

---

Subject: Re: Similarity analysis using "find similar compounds..." - slow analysis of libraries

Posted by [thomas](#) on Thu, 11 Feb 2021 09:55:58 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

Sorry for the delay. Since the MacOS and the Linux version use exactly the same jar file, they both use the same procedure. However, it slipped my attention that your version didn't use multiple threads for the similarity calculation unless you used the flexophore descriptor. With that version I made a comparison on my Linux desktop and my MacBookPro (87.000 against 105.000 compounds), which took about 20 minutes (FragFp) and >2 hours (SkelSpheres) on both computers. I had expected the Linux machine to be faster, because it has a desktop i7 compared to the some years older MacBookPro laptop i7. I don't have an explanation for the Mac being equally fast.

I now updated the code (and dev version) to use all threads when calculating similarities independent of the descriptor type, which now increased the speed by factor 4 on my hexacore Linux. The speed gain is not higher, because not everything is multithreaded and because of the overhead to launch threads for every molecule from the second file set. In general the needed processing time should grow more or less linearly with the number of compounds in file 1 and 2. 30 or even 170 hours seem very high to me. Which descriptor did you use? The molecule size does not play a role, because the calculation is running on descriptors only. If the second file doesn't contain the needed descriptors, they are calculated on the fly, which then adds to the time needed. I will make some more tests with larger files...

---