



Atlas Search



Atlas Search

Objectives

Gain a basic understanding of Atlas Search

Know when to choose Atlas Search and how to get the most from it.

Understand when Atlas search might be suboptimal.



Atlas Search

Introductions

Integrated full-text search in MongoDB
Atlas

Eliminates the need to run a separate
search system alongside your database.

Further Training: DS110 - Atlas Search

[Getting Started](#)

Agenda

Introduction to Atlas Search (10 mins)

Demo (10 mins)

Advanced features (5 mins)

Performance & Sizing (3 mins)

Q&A (5 mins)



"With a database query - you know exactly what you want. When you use search you are open to suggestions." - unknown

```
db.movies.find({ plot : "Raising Arizona"}) - plot IS EXACTLY "Raising Arizona"
```

```
db.movies.find({ plot: /Raising Arizona/}) - plot CONTAINS the phrase "Raising Arizona"
```

```
db.movies.aggregate({$search : { text : { query : "Raising Arizona' , path: "plot" }}}) - plot  
mentions some variations of "to Raise", "Arizona" or both.
```



Scoring and Relevance



Search:
Javascript Book

Title	Score
Javascript the Definitive Guide	75
Javascript the Good Parts	60
The Beginners Guide to Node.js	35
Angular for Dummies	25
Introducing MongoDB	10

More relevant
results should
appear first

Mountain
Bike Gloves

**WFX Mountain Bike
Bicycle Cycling Full
Finger Gloves BMX
MTB Riding
Touchscreen**

The screenshot shows an eBay search results page for 'mountain bike gloves'. At the top, there's a navigation bar with links for 'Hello John (Sign in)', 'Daily Deals', 'Brand Outlet', 'Help & Contact', and a promotional message '15% off selected sellers. Use code BAG15OFF →'. Below the search bar, it says 'Related: mountain bike gloves full finger fox mountain bike gloves mtb gloves mountain bike helmet fox gloves cycling gloves winter mountain bike gloves mountain...'. The search bar itself contains 'mountain bike gloves'. The results count is '1,173 results for mountain bike glo...'. There are filters for 'All Listings', 'Accepts Offers', 'Auction', 'Buy it now', 'Condition', 'Item location', 'Local', 'Best Match', and 'Postage to: G759NW'. The results section starts with a sponsored listing for 'WFX Mountain Bike Bicycle Cycling Full Finger Gloves BMX MTB Riding Touchscreen' priced at £6.99, followed by another sponsored listing for 'WFX Mountain Bike Bicycle Full Finger Gloves BMX MTB Touch Screen Dirtpaw Men UK' also at £6.99, and a third sponsored listing for 'Full Finger Gloves All Weather Mountain Bike Hard Knuckle MTB BMX Off Road Sport' ranging from £12.99 to £17.99.



Recommendation engines

The screenshot shows the Netflix homepage with a dark background. At the top, there's a navigation bar with three colored dots (red, yellow, green) on the left, the Netflix logo, and links for Home, Series, Films, New & Popular, and My List. On the right are icons for search, notifications, and account settings.

Exciting Sci-Fi TV

Shows listed include: MISFITS, STAR TREK VOYAGER, ORPHAN BLACK, TRAVELERS, BLACK LIGHTNING, and ZOO.

Competition Reality TV

Shows listed include: INSTANT HOTEL, CABINS IN THE WILD, CRAZY DELICIOUS, INTERIOR DESIGN MASTERS, and THE AMERICAN BARBECUE SHOWDOWN.

US Crime TV Dramas

Shows listed include: S.W.A.T., L.A.'S FINEST, SONS OF ANARCHY, BREAKING BAD, OZARK, and STARTUP.



What is "Atlas" Search

 mongoDB
Native

Powered by


Easy Deployment

Aggregation Pipeline

Sharded Cluster Support

Performant Text Indexes

Powerful Query Operators

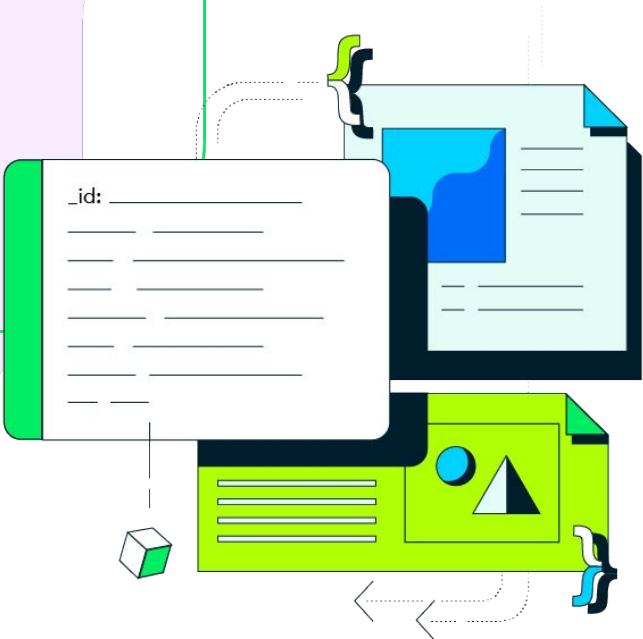
Configurable Language Analyzers

Built-in Highlighting

Flexible Scoring



Atlas Search



Integrated and fully managed

Available with every 4.2+ cluster on Atlas

No need for a separate search architecture

Simple to use

Create indexes with the Atlas UI, CLI, or API

\$search queries are MongoDB aggregations

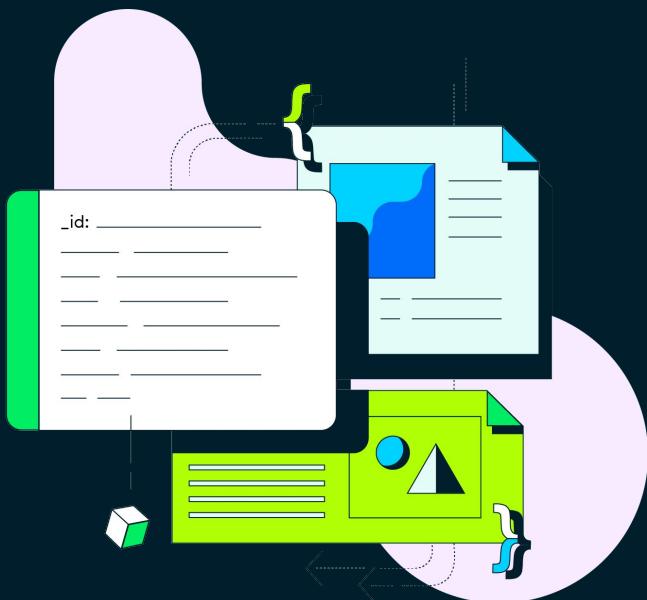
Rich search capabilities

Built on Apache Lucene

Fine-tune relevance with dozens of features



Atlas Search advantages



- Native MongoDB support
- Choice of dynamic or static field mappings to balance performance and ease-of-use
- More than just text search
- Access via the same familiar drivers, tools and syntax as the MongoDB server
- Full integration with the MongoDB aggregation pipeline
- No additional operational overhead



Who is using Atlas Search and why?

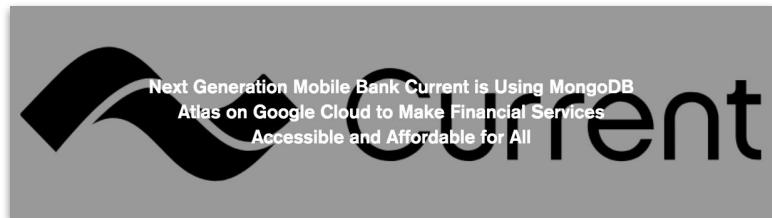
**Event ticketing
that's better for
Your Attendees**

100% of our profits fund education for disadvantaged kids.

