

Solrの重複uniqueKeyの扱い

- 単一のshard
- 複数のshard
- 出入口
- いろいろ

単一のshard

uniqueKeyの重複はない

複数のshard

DistributedSearch – Solr Wiki

- Distributed Searching Limitations
 - The unique key field must be unique across all shards.

複数のshard: 実際の処理

`src/java/org/apache/solr/handler/component/QueryComponent.java` で処理されている

- `uniqueKey`の重複を検出したら結果に含めない
 - 一番最初に処理されたものが結果に入る
それ以外は結果に含まれない

拡張

- shardの名前の大小で優先順位を付ける
- timestampのようなフィールドで優先順位をつける
- 作ってみた
 - <https://github.com/haruyama/solr/>
 - duplicate_unique_key
 - 実装はちょー適当
 - Apache Licence 2.0