MongoLantern - Fulltext Search Server

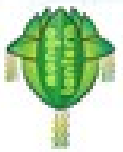
From

Download Link: http://sourceforge.net/projects/mongolantern/

Presented By: Sougata Pal (Skall)

skall.paul@techunits.com

February 12, 2012



Introduction to MongoLantern

MongoLantern is an open source full text search server using MongoDB as index storage, which allows MongoLantern to migrate any changes very easily into account using MongoDB API. It's written originally written in PHP can be migrated to any desired language as required using it's future APIs.

Basic Features:

Token based search algorithm, Case insensitive Fulltext search, Different Search match mode, Field based search algorithms, Pre-built Index support.

Note: MongoLantern is also listed in trusted apps in MongoDB official site, you can check it here.

MongoLantern Latest Release Notes (Version: 0.7)

1. Namespaces updated.
2. ErrorReporting type update handler added. [user: prodigyview]
3. CSV Indexer finalized.

Class: MongoLanternIndexer

Functional Guide:

MongoLanternIndexer::$indexName	:	Set Index name on MongoDB
MongoLanternIndexer::dictionaryEnabled()	:	Enable or disable dictonary: true / false. default: true. This method must be called before MongoLanternIndexer::Connect()
MongoLanternIndexer::Connect()	:	Connect to MongoLantern target Index
MongoLanternIndexer::setFields()	:	Set fields to be searched later. On this field list the index will be optimized.
MongoLanternIndexer::setDocument()	:	Set document to MongoLantern Indexer instance to commit to index.
MongoLanternIndexer::Commit()	:	Save dataset to index for searching.
MongoLanternIndexer::Optimize()	:	Optimize Index with few required fields & also fields specified by MongoLanternIndexer::setFields()
MongoLanternIndexer::totalDocs()	:	Get total number of documents available on index.
MongoLanternIndexer::validateDocumentID()	:	Check for document existance with documentID.
MongoLanternIndexer::dropFieldIndex()	:	Remove any pre-built field Index specified by MongoLanternIndexer::setFields(). Note that the required indexes must not be deleted, which may cause issue in search.

Class: MongoLanternDocument

Functional Guide:

MongoLanternDocument::setField() : Set fields to document instance to pass to MongoLanternIndexer::setDocument()

Class: MongoLanternField

Functional Guide:

MongoLanternField::Keyword() : Set field which will required to be used as Keyword for search, this field value will not be tokenized and will be used as search term as a whole.
e.g . Books ISBN number

MongoLanternField::Text() : TEXT content must be tokenized OR used as tokenized data as tokens. Also keep the whole content into index for field level filter.
e.g . Blog posts title

MongoLanternField::UnStored() : This field value will only be tokenized and but will not be stored to index as a whole. You can't run subquery on this field.
e.g. Blog posts body

MongoLanternField::UnIndexed() : UnIndexed content must not be tokenized OR used as tokens. Just keep the content into index. This can be used for sorting or subquery.
e.g. Blog post author

MongoLanternField::Binary() : BINARY content must not be tokenized OR used as tokens. Just keep the content into index. Not implemented yet.

Class: MongoLanternQuery

Functional Guide:

MongoLanternQuery::$indexName : Set Index name on MongoLantern

MongoLanternQuery::$debug : true / false. It enables or disables the debug mode. If enabled it will print the debug data but no results could be obtained.

MongoLanternIndexer::Connect() : Connect to MongoLantern target Index

MongoLanternIndexer::setQuery() : Set keyword for search.

MongoLanternIndexer::setMatchMode() : Set result match mode for search. There are 3 types of Query Mode: BESTMATCH, SUGGESTED, ANY

MongoLanternIndexer::setSortMode() : Set result sorting mode for returing results. There are 3 types of Sorting Mode: RANK, CREATED, DOCUMENT

MongoLanternIndexer::setLimit() : Limit the result.

MongoLanternIndexer::setSkip() : Paginate the result along with MongoLanternIndexer::setLimit()

MongoLanternIndexer::setIntelligentQueryMode() : Resolve parital query typo as mentioned in release notes.

MongoLanternIndexer::Execute() : Execute the query and result results.

MongoLanternIndexer::getStats() : Get query execution stats.

MongoLanternIndexer::Clear() : **Destroy MongoLanternQuery instance**.

3rd Party Class: MongoLanternIndexCSV

Functional Guide:

MongoLanternQuery::dictionaryEnabled()	**:**	Enable or Disable dictionary support for CSV indexer.
MongoLanternQuery::setCSV()	**:**	Set valid CSV file or CSV String.
MongoLanternIndexer::setFields()	**:**	Set fields to be indexed and optimize.
MongoLanternIndexer::setDocumentIDField()	**:**	Set field for unique documentID which exists in CSV.
MongoLanternIndexer::setFieldsType()	**:**	Set fields type for best index & results.
MongoLanternIndexer::Commit()	**:**	Save document to MongoLantern Index.

*** Note:** Please use Examples/CSV/CSVIndexer.php to test CSV indexer. The search mechanism is same as genral indexer.

How to create/update fulltext search index with MongoLantern ?