[Search events](#)

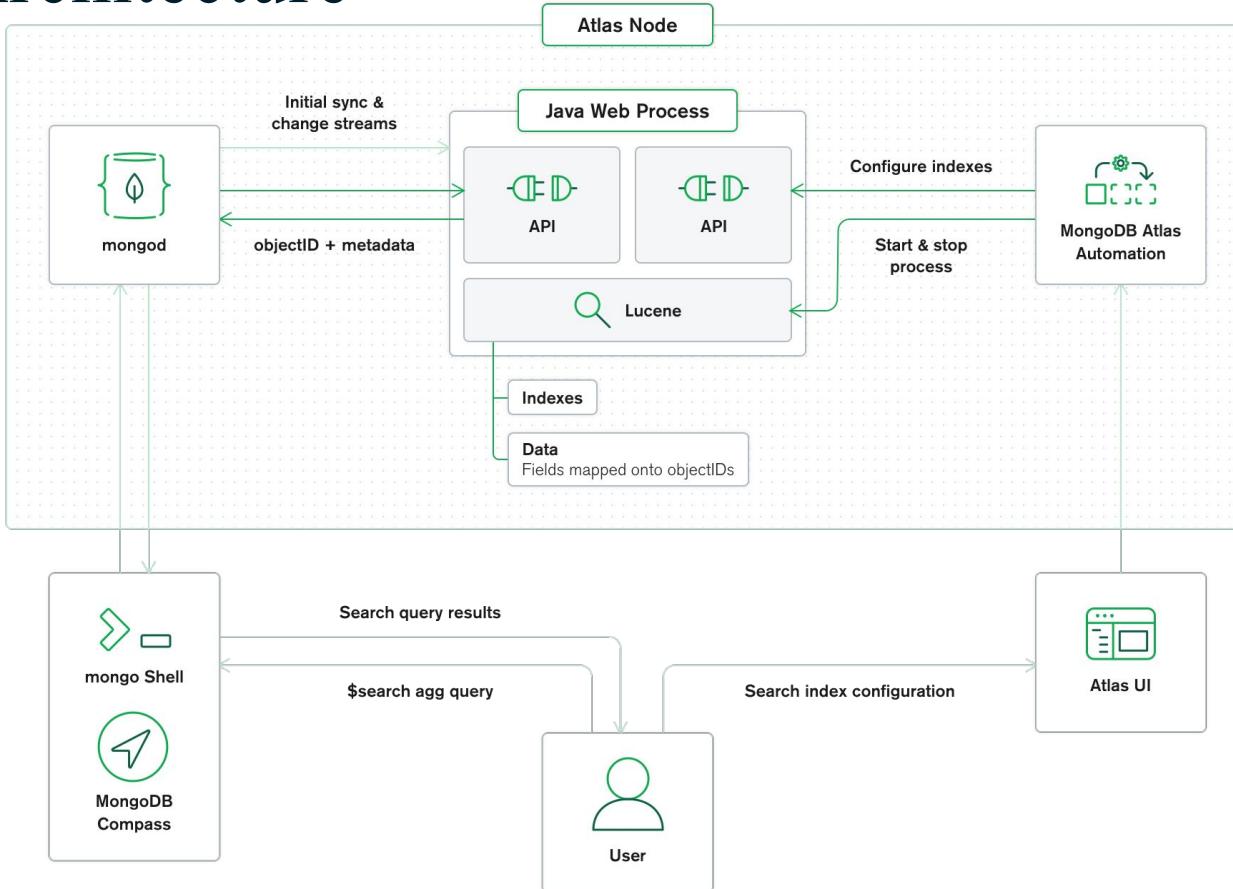
[Host events](#)

From Banking, to Concert Tickets to Online fraud detection - Search is a powerful enhancement to Atlas





Architecture





How to use it?

Search Indexes

Go to the “search” tab of your Atlas Cluster

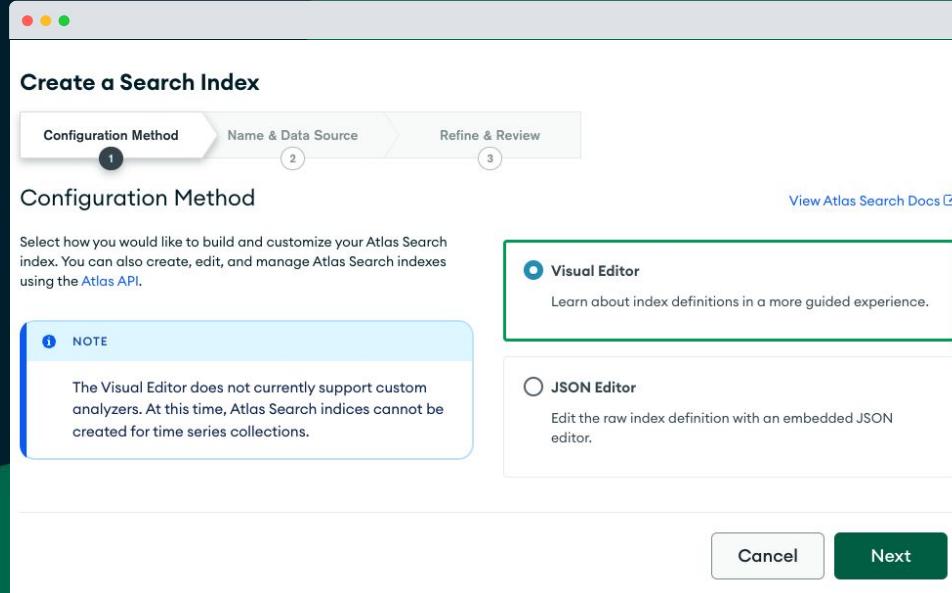
The screenshot shows the MongoDB Atlas Cluster search interface. The top navigation bar includes 'Professional Services', 'Access Manager', 'Billing', 'All Clusters', 'Admin', 'Get Help', and 'Ping'. The left sidebar shows 'Ping Kin Ma' under 'ACTIVE', 'Alias', 'Realm', and 'Charts'. Under 'DATA SERVICES', it lists 'Database', 'Data Lake', 'Triggers', 'Data API' (selected), and 'PREVIEW'. Under 'SECURITY', it lists 'Database Access', 'Network Access', and 'Advanced'. A 'New On Atlas' button is also present. The main content area is titled 'Cluster0' and has tabs for 'Overview', 'Real Time', 'Metrics', 'Collections', 'Search' (selected), 'Profiler', 'Performance Advisor', 'Online Archive', and 'Cmd Line Tools'. A search bar at the top says 'Find a Search Index...'. Below it, there are two sections: 'searchtraining.example2' and 'searchtraining.example3'. Each section contains a table with columns: Name, Index Fields, Status, Size, Documents, and Actions. In 'searchtraining.example2', there is one index named 'default2' with status 'ACTIVE', size 25.39KB, and 11 documents. In 'searchtraining.example3', there is one index named 'default3' with status 'ACTIVE', size 20.92KB, and 9 documents. Both sections have 'QUERY' and '...' buttons in the Actions column.

Name	Index Fields	Status	Size	Documents	Actions
default2	[dynamic]	ACTIVE	25.39KB	11	<button>QUERY</button> ...

