Review of: "Biofuels and nanocatalysts: A Data Mining study"

Rafal Gieleciak

Potential competing interests: The author(s) declared that no potential competing interests exist.

The authors of this manuscript presented an interesting approach to data mining of more than thousand papers on biofuels and nanocatalysts. According to their research, there is a good chance that research into microalgae for fuel production and exhaust emission systems based on cerium oxide nanoparticles may yield some benefits in the future use of renewable resources. However, I would like to share the following comments and suggestions, hoping that these may help the authors to improve the overall quality of the paper:

- Page 2: I agree that nanocatalysts may be relatively 'easily separable' for processes such as biodiesel productions, but it will be much more complicated in case of 'nasty' biocrude/biooil upgrading processes.
- Figure 1: Please use the right number format for the year on x-axis.
- Figure 2: Consider adding the shared percentage of documents (in % in brackets) after subject area
- Figure 3: How Scopus deal with multiple authors from different countries? Does it take into consideration only corresponding authors or all the authors' affiliation?
- Page 6: It is interesting that Saudi Arabia is so high in this ranking. Traditionally, this country is seen/perceive as petroleum oriented country. Do you have any explanation for this observation (the relatively high number of papers written by researchers from Saudi Arabia)?
- Page 9: '… hydrogenation (cluster … videlicet (cluster 7; Occ. 13).' - I am not sure if this text is necessary here. This information may be redundant especially when we have all these details in Figure 6.
- Figure 6: Please add to the manuscript the Abbreviation list section, or include (in the caption) the expansion on the acronyms appearing in the figure.
- (a sad comment here) The manuscript reads smoothly until page 13 where the results of LSBI is presented and discussed (Figure 7 and Table 1). I noticed that the concept of LSBI and classification treemaps was not introduced in a proper and easy to understand way. Therefore, many potential readers may be confused at this point. The proper introduction and presentation these results seems to be very important especially in case as authors said 'These Treemaps, as well as the results shown in Table 1, are the direct result of the software developed especially for this work ...'
- Figure 7: (again a sad comment) I do not think that comparison identical terms pairs is appropriate and meaningful here. For example 'htl vs hydrothermal liquifaction', or 'egt vs. exhaust gas temperature', or 'dimethylofuran vs. dmf'. What clear and practical conclusion can we draw from comparing the terms with their abbreviations? Please redo that part or provide good justification here.