



<  
Platform  
<  
Use cases

Try free

Have questions? [Contact us](#)

Already have an account? [Log in](#)

Elasticsearch Guide:

8.6 (current) ▾

- What is Elasticsearch? >
- What's new in 8.6
- Set up Elasticsearch >
- Upgrade Elasticsearch >
- Index modules >
- Mapping >
- Text analysis >
- Overview**
- Concepts >
- Configure text analysis >
- Built-in analyzer reference >
- Tokenizer reference >
- Token filter reference >
- Character filters reference >
- Normalizers
- Index templates >
- Data streams >
- Ingest pipelines >
- Aliases
- Search your data >
- Query DSL >
- Aggregations >

[Elastic Docs](#) > [Elasticsearch Guide \[8.6\]](#) > [Text analysis](#)

## Text analysis overview

[edit](#)

Text analysis enables Elasticsearch to perform full-text search, where the search returns all *relevant* results rather than just exact matches.

If you search for `Quick fox jumps`, you probably want the document that contains `A quick brown fox jumps over the lazy dog`, and you might also want documents that contain related words like `fast fox` or `foxes leap`.

### Tokenization

[edit](#)

Analysis makes full-text search possible through *tokenization*: breaking a text down into smaller chunks, called *tokens*. In most cases, these tokens are individual words.

If you index the phrase `the quick brown fox jumps` as a single string and the user searches for `quick fox`, it isn't considered a match. However, if you tokenize the phrase and index each word separately, the terms in the query string can be looked up individually. This means they can be matched by searches for `quick fox`, `fox brown`, or other variations.

### Normalization

[edit](#)

Tokenization enables matching on individual terms, but each token is still matched literally. This means:

- A search for `Quick` would not match `quick`, even though you likely want either term to match the other
- Although `fox` and `foxes` share the same root word, a search for `foxes` would not match `fox` or vice versa.
- A search for `jumps` would not match `leaps`. While they don't share a root word, they are synonyms and have a similar meaning.

To solve these problems, text analysis can *normalize* these tokens into a standard format. This allows you to match tokens that are not exactly the same as the search terms, but similar enough to still be relevant. For example:

- `Quick` can be lowercased: `quick`.

1/11/23, 5:24 PM		Text analysis overview   Elasticsearch Guide [8.6]   Elastic	
EQL	>	<ul style="list-style-type: none"><li>foxes can be <i>stemmed</i>, or reduced to its root word: fox .</li><li>jump and leap are synonyms and can be indexed as a single word: jump .</li></ul>	
SQL	>	To ensure search terms match these words as intended, you can apply the same tokenization and normalization rules to the query string. For example, a search for Foxes leap can be normalized to a search for fox jump .	
Scripting	>		
Data management	>		
Autoscaling	>	<h2>Customize text analysis</h2>	<a href="#">edit</a>
Monitor a cluster	>	Text analysis is performed by an <i>analyzer</i> , a set of rules that govern the entire process.	
Roll up or transform your data	>	Elasticsearch includes a default analyzer, called the <i>standard analyzer</i> , which works well for most use cases right out of the box.	
Set up a cluster for high availability	>	If you want to tailor your search experience, you can choose a different <i>built-in analyzer</i> or even <i>configure a custom one</i> . A custom analyzer gives you control over each step of the analysis process, including:	
Snapshot and restore	>	<ul style="list-style-type: none"><li>Changes to the text <i>before</i> tokenization</li><li>How text is converted to tokens</li><li>Normalization changes made to tokens before indexing or search</li></ul>	
Secure the Elastic Stack	>		
Watcher	>		
Command line tools	>		
How to	>	<a href="#">« Text analysis</a>	<a href="#">Text analysis concepts »</a>
Troubleshooting	>		
REST APIs	>		
Migration guide	>		
Release notes	>		
Dependencies and versions			

On this page

- Tokenization
- Normalization
- Customize text analysis

Register now for free for ElasticON Global 2023

Our biggest user conference of the year is happening March 7-8. Join us virtually for all new technical deep dives, live workshops, demos, and more!

Learn more

Subscribe to our newsletter

Email address

Sign up

Follow us

PRODUCTS & SOLUTIONS

- Enterprise Search
- Observability
- Security
- Elastic Stack
- Elasticsearch
- Kibana
- Integrations
- Subscriptions
- Pricing

COMPANY

- Careers
- Board of Directors
- Contact

RESOURCES

- Documentation
- What is the ELK Stack?
- What is Elasticsearch?
- Migrating from Splunk
- OpenSearch vs. Elasticsearch
- Public Sector



[Trademarks](#) | [Terms of Use](#) | [Privacy](#) | [Sitemap](#)

© 2023. Elasticsearch B.V. All Rights Reserved

Elasticsearch is a trademark of Elasticsearch B.V., registered in the U.S. and in other countries.

Apache, Apache Lucene, Apache Hadoop, Hadoop, HDFS and the yellow elephant logo are trademarks of the Apache Software Foundation in the United States and/or other countries.