Name	Index Fields	Status	Size	Documents	Actions
default3	[dynamic]	ACTIVE	20.92KB	9	<button>QUERY</button> ...

Create Search Index

Click “Create Index” button on the top-right hand corner



Visual Editor

Simple to use GUI

Filled with default values

The screenshot shows a step-by-step wizard titled "Create a Search Index". The current step is "Name & Data Source" (step 2 of 3). The previous step, "Configuration Method", is marked as completed with a green checkmark. The next step, "Refine & Review", is shown with a number 3.

Index Name and Data Source

Atlas Search indexes are specific to a database and collection. Give your search index a name for easy reference, and select the database and collection you want to draw data from.

Index Name: default

Database and Collection

Search for database or collection ...

▼ searchtraining

example

example3

example2

Back **Cancel** **Next**

JSON Editor

Easy to place the whole index setting in JSON format

Useful tool for preparing deployment script using API

The screenshot shows the "Create a Search Index" wizard in the MongoDB Atlas interface. The current step is "JSON Editor" (step 2). The interface includes:

- A "Database and Collection" section with a search bar and a list of available databases and collections: "searchtraining" (expanded), "example" (selected), "example3", and "example2".
- An "Index Name" field containing "default".
- A code editor area displaying the following JSON configuration:

```
1 {  
2   "mappings": {  
3     "dynamic": true  
4   }  
5 }
```
- Navigation buttons at the bottom: "Back", "Cancel", and "Next".

Review

Review the Index setting

Refine the index

Click “Create Search Index” to confirm the index creation

The screenshot shows a 'Create a Search Index' wizard with three steps: 'Configuration Method', 'Name & Data Source', and 'Refine & Review'. The 'Refine & Review' step is active, indicated by a green checkmark and the number '3' in a circle. The title 'Create a Search Index' is at the top. Below it, a sub-header says 'Review "default" for searchtraining.example' with a 'VIEW JSON' button and a 'Learn How to Refine your Index' link. A note states: 'By default, your Atlas Search index will have the following configurations. We recommend starting with this and refining it later if you need to.' A table lists five configuration items:

Setting	Description	Value
Index Analyzer	Creates searchable terms from data to be indexed.	lucene.standard
Search Analyzer	Parse \$search queries into searchable terms.	lucene.standard
Dynamic Mapping	Automatically index common data types in a collection	On
Store Full Document	Make all data available for lookup on Atlas Search side. See use cases and Performance Considerations for details.	Off
Field Mappings	Define data types and input parameters for specific fields.	None

At the bottom, a note says: 'You can edit these defaults at any time to fine-tune relevance and improve search performance.' There are 'Refine Your Index', 'Back', 'Cancel', and 'Create Search Index' buttons.



Creating Index...

Atlas Search will index the contents in the background, you can check the status in the indexes view

searchtraining.example					
Name	Index Fields	Status ⓘ	Size	Documents	Actions
default	[dynamic]	NOT STARTED			<button>QUERY</button> <button>...</button>

searchtraining.example					
Name	Index Fields	Status ⓘ	Size	Documents	Actions
default	[dynamic]	ACTIVE View status details	Primary Node: 13.64KB	Primary Node: 11 (61%) indexed of 18 total	<button>QUERY</button> <button>...</button>

searchtraining.example					
Name	Index Fields	Status ⓘ	Size	Documents	Actions
default	[dynamic]	ACTIVE View status details	Primary Node: 17.14KB	Primary Node: 18 (100%) indexed of 18 total	<button>QUERY</button> <button>...</button>



Query Tester

The screenshot shows the MongoDB Atlas Search Tester interface. The left sidebar has links for Index Overview, Search Tester (which is selected and highlighted in green), and Quick Start Tutorials. The main area is titled "Search Tester" and contains instructions about running a search query on sample documents. A search bar at the top contains the query "鲷魚湧". Below the search bar, it says "This search took 0.19 seconds". Two search results are displayed in cards:

- Result 1:** SCORE: 4.3396148681640625 _id: "62664b188f0af33dd083a542"
我公司縣
Matching fields: text
- Result 2:** SCORE: 4.3396148681640625 _id: "6268b8d346e6515aa9ac374e"
我公司縣
Matching fields: text



View Query Syntax

To execute the full \$search query, [connect to the mongo shell](#) and copy the code below into an aggregation pipeline. You can also export the code to a programming language.

MongoDB API

Export to Language

```
1 []
2 {
3     $search: {
4         index: 'default',
5         text: {
6             query: '鰯魚涌',
7             path: {
8                 'wildcard': '*'
9             }
10        }
11    }
12 }
13 ]
```



Look very simple, is there any trap?

Close

[View Atlas Search Docs](#)

Try this query on **training.example** using your search index and see the [wildcard queries](#). To run more complex queries, visit the [Atlas search documentation](#).

[RESET](#) [Search](#) [View Query Syntax](#)

>

>



Query scenarios

1. Wildcard query
2. Query by path
3. Query using Analyzers
4. Query using multiple Analyzers

Settings



Data schema

```
{  
  "title": "Search Data",  
  "description": "An entry for search",  
  "type": "object",  
  "properties": {  
    "_id": {  
      "description": "The unique identifier",  
      "type": "ObjectId"  
    },  
    "location": {  
      "description": "The location of the item for GEO search",  
      "type": "object"  
    },  
    "content": {  
      "description": "The content for search",  
      "type": ["object", "string"]  
    },  
    "required": [ "_id", "content" ]  

```



Search Index Definition

```
{  
  "mappings": {  
    "dynamic": false,  
    "fields": {  
      "content": [  
        { "dynamic": true, "type": "document" },  
        { "multi": {  
          "chinese": { "analyzer": "lucene.chinese",  
                      "searchAnalyzer": "lucene.chinese",  
                      "type": "string"  
                    },  
          "english": { "analyzer": "lucene.english",  
                      "searchAnalyzer": "lucene.english",  
                      "type": "string"  
                    }  
        },  
        "type": "string"   
      ],  
      "location": { "type": "geo" }  
    }  
  }  
}
```



Edit "default" for search.data

[Learn How to Refine your Index](#) 



Index Analyzer	<i>Creates searchable terms from data to be indexed.</i>	lucene.standard	
Search Analyzer	<i>Parse \$search queries into searchable terms.</i>	lucene.standard	
Dynamic Mapping	<i>Automatically index common data types in a collection</i>		
Store Full Document	<i>Make all data available for lookup on Atlas Search side. See use cases and Performance Considerations for details.</i>		

Field Mappings

Configure specific fields within your search index with data types, analyzers, and input parameters. Field-level definitions will override the defaults above.

» content	Data Types: String, Multi, Multi	Dynamic Mapping: On	Store Original Value: Off	
» location	Data Types: Geo	Dynamic Mapping: Off	Store Original Value: Off	

 [Add Another Field](#)



Field Name

If the field is a nested document, enter the full path name (e.g., name.first)

[DELETE FIELD](#)

Enable Dynamic Mapping

Automatically index all fields in a document at this path.



Store Original Value

Save the exact value at this path for data lookup. See [Performance Considerations](#) for details.



Data Type Configuration

Specify data type(s), analyzers, and other optional input parameters for this field. See [Index Definition Docs](#) for details.



