I have submitted five source codes. All are in python programming language.

1. Web Crawler: This program crawls the needful from the website. It can directly be executed.
2. PLSA\_Matrix: To run this program, there is a set of prerequisites should be followed.
3. The path must be specified before running the program line number 20 and 203 , the path of the input data set must be specified.
4. In line number 200, one output directory must be ready before running program. The name and the path must be same as in the program
5. PLSA\_Dictionary: To run this program, there is a set of prerequisites should be followed.
6. The path must be specified before running the program line number 16 and 226 , the path of the input data set must be specified.
7. In line number 186, one output directory must be ready before running program. The name and the path must be same as in the program
8. PLSA\_Dictionary\_Reversed: To run this program, there is a set of prerequisites should be followed.
9. The path must be specified before running the program line number 21 and 255 , the path of the input data set must be specified.
10. In line number 237, one output directory must be ready before running program. The name and the path must be same as in the program

Due to shortage of time I have not used the converging point in reversed program. The converging logic I have implemented in the PLSA\_Dictionary program

Though, in the reversed program I have made use of a python’s in-built tool to show the output graphically.

1. Naming: This program will work on the result of the any of the program which is doing clustering

Apart from this five programs, I have submitted a zipped folder named full\_data, containing the dataset I have crawled from the dblp website.

Also I have submitted one sample output in a zipped folder named output.