Problem statement

Users need to perform full-text search over instances, items, holdings. The following requirements should be fulfilled:

• **Assumption:** For over 20+ millions of records search request should be executed in less than 500ms

> 我们用es可以做到2000W数据在500ms以内

**We can use ElasticSearch to achieve 2000W data within 500ms**

• Multi-tenant search (e.g. for consortia) should allow to share metadata view permission from one tenant to another

> 只需要根据权限查询条不同索, 我们就能查询到其他租户的数据。

**We can query the data of other tenants just according to the different indexes**

• Search should be efficient in terms of multi-tenancy, namely the amount of data in each tenant shouldn’t affect request execution time for other tenants

> 每个租户可以分成不同索引, 相互独立, 不会影响另外的租户的查询操作

**Each tenant can be divided into separate index, independent of each other, without affecting the query operations of other tenants**

• Result counts should be precise for any query and amount of data

> 已实现

**Done**

• Auto-complete should be based on titles. The response time should be less than 100ms

> 尚未实现，需要新增接口。因目前项目暂无此需求

• Facets (e.g. how many Print-books are found for the search request) should be precise

> 尚未实现，需要新增接口。因目前项目暂无此需求

• **Assumption:** Rich full-text functionality which will be beneficial for user, should be provided, namely:

> 已实现

**Done**

• Stemming for words (e.g. find record with term "books" for query "book")

> 已实现

**Done**

• Stop-words (e.g. and. or for English language) should not affect relevancy

> 已实现

**Done**

• Relevancy scoring should be based on TF-IDF frequencies, in order to provide the most relevant records at the top

> 已实现

**Done**

• **Didyoumean for input string spelling correction should be implemented and showed as tip if there is more significantly more relevant query**

> 尚未实现。因目前项目暂无此需求

• **All language depended features should support certain predefined list of languages**

> 尚未实现自动语种识别，只能为字段配置固定的语种。因目前项目暂无此需求
• There should be support for CQL queries from input string

尚未实现。因目前项目暂无此需求