[DELETE DATA TYPE](#)

Index Analyzer	<i>Creates searchable terms from data to be indexed.</i>	lucene.standard	
Search Analyzer	<i>Parse \$search queries into searchable terms.</i>	lucene.standard	
Index Options	<i>Specify information to index</i>	offsets	
Store	<i>Index exact document text</i>	true	
Ignore Above	<i>Set character limit for indexing</i>	None	
Norms	<i>Include field length in result for scoring</i>	include	

> Multi

> Multi



▼	Multi	▼	chinese	DELETE DATA TYPE
---	-------	---	---------	----------------------------------

Index Analyzer	<i>Creates searchable terms from data to be indexed.</i>	lucene.chinese	▼
Search Analyzer	<i>Parse \$search queries into searchable terms.</i>	lucene.chinese	▼
Index Options	<i>Specify information to index</i>	offsets	▼
Store	<i>Index exact document text</i>	true	▼
Ignore Above	<i>Set character limit for indexing</i>	None	
Norms	<i>Include field length in result for scoring</i>	include	▼

▼	Multi	▼	english	DELETE DATA TYPE
---	-------	---	---------	----------------------------------

Index Analyzer	<i>Creates searchable terms from data to be indexed.</i>	lucene.english	▼
Search Analyzer	<i>Parse \$search queries into searchable terms.</i>	lucene.english	▼
Index Options	<i>Specify information to index</i>	offsets	▼
Store	<i>Index exact document text</i>	true	▼
Ignore Above	<i>Set character limit for indexing</i>	None	
Norms	<i>Include field length in result for scoring</i>	include	▼



Dynamic Mapping

If set to true, Atlas Search recursively indexes all fields and embedded documents in the document except:

Fields of certain data types. To learn more, see [BSON Data Types](#).

Any fields explicitly excluded by the `mappings.fields` parameter.

If set to false, you must specify individual fields to index using `mappings.fields`.

If omitted, defaults to false.

Wildcard query

```
{ 'text': {  
  'query': "Tenants Service",  
  'path': [{wildcard': '*' } ]  
}}
```

Pros: easy to use, use all index defined, match everything dynamically, therefore no need to consider the data structure

Cons: Relevancy of the results is depends on the data structure and quality

The screenshot shows two separate search results pages from a system.

Top Search Results: The search term is "Tenants Service". The results are as follows:

- matched path: content.District Tenancy Management Office/Estate Office.en
Yuen Long District **Tenancy** Management Office (2)
G/F, Yan Lok House, Tin Yan Estate, Tin Shui Wai
Telephone: 3152 2672
Fax: 3152 3780
- matched path: content.Further Information.en
Tenancy Management: China Overseas Property **Services** Ltd.

Search score: 2.5975499153137207

Bottom Search Results: The search term is "租約事務". The results are as follows:

- matched path: content.District Tenancy Management Office/Estate Office.en
Tai Po, North & Shatin District **Tenancy** Management Office (2)
Market Podium, Pok Hong Estate, Shatin
Telephone : 2648 1312
Fax : 2649 1648
- Tenants Service** Management Office:
China Overseas Property **Services** Ltd.

Search score: 2.5975499153137207

Bottom Left Search Results: The search term is "租約事務". The results are as follows:

- matched path: content.Name of Block(s).en
Yung Sui House< **br**>Yung **We** House< **br**>Ancillary Facilities Block
- matched path: content.Region Name.zh-Hans
新界

matched path: content.Region Name.zh-Hant
新界

matched path: content.Carpark Management.zh-Hant
由私營機構擁有及營運

matched path: content.Carpark Management.zh-Hans
由私營机构拥有及营运

matched path: content.District Name.zh-Hans



Wildcard query

The object containing the wildcard character * to match any character in the name of the field to search, including nested fields. A wildcard path:

- Must be defined as an object.
- Cannot contain the value or multi option.
- Cannot contain multiple consecutive wildcard characters such as **.
- Wildcard path is only accepted by the following operators:
 - phrase
 - regex
 - text
 - wildcard

Wildcard path is also accepted for highlighting.

Query by path

```
{ 'text': {  
    'query': "Tenants Service",  
    'path': ["content"]  
}}
```

Pros: More accurate result as you narrow down the searching range, more efficient

Cons: Exact path is required even dynamic mapping is on, thus developer need to know the schema.

The screenshot shows a search interface with two results. The first result is for 'Tenants Service' and includes a link to a detailed page. The second result is for 'School of General Nursing' and also includes a link to a detailed page. Both results show a 'matched path: content' label and a search score.

Search Results
10 results were found for the search for **Tenants Service**

"Opened in 1957 and formerly operated by the Hong Kong Tuberculosis, Chest and Heart Diseases Association, the Hospital is a major tertiary referral centre specialising in comprehensive medical treatment for adult heart and lung diseases, as well as providing geriatric medicine and palliative medicine **services**.
matched path: content
Opened in 1957 and formerly operated by the Hong Kong Tuberculosis, Chest and Heart Diseases Association, the Hospital is a major tertiary referral centre specialising in comprehensive medical treatment for adult heart and lung diseases, as well as providing geriatric medicine and palliative medicine **services**.

It also works closely with the University of Hong Kong to provide specialist training for nurses from both the public and private healthcare sectors, and to set up a Cataract Centre in 2009, providing day surgery **service** to Cataract patients.
matched path: content
The School of General Nursing has been re-opened in 2008 to educate and train up Enrolled Nurses and the hospice centre under the Li Ka Shing Foundation Hospice **Service** Programme was set up to provide holistic care to cancer patients and their families.
matched path: content
Search score: 0.873129665851593

The screenshot shows a search interface with two results for '租約事務'. The first result is for 'Queen Elizabeth Hospital (QEH)' and includes a link to a detailed page. The second result is for '伊院' and also includes a link to a detailed page. Both results show a 'matched path: content' label and a search score.

Search Results
10 results were found for the search for **租約事務**

"Queen Elizabeth Hospital (QEH) is a major acute hospital with
matched path: content
伊利沙伯醫院（伊院）於1963年啟用，是急症全科醫院，病床逾2,000張，職員約7,000名，提供24小時急症室服務，共設15個臨床部門、3間專科診所，及6間普通科門診診所。
matched path: content
伊院為各個臨床專科的病人提供高層次及高強度護理，亦在許多所設的臨床專科中，扮演著第三層導介中心的角色，特別是癌症科、胸肺及心臟科、腦神經科，及愛滋病科。
matched path: content
伊院是醫生、護士及專職醫療人員的基本及研究生訓練的教學中心，與各有關的大專院校關係密切。
matched path: content
伊院與公共及其他醫療機構合作，為本港開創了數個健康服務及專業發展，如病人資源中心、青少年醫療中心、醫院員工服務、護理教育、公眾健康教育等。
matched path: content
伊院为各个临床专科的病人提供高层次及高强度护理，亦在许多所设的临床专科中，扮演

Query by path

```
{ 'text': {  
    'query': "Tenants Service",  
    'path': ["content.District Tenancy  
Management Office/Estate  
Office.en"]  
}}
```

```
{ 'text': {  
    'query': "Tenants Service",  
    'path': ["content.District Tenancy  
Management Office/Estate  
Office.zh-Hant"]  
}}
```

Search Results
10 results were found for the search for **Tenants Service**

```
{  
  "Estate Name": {  
    "en": "Pok Hong Estate",  
    "zh-Hant": "博康邨",  
    "zh-Hans": "博康邨"  
  },  
  "District Name": {  
    "en": "Sha Tin",  
    "zh-Hant": "沙田",  
    "zh-Hans": "沙田"  
  },  
}  
  
{  
  "Estate Name": {  
    "en": "Hin Keng Estate",  
    "zh-Hant": "顯徑邨",  
    "zh-Hans": "显径邨"  
  },  
  "District Name": {  
    "en": "Sha Tin",  
    "zh-Hant": "沙田",  
    "zh-Hans": "沙田"  
  },  
}
```

matched path: content.District Tenancy Management Office/Estate Office.en
Tai Po, North & Shatin District **Tenancy** Management Office (2)
Market Podium, Pok Hong Estate, Shatin
Telephone : 2648 1312
Fax : 2649 1648

