

Edit - bug #7304

Add warning to parsed references if reference type is unclear

03/07/2018 09:54 AM - Andreas Müller

Status:	New	Start date:	03/07/2018
Priority:	New	Due date:	
Assignee:	Andreas Müller	% Done:	0%
Category:	cdmlib	Estimated time:	0.00 hour
Target version:	Unassigned CDM tickets	Found in Version:	Release 4.14
Severity:	normal		

Description

AM:

ich habe mir das jetzt nochmal im Parser angeschaut. Der Fall

xxx in Acta Univ.Carol., Biol. 43(2): 105. 1999

ist wirklich einer der schwierig zu parsenden. Das hängt mit dem Komma im Titel zusammen. Das ist auch die Standardform für Booksections mit dem Autor vor dem Komma. Bisher war die einfache Regel, wenn das Komma spätestens an der 4. Stelle kommt, der Autor also maximal aus 3 Teilen besteht, dann wird der Teil vor dem Komma als Autor interpretiert, ansonsten gehen wir von einem Journal Title aus. Das scheint hier nicht zu passen.

Ich habe das jetzt mal auf die 3. Stelle reduziert, womit zumindest dieser Fall abgedeckt wird. Ich hoffe, dass dadurch nicht zu viele richtige Autoren NICHT erkannt werden.

Es ist die Frage, was schlimmer ist, wenn mal eine Booksection als Article interpretiert wird und der Journaltitle dann den Autor beinhaltet, oder wenn ein Article als Booksection angelegt wird und versehentlich der erste Teil des Journaltitle als Autor interpretiert wird.

In Zukunft könnte man vielleicht auch einen Graubereich einfügen, z.B. bei 3 oder 4 Wörtern vor dem Komma wird eine Section einer Generic Referenz verwendet, deren Titel / Autor dann nicht parsebar ist und die mit einer Warnung versehen wird.

Noch besser wäre es natürlich, irgendwann Lookups einzuführen, bei denen in kritischen Fällen dann aus einer hinterlegten Liste (hardcodiert, aus Datenbank oder aus einer webzugängigen Liste mögliche Journal Titel bzw. Autoren abgefragt werden.

=====

KL:

Habe jetzt nochmal debugged und es ist tatsächlich so, dass der Parser „Acta Univ.Carol.“ als Autorenstring erkennt.

=====

KL:

Also bei dem gleichen Event wurde der Name „Bufonia kotschyana subsp. calycina (Boiss. & Hausskn.) Chrtek & Křísa in Acta Univ.Carol., Biol. 43(2): 106. 1999“ erzeugt, aber auch das book und booksection und die vier Autoren.

Bei „Bufonia kotschyana subsp. densa Chrtek & Křísa in Acta Univ.Carol., Biol. 43(2): 105. 1999“ ist dann das gleiche nochmal passiert und bei

„Bufonia calderae Chrtek & Křísa in Acta Univ.Carol., Biol. 38(2): 89. 1995“,

„Bufonia paniculata subsp. orientalis Chrtek & Křísa in Acta Univ.Carol., Biol. 43(2): 87. 1999“,

„Bufonia pabotii Chrtek & Křísa in Acta Univ.Carol., Biol. 43(2): 99. 1999“,

„Bufonia anatolica var. elatior Chrtek & Křísa in Acta Univ.Carol., Biol. 43(2): 95. 1999“,

„Bufonia ephedrina subsp. musilii Chrtek & Křísa“,

„Bufonia ephedrina var. minor Chrtek & Křísa in Acta Univ.Carol., Biol. 43(2): 98. 1999“,

„Bufonia anatolica Chrtek & Křísa in Acta Univ.Carol., Biol. 43(2): 94. 1999“

Associated revisions

Revision c70266b8 - 03/07/2018 10:00 AM - Andreas Müller

ref #7304 reduce author part in booksections to 2 words in name parser