BIEN 3 - Task #636

support TNRS-matching names in "", insted of encoding them and letting TNRS return no match

06/14/2013 09:26 AM - Aaron Marcuse-Kubitza

Status:	New		**Start date:**	06/14/2013
Priority:	Normal		**Due date:**	
Assignee:	Aaron Marcuse-Kubitza		**% Done:**	0%
Category:			**Estimated time:**	0.00 hour
Target version:				
Activity type:				

Description

- names like "Acer rubrum" (*with* quotes) currently get encoded as !quo Acer rubrum !quo , which TNRS can't parse (but will at least pass through verbatim)
- unfortunately, the encoding/decoding is required in order to make TNRS's Name_submitted match the actual submitted name
  - otherwise, some input names would not be entered verbatim in the TNRS cache, causing them to be repeatedly re-scrubbed

- fixing this requires matching TNRS output names to input names by Name_number instead of Name_submitted, to avoid relying on Name_submitted
  - in this case, Name_submitted would instead contain a TNRS-preprocessed version of the input name, rather than the exact encoded input name
  - some characters would no longer need to be encoded, because it will be OK for TNRS to strip them out
  - however, some characters break TNRS completely, so those will still need to be encoded or stripped out in advance

- see lib/tnrs.py > encode_map

History

#1 - 06/14/2013 09:27 AM - Aaron Marcuse-Kubitza

- Description updated

#2 - 06/14/2013 09:32 AM - Aaron Marcuse-Kubitza

- Description updated

#3 - 04/28/2017 05:13 AM - Aaron Marcuse-Kubitza

- Description updated