V-Spark 3.4.3 Regular Expression Guide

V-Spark 3.4.3 Regular Expression Guide

© Copyright 2019 Voci All Rights Reserved.

The information contained in this document is the proprietary and confidential information of Voci Technologies, Inc. incorporated. you may not disclose, provide or make available this document, or any information contained in this document, to any third party, without the prior written consent of Voci.

The information in this document is provided for use with V-Spark Voice Analytics. No license, express or implied, to any intellectual property associated with this document or such products is granted by this document.

All Voci Technologies, Inc. products described in this document, including V-Spark and others prefaced by Voci are owned by Voci (or those companies that have licensed technology to Voci) and are protected by patents, trade secrets, copyrights or other industrial property rights. The Voci products described in this document may still be in development. The final form of each product and release date thereof is at the sole and absolute discretion of Voci. Your purchase, license and/or use of Voci products shall be subject to Voci's then current sales terms and conditions.

Trademarks

The following terms used in this document are trademarks of Voci Technologies, Inc. in the United States and other countries:

- Voci
- V-Blaze
- V-Cloud
- V-Discovery
- V-Ferno
- V-Purify
- V-Spark
- Voci

Other third party disclaimers or notices may be set forth in Voci's online or printed documentation. All other product and service names, and trademarks not owned by Voci are the property of their respective owners.

Table of Contents

1. Overview 1
    1.1. Anchoring 1
    1.2. Allowed Characters 1
2. Regular Expression Search Characters 3

Chapter 1. Overview

Regular expression (regex) queries search text using pattern matching. V-Spark regex syntax is briefly described below. Regex searches can be performed in the Dashboard File View by selecting the Regex option for search text. Regex queries in V-Spark will only match whole terms as with plain text queries. There are no partial matches.

Note

Regex queries operate on individual terms and cannot be used to match multiword phrases. For each regex query, the search engine scans the list of terms in the inverted index to find all matching terms, and retrieves all documents for each term. This means that running a regex query that matches many unique terms can be very resource intensive. Users should avoid using a pattern that starts with a wildcard (for example, .*foo).

For more information on regex syntax used in V-Spark, see Elasticsearch 1.4 Documentation [https://www.elastic.co/guide/en/elasticsearch/reference/1.4/query-dsl-regexp-query.html].

1.1. Anchoring

Most regex search engines will match any part of a word. In these cases ^ and $ are used to anchor searches to the beginning and end of a word, respectively. However, since V-Spark regex searches will only match whole words, these special anchors are not required and not valid except as literal characters.

As an example, for the word "abcde":

```
ab.*    # match
abcd    # no match
^abcd   # no match
abcd$   # no match
```

1.2. Allowed Characters

Any Unicode characters may be used in the pattern, but certain characters are reserved. The reserved characters are:

```
. ? + * | { } [ ] ( ) # @ & < > ~" \
```

Note

^ and $ are not reserved characters.

Chapter 2. Regular Expression Search Characters

Character	Meaning	Example Text	Example Queries	Match?
.	Match any character	ab	a.	Yes
+	Match preceding shortest pattern 1 or more times	aaabbb aaabbb	a+b+ a+b+c+	Yes No
*	Match the preceding shortest pattern 0 or more times	aaabbb	a*b* a*b*c*	Yes No
?	Match the preceding shortest pattern 0 or one times	aaabbb	aaabbbc?	Yes
{n}	Indicate a minimum number of times the preceding shortest pattern should match	aaabbb	a{3}b{3} a{4}a{4}	Yes No
{n,n}	Indicate a minimum and maximum number of times the preceding shortest pattern should match	aaabbb	a{2,3}b{2}	Yes
()	Group characters to form subpatterns	ababab	(ab)+	Yes
|	OR (applies to the *longest* pattern on either side)	aaabbb	a+|ccc aaa|bbb	Yes No
[]	Indicate lists or ranges of characters. ^ negates characters	abc	[a-z]* [^a-z]	Yes No
~	Negate the following shortest pattern	abcdef	abc~df ab~ef	Yes No
<>	Range of numeric values	99	<1-100> <100-101> <99-100>	Yes No Yes

Character	Meaning	Example Text	Example Queries	Match?
&	AND - text must meet conditions on both sides	aaabbb	aaa.+&.+bbb aaa&bbb	Yes No
@	Match any word	hello	@ @&~(hello)	Yes No

 © 2019, Voci Technologies, Inc.