

Archivematica - Feature #1889

Add underscores and periods as tokenizers to value list in ElasticSearch

10/03/2012 12:00 AM - Evelyn McLellan

Status:	New	Start date:	
Priority:	Low	Due date:	
Assignee:	Mike Cantelon	% Done:	0%
Category:	Index/Search	Estimated time:	0.00 hour
Target version:		Pull Request:	
Google Code Legacy ID:	archivematica-1234	Requires documentation:	
Sponsored:	No		

Description

See discussion at <https://groups.google.com/forum/?hl=en&fromgroups=#!topic/archivematica/u1TSa2i9B.JM> .

[g] Legacy categories: Search

History

#1 - 11/03/2012 12:00 AM - Mike Cantelon

Yeah, mentioning to users that they can do wildcard searches might be good rather than changing the tokenizer.

If a user enters ""*master*" or "Members_Master2009.xls" now they'd end up seeing the entry for Members_Master2009.xls. If we change the tokenizer so it splits words up by underscores, etc., we might lose the ability to search for "Members_Master2009.xls" precisely.

I've posted a question online to see if anyone has any tokenizer advice on how we could tokenize "Members_Master_2009.xls" into "members", "master", and "Members_Master_2009.xml".

#2 - 12/17/2012 02:41 PM - Mike Cantelon

- Sponsored set to No

Bump to 1.1?

#3 - 12/17/2012 05:08 PM - Mike Cantelon

- Target version changed from Release 0.10-beta to Release 1.0.0

#4 - 02/18/2013 12:51 PM - Evelyn McLellan

- Category set to Index/Search

#5 - 05/30/2013 04:37 PM - Mike Cantelon

Updating analysis then testing...

```
curl -XPOST 'http://192.168.1.70:9200/aips/_close'
curl -XPUT 'http://192.168.1.70:9200/aips/_settings' -d '{
  "index": {
    "analysis" : {
      "analyzer": {
        "default": {
          "tokenizer" : "standard",
          "filter" : ["preserve_hyphens_filter", "lowercase", "stop"]
        },
      }
    }
  }
}
```

```
"filter" : {  
  "preserve_hyphens_filter" : {  
    "type" : "word_delimiter",  
    "generate_word_parts": false,  
    "catenate_words": true  
  }  
}  
}  
}  
}  
}
```

```
curl -XPOST 'http://192.168.1.70:9200/aips/_open'
```

```
curl -XGET '192.168.1.70:9200/aips/_analyze?field=msg&pretty=1' -d "Run to the hills and rock-around-town."
```

#6 - 10/07/2013 12:52 PM - Courtney Mumma

- Target version changed from Release 1.0.0 to Release 1.1.0

#7 - 03/18/2014 04:05 PM - Justin Simpson

- Target version deleted (Release 1.1.0)

We do not have advanced search functionality on the Archivematica roadmap at the present time. We would like to provide a better search interface, at some point, I am moving this ticket out of any targetted version queues until the feature can have requirements generated.

#8 - 09/15/2016 09:37 PM - Justin Simpson

- Priority changed from High to Low