WP2011-25 NRDC 2011

Spelling of nuclides and mathematical expression in free text (A54)

M. Bossant, NEA Data Bank

The in-house code (see WP2011-24) developed to compare EXFOR references with information provided on the publisher websites shows differences in titles due to spelling of nuclides and mathematical expressions. Some examples of discrepant titles are given in the following table.

Entry		Titles
22406.001	Web title	Excitation functions of (n,2n), (n,p), (n,np+pn+d), and (n, α) reactions on
		isotopes of chromium
	EXFOR TITLE	Excitation Functions of (N,2N), (N,P), (N,NP+PN+D), and (N,ALPHA)
		Reactions on Isotopes of Chromium
T0172.001	Web title	92Zr(p→, a)89Y reaction at 17 MeV
	EXFOR TITLE	92Zr(p(pol),t)89Y reaction at 17 MeV
12540.001	Web title	Fission Cross Section of Plutonium-242
	EXFOR TITLE	FISSION CROSS SECTION OF PU242.
C0990.001	Web title	Excitation Function for Na22 from Deuterons on Aluminum
	EXFOR TITLE	Excitation function for 22Na from deuterons on aluminium
M0680.001	Web title	Simultaneous measurement of the photodisintegration of 4He in the giant
		dipole resonance region
	EXFOR TITLE	Simultaneous measurement of the photodisintegration of 4He in the giant
		dipole resonance region. The title "Photonuclear reactions of light nuclei
		studied with high-intensity real photon beams" is presented in
		C,2006CERN,176,2006
20426.001	Web title	Isospin and strong coupling effects in neutron scattering from even-A Se
		isotopes
	EXFOR TITLE	-ELASTIC AND INELASTIC SCATTERING OF FAST NEUTRONS ON EVEN
		ISOTOPES OF SE IN FRENCH

In this last example, the translation indication may be better indicated in the REFERENCE keyword free text rather than directly in the TITLE keyword. In the case of nuclides, various spelling rules are observed in EXFOR (e.g. Na22 vs. 22Na or Plutonium-242 vs. PU242). Some material names have different spellings between American and English e.g. Aluminium / Aluminum.

Experience from developing our in-house code for retrieving citations from the Internet (see WP2011-24) indicates that defining coding rules for nuclides and mathematical expressions in free text is not practical and cannot be exhaustive. Indeed, there are more than one hundred characters used in published journal titles that are not allowed in the EXFOR character set. Furthermore, the code already detects discrepant titles without using any transcoding rule.

Finally, information such as nuclide and polarization must be coded in the REACTION keyword and free text should not be used to store/retrieve coded information.

As a general rule, it is preferable to keep the title as close as given in the publication. For mathematical expressions, it is not possible to cover all cases without resorting to high-level languages such as LaTeX, HTML, etc. To conclude, the analysis of EXFOR titles shows that more important coding mistakes should be corrected first.

Example of title for which it would be very difficult to propose an usable coding rule: (ARI,27,675)