Creating/Updating a New Index

```
MongoLanternUtility::setErrorReporting(E_ALL | E_STRICT);

$indexer = new MongoLanternIndexer();

$indexer->indexName = 'People';

//  Enable or disable dictonary: true / false.  default: true;
$indexer->dictionaryEnabled(false);

$indexer->Connect();

$indexer->setFields(array(
        'firstname',
        'lastname',
        'age',
        'email',
        'height',
));
```

Note: MongoLanternIndexer::setFields() used to preserve & optimize fields to the index to search with those specific fields later on. Fields specified in the function can only be used in subquery.

*** Deprecated Method:**

```
$indexer->setDocument($uniqueIndexID, array(
        'firstname'             =>  $doc['firstname'],
        'lastname'              =>  $doc['lastname'],
        'email'                 =>  $doc['email'],
        'age'                   =>  $doc['age'],
        'height'                =>  $doc['height'],
        'joined_on'             =>  $doc['joined_on'],
));
```

How to create fulltext search index with MongoLantern ?

Creating/Updating a New Index(continue...)

*** Current Advanced Method:**

```
$docObj = new MongoLanternDocument();

$docObj->setField(MongoLanternField::Keyword('email', $doc['email']))
        ->setField(MongoLanternField::Text('firstname', $doc['firstname']))
        ->setField(MongoLanternField::UnStored('lastname', $doc['lastname']))
        ->setField(MongoLanternField::UnIndexed('age', $doc['age']))
        ->setField(MongoLanternField::UnIndexed('height', $doc['height']))
        ->setField(MongoLanternField::UnIndexed('joined_on', $doc['joined_on']));

$indexer->setDocument($uniqueIndexID, $docObj);
```

Note: To update data into mongolantern index you must pass the correct value of $uniqueIndexID to the function. The document will be identified via this unique ID.

Also note that the new document object method is backward compatible so you can always index using array instead using document object. Using array is highly discouraged due to result quality issue.

How to create fulltext search index with MongoLantern ?

Creating/Updating a New Index(continue...)

```
$indexer->Commit();
```

Note: Data will only be stored to index if and only if you commit it.

```
$indexer->Optimize();
```

Note: MongoLanternIndexer::Optimize() is used to set indexes on the fields specifiedvia setFields. It will also created default search & sort indexes.

How to validate a document from index with DocumentID ?

Creating/Updating a New Index(continue...)

$indexer->validateDocumentID($docID);

Note: This will check whether the documentID suppiled exists into index. Output type is boolean.

Success: true
Failure: false

Now you can also get info on total available docs into the index.

$indexer->totalDocs();

How to search from pre-built index with MongoLantern ?

Searching an Index

```
$search = new MongoLanternQuery();

$search->indexName = 'People';

$search->Connect();

$search->setQuery($keyword);


$search->setMatchMode('BESTMATCH');
```

Note: There are 3 types of Query Mode: RANK, CREATED, DOCUMENT yet.

```
$search->setSortMode('RANK');
```

Note: There are 3 types of Query Mode: RANK, CREATED, DOCUMENT yet.

```
$search->setIntelligentQueryMode(true);
```

Note: Resolve parital query typo detection.
e.g. Search Term: millban tower will return you "Millbank Tower" from Place index.
Search Term: "markku nikkane" will return you "Markku Nikkanen" from People index.

Use of this kind of query is highly discouraged for larger databases. MongoLantern team is trying to optimize it.

How to search from pre-built index with MongoLantern ?

Searching an Index(continue...)

```
$resultList = $search->Execute();

print_r ( $resultList );

Array
(
    [0] => Array
        (
            [_id] => MongoId Object
                (
                    [$id] => 4f106729293d861805000018
                )

            [documentID] => 8
            [document] => Array
                (
                    [firstname] => subhajit
                    [lastname] => mjee
                    [email] => subhajit@techunits.com
                    [age] => 25
                    [height] => 6.1
                    [joined_on] => 23
                )

            [created] => 1326904795
            [phraseDistanceRank] => 30
        )
)
```

How to search from pre-built index with MongoLantern ?

Searching an Index(continue...)

```
print_r ( $search->getStats() );

Array
(
        [executionTime] => Array
                (
                        [query] => 0.004927
                        [sort] => 0.0023660000000001
                        [total] => 0.0072930000000001
                )

        [resultsCount] => 2
)
```

Note: MongoLanternQuery::getStats() must be called after MongoLanternQuery::Execute(). It will just return the stats info for the last query.

How to search from pre-built index with MongoLantern ?

Query Construction API

MongoLatern is also having support for query construction via subquery & range query methods. Example as follows:

```
$parser = new MongoLanternQueryParser();

$parser->setQueryTerm('techunits');

$parser->setSubqueryTerm('email', 'techunits.com', false);
```

Note: 3rd parameter to the MongoLanternQueryParser::setSubqueryTerm() is used to specify whether the match mode will be exact or partial match.

true: enable exact match
False: enable partial match

e.g. If “false” then it will match techunits.com to any email address containing techunits.com in the above example.

```
$parser->setRange('height', 5.0, 6.0);
```

Finally the the Query parser object instance can be passed to MongoLanternQuery::setQuery().

```
$search->setQuery($parser);
```

MongoLantern - Fulltext Search Server

Thank you using MongoLantern

Developer's Email: skall.paul@techunits.com

Community Discussion:

mongolantern@googlegroups.com

Download Link: http://sourceforge.net/projects/mongolantern/