Tenants Service Management Office:
China Overseas Property **Services** Ltd.
Search score: 1.5197465419769287

matched path: content.District Tenancy Management Office/Estate Office.en
Tai Po, North & Shatin District **Tenancy** Management Office (2)
Market Podium, Pok Hong Estate, Shatin
Telephone : 2648 1312
Fax : 2649 1648

Hin Keng Estate **Tenants Service** Management Office
Creative Property **Services** Consultants Ltd.
Search score: 1.4654370546340942

Search Results
10 results were found for the search for **租約事務**

```
{  
  "Estate Name": {  
    "en": "Po Tin Estate",  
    "zh-Hant": "寶田邨",  
    "zh-Hans": "宝田邨"  
  },  
  "District Name": {  
    "en": "Tuen Mun",  
    "zh-Hant": "屯門",  
    "zh-Hans": "屯门"  
  },  
}  
  
{  
  "Estate Name": {  
    "en": "Fu Tai Estate",  
    "zh-Hant": "富泰邨",  
    "zh-Hans": "富泰邨"  
  },  
  "District Name": {  
    "en": "Tuen Mun",  
    "zh-Hant": "屯門",  
    "zh-Hans": "屯门"  
  },  
}
```

matched path: content.District Tenancy Management Office/Estate Office.zh-Hant
屯門北區租約事務管理處 屯門富泰邨服務設施大樓5樓 電話 : 2644 0201 傳真 : 2465 0561
Search score: 0.004576048813760281

matched path: content.District Tenancy Management Office/Estate Office.zh-Hant
屯門北區租約事務管理處 屯門富泰邨服務設施大樓5樓 電話 : 2644 0201 傳真 : 2465 0561

Query by path

```
{ 'text': {  
    'query': "延續護理服務",  
    'path': ["content"]  
}}
```

Problem: There are English result
while I am searching Chinese

The screenshot shows a search interface with the following details:

- Search Results**: 10 results were found for the search for 延續護理服務.
- Result 1:** 匹配路径: content
醫院專門提供老人科延續護理服務，如為內科和骨科的長者病人提供復康和療養服務
The Hospital forms part of the integrated Wong Chuk Hang Complex for the Elderly, and provides extended-care, infirmary and rehabilitation services to elderly patients who are referred to it.
黃竹坑醫院是黃竹坑老人服務綜合大樓的其中一部份，為轉介的長者提供延續護理、復康和療養服務。
Search score: 1.7339303493499756
- Result 2:** 匹配路径: content
The Hospital forms part of the integrated Wong Chuk Hang Complex for the Elderly, and provides extended-care, infirmary and rehabilitation services to elderly patients who are referred to it.
黃竹坑醫院是黃竹坑老人服務綜合大樓的其中一部份，為轉介的長者提供延續護理、復康和療養服務。
Search score: 1.726996660232544
- Result 3:** 匹配路径: content
九龍醫院提供多個專科的急症和延續護理服務，更具有本港其中一所最大的胸肺專科中心，同時為伊利沙伯醫院和基督教聯合醫院康復中的病人，提供延續護理服務。
九龍醫院是全港首數間設立康復科部門的醫院之一，也是首間提供胸肺康復計劃、精神科老人日間醫院和藥物濫用診所的醫院。

Query using analyzers

```
{ 'text': {  
    'query': "special",  
    'path': [{"value": "content", "multi":  
"english"}]  
}}
```

The screenshot shows a search results page with the title 'Search Results' and a message indicating 5 results were found for the search term 'special'. The results are listed in descending order of search score.

- matched path: content, analyzer: english**
Opened in 1957 and formerly operated by the Hong Kong Tuberculosis, Chest and Heart Diseases Association, the Hospital is a major tertiary referral centre **specialising** in comprehensive medical treatment for adult heart and lung diseases, as well as providing geriatric medicine and palliative medicine services.
Furthermore, the Hospital partners with other organisations to promote heart and chest health, and undertakes innovative research projects, striving for continuous quality breakthroughs in its areas of **specialisation**.
Search score: 1.1573679447174072
- matched path: content, analyzer: english**
Founded by the Maryknoll Sisters in 1961, the Hospital has a Catholic ethos that focuses on total care for the physical, **social**, psychological and spiritual needs of its inpatients, outpatients and ambulatory patients.
Search score: 0.817704439163208
- matched path: content, analyzer: english**
Established by the Tung Wah Group of Hospitals in 1965, this **special** programmes to care for people with chronic chest diseases, physical handicaps and complicated geriatric problems.
The Centre provides comprehensive healthcare services to the patients with end-stage organ failure and cancer to meet the physical, psycho- **social** and spiritual needs of patients and families.
Search score: 0.8120408058166504

Query using analyzers

```
{ 'text': {  
    'query': "延續護理服務",  
    'path': [{"value": "content", "multi":  
"chinese"}]  
}}
```

What if there are multi-language in same field?

The screenshot shows a search results page titled 'Search Results' with the message '10 results were found for the search for 延續護理服務'. The results are listed in two columns:

matched path: content, analyzer: chinese	matched path: content, analyzer: chinese
黃竹坑醫院是黃竹坑老人服務綜合大樓的其中一部份，為轉介的長者提供延續護理、復康和療養服務。	黃竹坑医院是黄竹坑老人服务综合大楼的其中一部份，为转介的长者提供延续护理、复康和疗养服务。
Search score: 5.0609331130981445	
matched path: content, analyzer: chinese	matched path: content, analyzer: chinese
九龍醫院提供多個專科的急症和延續護理服務，更具有本港其中一所最大型的胸肺專科中心，同時為伊利沙伯醫院和基督教聯合醫院康復中的病人，提供延續護理服務。	九龙医院提供多个专科的急诊和延续护理服务，更具有本港其中一所最大型的胸肺专科中心，同时为伊利沙伯医院和基督教联合医院康复中的病人，提供延续护理服务。
九龍醫院是全港首數間設立康復科部門的醫院之一，也是首間提供胸肺康復計劃、精神科老人日間醫院和藥物濫用診所的醫院。	九龙医院是全港首数间设立康复科部门的医院之一，也是首间提供胸肺康复计划、精神科老人日间医院和药物滥用诊所的医院。
matched path: content, analyzer: chinese	matched path: content, analyzer: chinese
醫院現正分階段進行翻新和重建，以便提供更多全面的服務，滿足社會大眾的醫療需求。	医院现正分阶段进行翻新和重建，以便提供更多全面的服务，满足社会大众的医疗需求。
matched path: content, analyzer: chinese	matched path: content, analyzer: chinese
九龍醫院提供多個專科的急症和延續護理服務，更具有本港其中一所最大型的胸肺專科中心，同時為伊利沙伯醫院和基督教聯合醫院康復中的病人，提供延續護理服務。	九龙医院提供多个专科的急诊和延续护理服务，更具有本港其中一所最大型的胸肺专科中心，同时为伊利沙伯医院和基督教联合医院康复中的病人，提供延续护理服务。
九龍醫院是全港首數間設立康復科部門的醫院之一，也是首間提供胸肺康復計劃、精神科老人日間醫院和藥物濫用診所的醫院。	九龙医院是全港首数间设立康复科部门的医院之一，也是首间提供胸肺康复计划、精神科老人日间医院和药物滥用诊所的医院。



Analyzers and Tokens

Analyzer	Tokens after Analysis
Original	José is much taller, his email is j.parillo@gmail.es.
Standard	josé much taller his email j.parillo gmail.es
Simple	josé is much taller his email is j parillo gmail es
Whitespace	José is much taller, his email is j.parillo@gmail.es
Keyword	"José is much taller, his email is j.parillo@gmail.es."(One value)
English	josé much taller hi email j.parillo@gmail.es
Spanish	josé is much taller his email is j.paril gmail.



