# CS214-PA2

Gang Qiao

- **Step1: find all files in a directory recursively and tokenize each file  (opendir and readdir)**
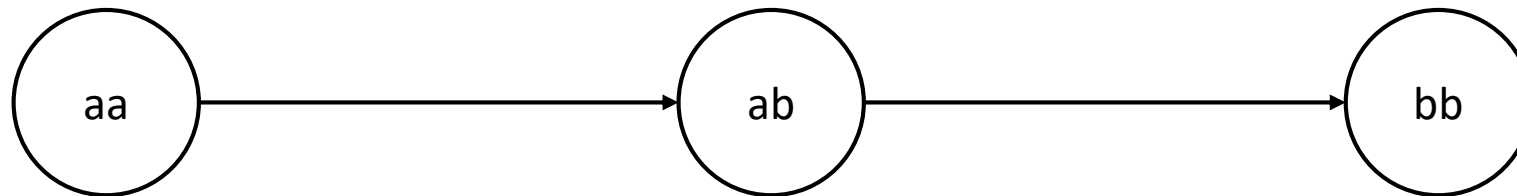
Will know all file names (filename1, filename2, …, filenameN)

Will know all tokens (aa, ab, bb, bc…, mm)

Sequential parsing in the first opendir run gives all filenames and all tokens

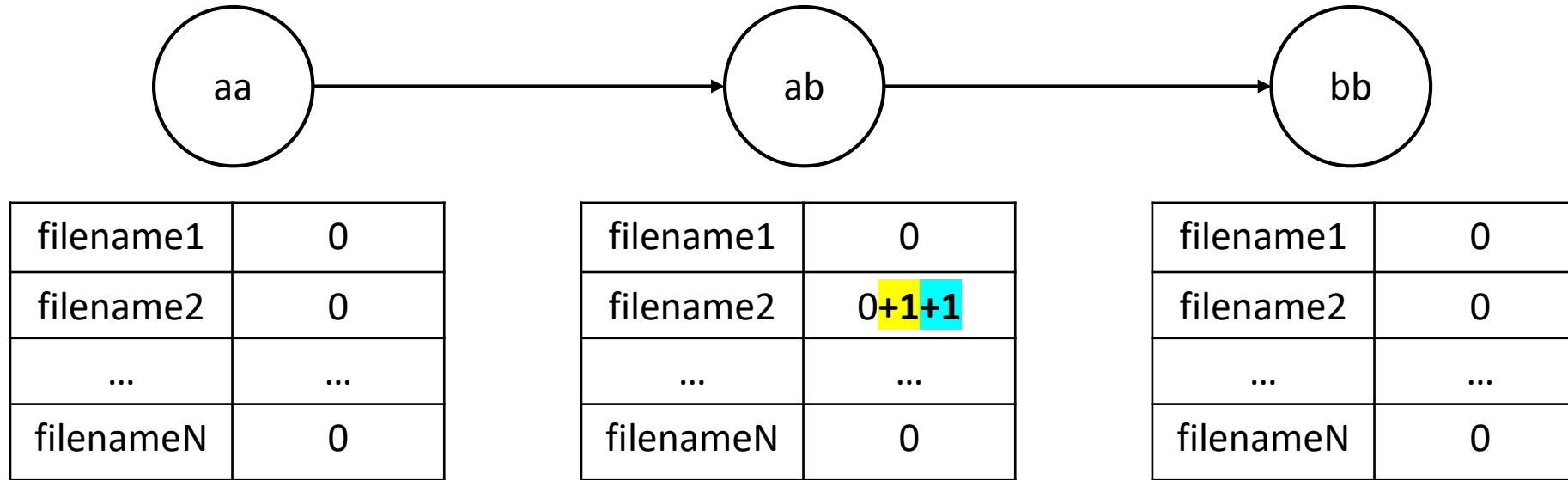- **Step2: construct a linked list, where each node is equipped with a 'complete' table of all filenames**



| aa | |
|---|---|
| filename1 | 0 |
| filename2 | 0 |
| … | … |
| filenameN | 0 |

| ab | |
|---|---|
| filename1 | 0 |
| filename2 | 0 |
| … | … |
| filenameN | 0 |

| bb | |
|---|---|
| filename1 | 0 |
| filename2 | 0 |
| … | … |
| filenameN | 0 |

- **Step3: find each file in the directory and tokenize it again**



| aa | | | ab | | | bb | |
|---|---|---|---|---|---|---|---|
| filename1 | 0 | | filename1 | 0 | | filename1 | 0 |
| filename2 | 0 | | filename2 | 0**+1+1** | | filename2 | 0 |
| … | … | | … | … | | … | … |
| filenameN | 0 | | filenameN | 0 | | filenameN | 0 |

e.g., for filename2, if token 'ab' appear the first time, increment +1 in row 'filename2' of the node 'ab'
for the same file, if token 'ab' appear again during parsing, increment +1 in row 'filename2' of the node 'ab'

**Sequential parsing in the second opendir run gives frequencies of all tokens**