|  |  |
| --- | --- |
| **How to identify records in Alma which have corrupt (invalid Unicode) characters** |  |
|  |



**Yoel Kortick**

**Senior Librarian**

For various reason there may be records in Alma which have invalid characters. Typically, this occurs if a record with invalid characters is imported to Alma, or if it arrived to Alma from a migration and already had invalid characters.

These records can be easily identified as follows.

**ONE**

In Alma metadata editor navigate to the indication rules community folder and find sample rule "EXL – Find corrupt characters YLK" and copy to your institution.



**TWO**

Check contents of rule and if desired amend



|  |
| --- |
| rule "Contains \_\_ (underline) character in 0XX or 1XX or 2xx or 3xx or 4xx or 5xx or 6XX or 7XX or 8xx or 9XX fields"when ((exists "0\*\*.\*.\*\_\_\*") OR (exists "1\*\*.\*.\*\_\_\*") OR (exists "2\*\*.\*.\*\_\_\*") OR (exists "3\*\*.\*.\*\_\_\*") OR (exists "4\*\*.\*.\*\_\_\*") OR (exists "5\*\*.\*.\*\_\_\*") OR (exists "6\*\*.\*.\*\_\_\*") OR (exists "7\*\*.\*.\*\_\_\*") OR (exists "8\*\*.\*.\*\_\_\*") OR (exists "9\*\*.\*.\*\_\_\*"))thenset indication."true"endrule "Contains <U+ in 0XX or 1XX or 2xx or 3xx or 4xx or 5xx or 6XX or 7XX or 8xx fields or 9xx fields"when ((exists "0\*\*.\*.\*<U+\*") OR (exists "1\*\*.\*.\*<U+\*") OR (exists "2\*\*.\*.\*<U+\*") OR (exists "3\*\*.\*.\*<U+\*") OR (exists "4\*\*.\*.\*<U+\*") OR (exists "5\*\*.\*.\*<U+\*") OR (exists "6\*\*.\*.\*<U+\*") OR (exists "7\*\*.\*.\*<U+\*") OR (exists "8\*\*.\*.\*<U+\*") OR (exists "9\*\*.\*.\*<U+\*"))thenset indication."true"endrule "Contains double dagger subfield sign ǂ in 0XX or 1XX or 2xx or 3xx or 4xx or 5xx or 6XX or 7XX or 8xx or 9xx fields"when ((exists "0\*\*.\*.\*ǂ\*") OR (exists "1\*\*.\*.\*ǂ\*") OR (exists "2\*\*.\*.\*ǂ\*") OR (exists "3\*\*.\*.\*ǂ\*") OR (exists "4\*\*.\*.\*ǂ\*") OR (exists "5\*\*.\*.\*ǂ\*") OR (exists "6\*\*.\*.\*ǂ\*") OR (exists "7\*\*.\*.\*ǂ\*") OR (exists "8\*\*.\*.\*ǂ\*") OR (exists "9\*\*.\*.\*ǂ\*"))thenset indication."true"endrule "Contains funny question mark in 0XX or 1XX or 2XX or 3XX or 4XX or 5XX or 6XX or 7XX or 8xx or 9xx fields"when ((exists "0\*\*.\*.\*�\*") OR (exists "1\*\*.\*.\*�\*") OR (exists "2\*\*.\*.\*�\*") OR (exists "3\*\*.\*.\*�\*") OR (exists "4\*\*.\*.\*�\*") OR (exists "5\*\*.\*.\*�\*") OR (exists "6\*\*.\*.\*�\*") OR (exists "7\*\*.\*.\*�\*") OR (exists "8\*\*.\*.\*�\*") OR (exists "9\*\*.\*.\*�\*"))thenset indication."true"end |

**THREE**

Make a set in Alma and then filter it. Here we made a set "Titles to check" and then choose "filter set".



Use the above indication rule which we copied from the community and call new set whatever you want (here we call it "Titles with invalid characters")



**FOUR**

See results of new set



Here for example are two records with an invalid character, so we see that the process has worked:



