Subject: feature suggestion "pdf2dwar"
Posted by nbehrnd on Thu, 06 Jun 2019 09:37:19 GMT
View Forum Message <> Reply to Message

Dear Thomas,

maybe the following feature might be implemented in future versions of the program.

Among the complementary .dwar files is the one about reactions associated with US Patents. Beside these, publications in scientific journals represent an additional source of information; frequently available as .pdf files. Given Lowe's thesis "Extraction of chemical structures and reactions from the literature" already mentioned, and picture-to-SMILES converters like OSRA on

https: // cactus.nci.nih.gov/cgi-bin/osra/index.cgi

perhaps DataWarrior may be enabled to harvest equally their information, too.

I speculate current publications already set to appear as .pdf might be easier to work with, than those scanned after their publication in print (e.g., Acta Chemica Scandinavica).

Well, it may sound like a resurrection of MDL's IsisBase seen in the later 1990s. To some extent, it is tangential to webreactions, too. The idea surfaced (again) while accessing my literature reference program, zotero. So far, however, zotero's indexing is limited to text-only information; the addition of a key reaction is constrained by pasting a figure from the publication as annotation then accessible in its browser or report by entry (cf. the two example files attached).

Harvesting this information might be eased if the relevant .pdf are all deposit into a dedicated partition, rather than multiple sub-folders of the webbrowser directory requiring an os.walk.

Norwid

File Attachments

1) example_report_zotero.html, downloaded 1120 times
2) Weiberth-2011.pdf, downloaded 1070 times