Analyzers and Tokens

You can create a custom analyzer by configuring

- Character filters
 - Removing Accents, Standardizing characters
- Tokenizing of strings.
 - Understanding language context
- Token Filtering/Modification
 - Synonyms, Removing short words etc.

<https://www.mongodb.com/docs/atlas/atlas-search/analyzers/custom/>

Define multiple analyzers on single field



<https://www.mongodb.com/docs/atlas/atlas-search/analyzers/multi/>

```
{"multi": {  
    "chinese": {  
        "analyzer": "lucene.chinese",  
        "searchAnalyzer": "lucene.chinese",  
        "type": "string"  
    }, "english": {  
        "analyzer": "lucene.english",  
        "searchAnalyzer": "lucene.english",  
        "type": "string"  
    }  
}, "type": "string" }
```

Multi		
Index Analyzer	<i>Creates searchable terms from data to be indexed.</i>	lucene.chinese
Search Analyzer	<i>Parse \$search queries into searchable terms.</i>	lucene.chinese
Index Options	<i>Specify information to index</i>	offsets
Store	<i>Index exact document text</i>	true
Ignore Above	<i>Set character limit for indexing</i>	None
Norms	<i>Include field length in result for scoring</i>	include

Multi		
Index Analyzer	<i>Creates searchable terms from data to be indexed.</i>	lucene.english
Search Analyzer	<i>Parse \$search queries into searchable terms.</i>	lucene.english
Index Options	<i>Specify information to index</i>	offsets
Store	<i>Index exact document text</i>	true
Ignore Above	<i>Set character limit for indexing</i>	None
Norms	<i>Include field length in result for scoring</i>	include

Note: The multi path option is available only to fields of type string. Sub-document is not available!!

Query using multiple analyzers

```
{ 'text': {  
    'query': "出嚟食lunch",  
    'path': [{"value": "content", "multi": "chinese"}, {"value": "content", "multi": "english"}]  
}}
```

The screenshot displays a search results page with three distinct sections, each representing a different analyzer path:

- matched path: content, analyzer: chinese**

"我哋講嘅嘅時候亦都好鍾意中英夾雜，例如你會叫個同事「print 份 report 出嚟」
「scan 啟份 document」、「出嚟食個 lunch」、「book 場」、「去 site out [trip] 等等……對於香港人，呢類型嘅表達方式正常不過，一聽就曉香港生活過嘅人，就真係可能「拗唔頭」，而香港人卻會經常不自覺咁嘅說話模式都不自知，毫不造作，是真性情嘅表達。
- matched path: content, analyzer: english**

"我哋講嘅嘅時候亦都好鍾意中英夾雜，例如你會叫個同事「print 份 report」「scan 啟份 document」、「出嚟食個 lunch」、「book 場」、「去 site out [trip] 等等……對於香港人，呢類型嘅表達方式正常不過，一聽就曉香港生活過嘅人，就真係可能「拗唔頭」，而香港人卻會經常不自覺咁嘅說話模式都不自知，毫不造作，是真性情嘅表達。
- Search score: 5.31962251663208**
- matched path: content, analyzer: chinese**

"Established following the enactment of the Tung Wah Hospital Act in 1870, the hospital provided medical services to inpatients and outpatients. Its subsequent launch of a pilot Integrated Chinese and Western Medicine Training Program in 1999 marked a milestone in Hong Kong's public hospital services.
- matched path: content, analyzer: english**

"Its subsequent launch of a pilot Integrated Chinese and Western Medicine Training Program in 1999 marked a milestone in Hong Kong's public hospital services.
- Search score: 2.331730842590332**
- matched path: content, analyzer: chinese**

"最後，由細到大都勁中英夾雜，程度係聽到casual 講嘢、但用正式中文翻譯（尤其他音譯外文 terms，你試吓將LV講做路易威登？）
- matched path: content, analyzer: chinese**

"最後，由細到大都勁中英夾雜，程度係聽到casual 講嘢、但用正式中文翻譯（尤其他音譯外文 terms，你試吓將LV講做路易威登？）

Query using multiple analyzers

```
{ 'text': {  
    'query': "出嚟食lunch",  
    'path': [{"value": "content", "multi":  
      "english"}]  
}}
```

The screenshot shows a search interface with three results listed. Each result includes the matched path, analyzer used, and a snippet of the document text.

- matched path: content, analyzer: english**
我哋講嘅時候亦都好鍾意中英夾雜，例如你會叫個同事「print 份 report 出嚟」、
「scan 咗份 document」、「出嚟食個lunch」、「book 場」、「去 shopping」、「
出 trip」等等……對於香港人，呢類型嘅表達方式正常不過，一聽就明，對於有過
香港生活過嘅人，就真係可能「拗唔頭」，而香港人卻會經常不自覺咁樣用咗呢個類型
嘅說話模式都不自知，毫不造作，是真性情嘅表達。
Search score: 2.0757384300231934
- matched path: content, analyzer: english**
Its subsequent launch of a pilot Integrated Chinese and Western Medicine Treatment
Centre in which Chinese and Western medicine practitioners collaborate to provide services
to in-patients marked a milestone in Hong Kong public hospital services.
Search score: 1.173741102218628
- matched path: content, analyzer: english**
Opened in 1957 and formerly operated by the Hong Kong Tuberculosis, Chest and Heart
Diseases Association, the Hospital is a major tertiary referral centre specialising in
comprehensive medical treatment for adult heart and lung diseases, as well as providing
geriatric medicine and palliative medicine services.
It also works closely with the University of Hong Kong to provide specialist training for nurses
from both the public and private healthcare sectors, and to set up a Cataract Centre in 2009,
providing day surgery service to Cataract patients.
Furthermore, the Hospital partners with other organisations to promote heart and chest
health, and undertakes innovative research projects, striving for continuous quality
breakthroughs in its areas of specialisation.

Query using multiple analyzers

```
{ 'text': {  
    'query': "出嚟食lunch",  
    'path': [{"value": "content", "multi":  
      "chinese"}]  
}}
```

The screenshot shows a search results page with the title 'Search Results' and a message indicating 10 results were found for the search term '出嚟食lunch'. The results are listed in descending order of search score.

