаря без префикса).

Типы фасетов перечислены в текстовом файле `dbname.cells` в виде префиксов. Для каждого типа фасетов при загрузке словаря строится бинарный инверсный файл `dbname.cellN`, где `N` – порядковый номер фасета в списке `dbname.cells` начиная с единицы.

Файл `dbname.cellN` имеет жёсткую структуру: `MFN INDEX`, где `MFN` – номер записи в БД ЭК, `INDEX` – индекс фасета – порядковый номер фасета (термина) в списке терминов данного типа. Одному `MFN` могут соответствовать несколько `INDEX`.

По окончании загрузки строится файл дополнительных терминов словаря, которые используются для расшифровки индекса фасета данного типа в конкретный термин словаря. Вид дополнительного термина: `=Префикс=INDEX=Термин`. Например, для автора `SHORT: =A=3203=SHORT`

Итак, алгоритм формирования данного типа фасетов вкратце можно описать так:

1. Сканирование терминов словаря данного типа фасетов, например, авторов с префиксом `A=`.

2. Формирование списка фасетов из терминов словаря, количество которых превышает MAX\_POSTINGS\_NUMBER\_IN\_CELL и длина больше MAX\_LENGTH\_CELL\_TERM, и назначение каждому термину (теперь – фасету) уникального индекса (порядковый номер в индексном файле).

3. На основе списка фасетов формирование списка дополнительных терминов словаря, которые должны быть внесены в инверсный файл, и дополнительного инверсного файла для поиска индекса фасета по номеру записи в ЭК.

4. По окончании сканирования производится добавление дополнительных терминов словаря в инверсный файл ЭК.

Пример поиска по ЭК ГПНТБ России:

Запрос: U=5\$.

Найдено: 164910 записей.

Время поиска 100 ms, время расчета фасетов 150 ms.

БРЮНО, А. Д. (93)	1990 (9435)	НАУКА(М.) (1993)
BEILSTEIN, F. K. (77)	1991 (8636)	URSS(М.) (1096)
ПЕРЕЛЬМАН, Я. И. (74)	2006 (7130)	ФИЗМАТЛИТ(М.) (1013)
ТРОФИМОВА, Т. И. (74)	2007 (6785)	SPRINGER(BERLIN ;
СТРЕЛЬЦОВ, В. Н. (74)	2008 (6737)	HEIDELBERG ) (971)
		SPRINGER(BERLIN;
		HEIDELBERG) (853)

---

***Konstantin Sboichakov**, senior researcher, Russian National Public Library for Science and Technology;*

*Ksboychakov@yandex.ru*

*17, 3rd Khoroshevskaya str., Moscow, 123298, Russia*