

### Analyzer

A Combination of **Character Filters**, **Tokenizers**, **Token Filters**

### Processing pipeline

**Input** --String--> **Character Filters** --String--> **Tokenizer** --Tokens--> **Token Filters** --Tokens--> **Output**

### Character Filters

The Character filter removes words from the input Strings.

e.g. **html\_strip char filter**:

**Input**: "<p> An example <br></p>"

**Output**: "An example"

### Tokenizer

The way the input is divided into parts (tokens)

e.g. **whitespace tokenizer**

**Input**: *The quick brown fox*

**Tokens**: [*The, quick, brown, fox*].

### Token Filters

A Tokenfilter returns a subset of given tokens.

e.g. **stop word filter** that removes all the **stop words** from the tokens, such as "*the*", "*and*", ...



By **hajuko**  
[cheatography.com/hajuko/](https://cheatography.com/hajuko/)

Not published yet.  
Last updated 17th February, 2015.  
Page 1 of 1.

Sponsored by **Readability-Score.com**  
Measure your website readability!  
<https://readability-score.com>