- matched path: content, analyzer: chinese**
我聽講嘅時候亦都好鐘意中英夾雜，例如你會叫個同事「print 份 report 出嚟」、「scan 咗份 document」、「出嚟食個lunch」、「book 場」、「去 shopping」、「出 trip」等等……對於香港人，呢類型嘅表達方式正常不過，一聽就明，對於冇細心聽過嘅人，就真係可能「拗唔頭」，而香港人卻會經常不自覺咁樣用咗呢個類型嘅說話模式都不自知，毫不造作，是真性情嘅表達。
Search score: 3.2438840866088867
- matched path: content, analyzer: chinese**
最後，由細到大都勤中英夾雜，程度係聽得到casual講嘅、但用正式中文翻譯（尤其音譯外文terms，你試吓將LV講做路易威登？）
會覺得好 cringe 老嫗。
- matched path: content, analyzer: chinese**
但著作嘅講，香港人嘅「我手寫我口」就只能自己 enjoy，好難發揚光大，可能要中英、口語二語出版？
- matched path: content, analyzer: chinese**
Sir Walter Scott 用 Scottish dialect 寫作，自成一家，一樣無損佢影響力；曾經嘅周星馳港式笑話都有唔少外圍影迷欣賞，究竟香港幾時可以再輸出文化？
Search score: 1.6440656185150146
- matched path: content, analyzer: chinese**
Established following the enactment of the Tung Wah Hospital :
Its subsequent launch of a pilot Integrated Chinese and Western Medicine Treatment Centre in which Chinese and Western medicine practitioners collaborate to provide services to in-patients marked a milestone in Hong Kong public hospital services.
Search score: 1.157989740371704



Search Operators

```
//This finds Bikes  
  
query = [ { $search: { text: { query: "hike", path: "activity" ,  
fuzzy : { maxEdits:1}} } }]  
  
db.example.aggregate(query).pretty()  
  
//This does not find the Document - why is that?  
  
db.example.insertOne({activity:"tracking  
Elephant",hours:3,people:["Alan Quartermain"]})  
  
//Note the mistake in the query - a typo by our user  
  
query = [ { $search: { text: { query: "Elephant", path: "activity" ,  
fuzzy : { maxEdits:2}} } }]  
  
db.example.aggregate(query).pretty()
```



Operator	Description
autocomplete	Performs a search-as-you-type query from an incomplete input string.
compound	Combines other operators into a single query.
equals	Works in conjunction with the boolean and objectId data types.
exists	Tests for the presence of a specified field.
geoShape	Queries for values with specified geo shapes.
geoWithin	Queries for points within specified geographic shapes.
near	Queries for values near a specified number, date, or geo point.
phrase	Searches documents for terms in an order similar to the query.
queryString	Supports querying a combination of indexed fields and values.
range	Queries for values within a specific numeric or date range.
regex	Interprets the <code>query</code> field as a regular expression.
search	Deprecated. Performs analyzed search. Use <code>text</code> operator instead.
span	Specifies relative positional requirements for query predicates within specified regions of a text field.
term	Deprecated. Performs unanalyzed search.
text	Performs textual analyzed search.
wildcard	Supports special characters in the query string that can match any character.



Search Operators

Using the compound operator we can combine search operations with boolean logic

must - All clauses have to match, equivalent to AND

mustNot - None of the clauses are allowed to match , equivalent to NOT AND

should - N terms must match from the set, where N defaults to 0

filter - The same as *must*, but does not affect the score.

Advanced Features





Autocomplete

A screenshot of a Google search interface. The search bar at the top contains the query "how do i get mongodb". Below the search bar is a list of ten autocomplete suggestions, each preceded by a small blue square icon:

- mongodb in terminal
- mongodb from inserting duplicate records
- mongodb on windows
- mongodb on mac
- mongodb on linux
- mongodb on windows 10
- mongodb in ubuntu
- mongodb locally
- mongodb compass
- mongodb logs



Autocomplete

The screenshot shows the eBay homepage with a search bar containing the partial query "dera". Below the search bar is a dropdown menu displaying a list of suggestions:

- derailleur hanger
- derailleur hanger – Cycling Equipment
- derailleur
- deranged
- derailleur hanger extender
- derailleur hanger alignment tool
- rear derailleur
- front derailleur
- derailleur jockey wheel
- de razor blade
- shimano rear derailleur

The eBay header includes a "GET 20% OFF EVERYTHING" banner with a "Use Code: SAVE20" button, a "Fast & Free" delivery badge, and a "Shop Now >" button. The sidebar features a "Grab mighty savings from small..." promotional section and a "Start shopping →" button.



Four tips for using Autocomplete

Consider the index size - it's easy to make one very large.

nGrams create larger indexes than edgeGrams.

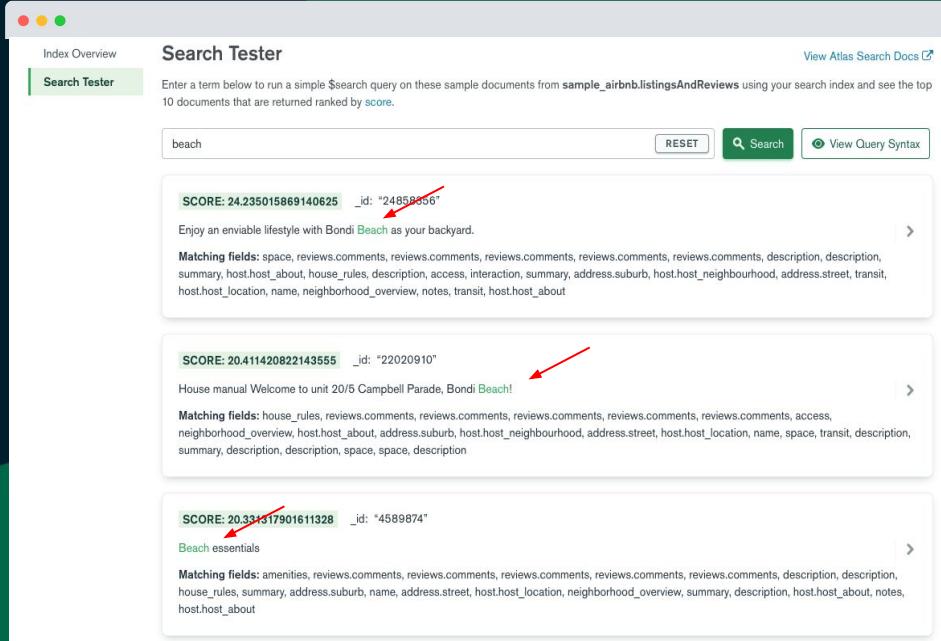
The smaller the minimum gram size - the larger the index.

To put exact matches at the top combine with a phrase query.

Highlighting

Highlighting is way of showing the user 'Why' their search matched that document.

It is particularly important where queries are against larger corpuses of text where the user may want to 'Jump to match'





Highlighting shows what was matched, not what you searched for.

Highlighting adds latency - 30ms or more to a result and uses more CPU

If you do not need highlighting you can make indexes smaller by removing the store option.

Highlighting does not currently work with autocomplete indexes.



Synonyms

Allow alternative words to find data that means the same thing.

Powerful context sensitive search capability.

Atlas implementation makes it easy to define synonyms.

Pharmacy

Chemist

Druggerie

Apothecary

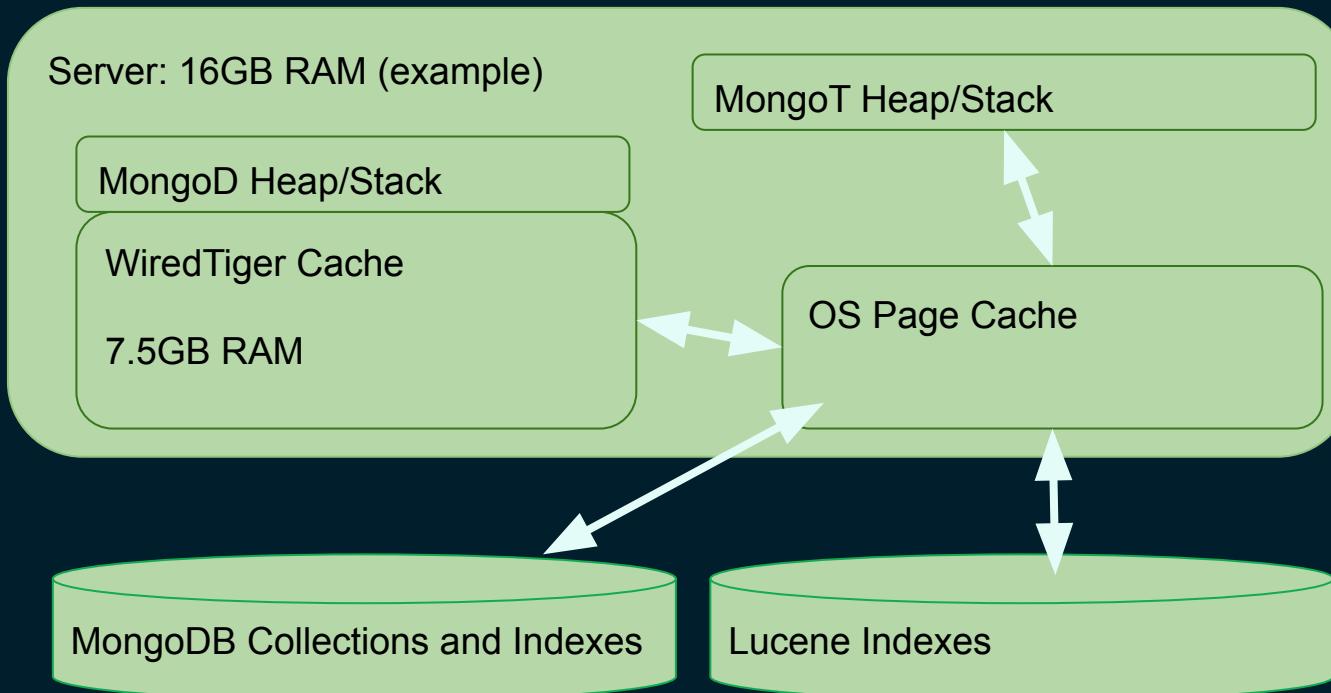


Performance and Sizing



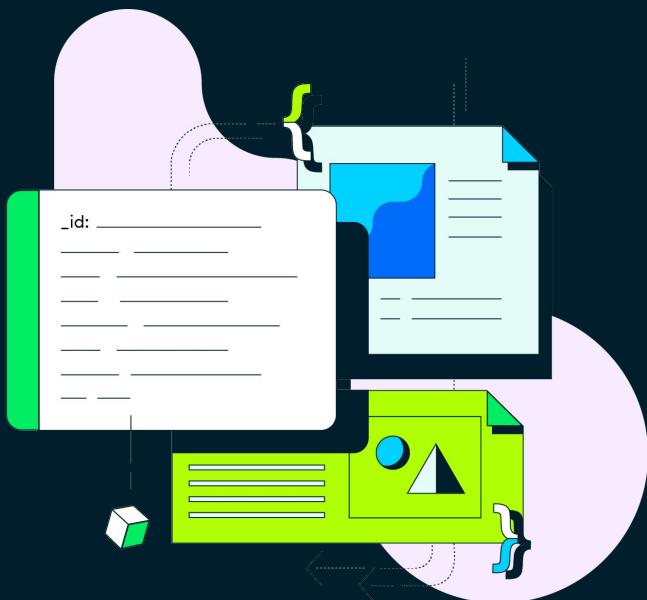


Factors that impact search performance





Index availability and consistency.



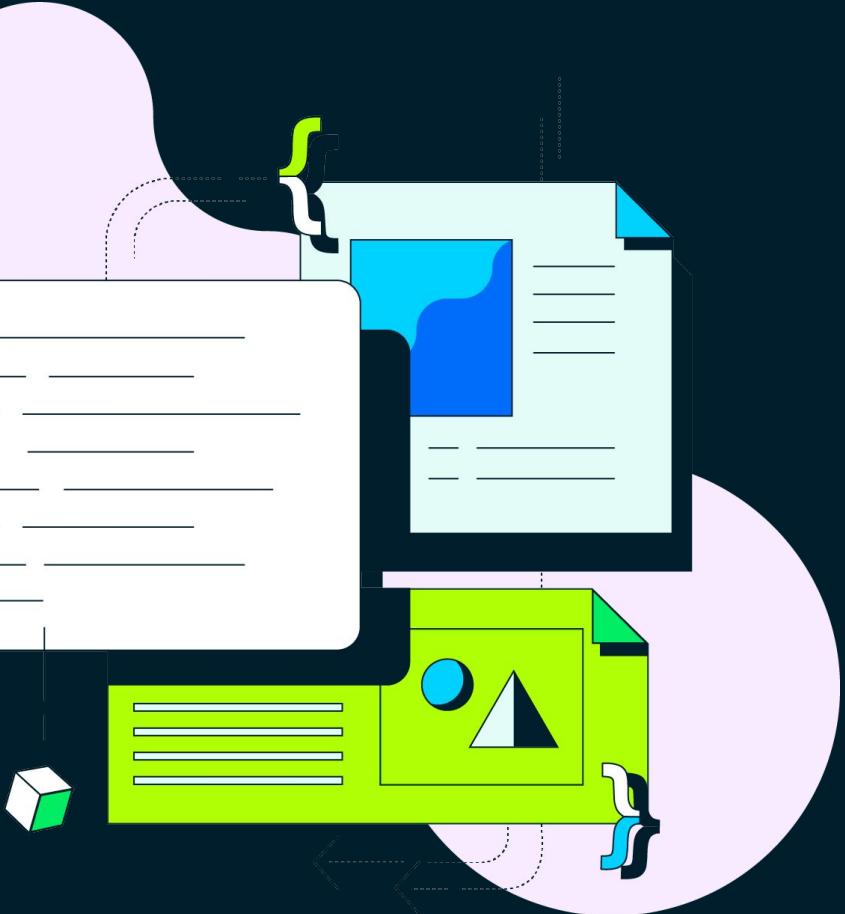
Indexing is Asynchronous

Indexing could fall a long way behind if the write load is high

An index will always be available even when re-indexing.



- Using a \$search then a \$match or \$sort in the pipeline is much slower
 - Combine the \$match into the \$search using filter
 - Use the near with boost operator to score in a way that is the same as your \$sort
- In-memory sort will use a lot of RAM and reduce Lucene index speed.
- Avoid using Dynamic mapping on large documents with many keys
- Results are sorted by score so larger result sets need more resources.
- On a sharded cluster - all \$search queries are sent to all shards.





Helpful Resources

Atlas Search Basics:

<https://www.youtube.com/watch?v=V-icerXJVZ8&t=8s>

Atlas Search Docs:

<https://www.mongodb.com/docs/atlas/atlas-search/>

Atlas Search Restaurant Project (and video):

<http://atlassearchrestaurants.com/>

<https://www.youtube.com/watch?v=s2kXYZRE7pA>



Atlas Free/Shared Cluster Limitations:

<https://www.mongodb.com/docs/atlas/atlas-search/limitations/#atlas-search-free-and-shared-cluster-limitations>

Performance Considerations:

<https://www.mongodb.com/docs/atlas/atlas-search/performance/#performance-considerations>

Atlas Search Changelog:

<https://www.mongodb.com/docs/atlas/atlas-search/changelog/>



Thank you!