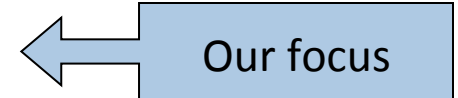


Introduction to **Information Retrieval**

Query processing with an inverted index

The index we just built

- How do we process a query?
 - Later - what kinds of queries can we process?



Query processing: AND

- Consider processing the query:

Brutus AND Caesar

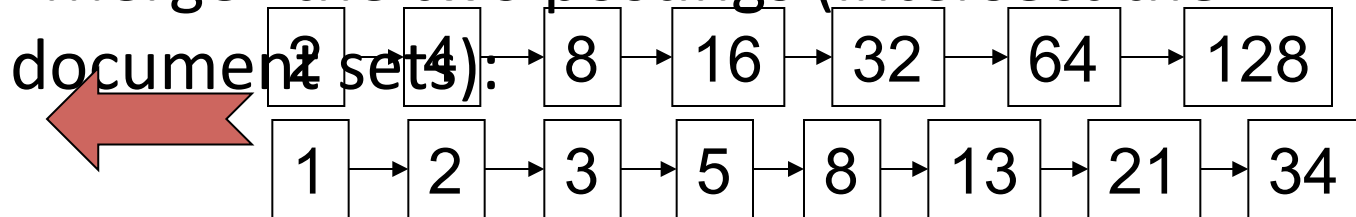
- Locate ***Brutus*** in the Dictionary;

- Retrieve its postings.

- Locate ***Caesar*** in the Dictionary;

- Retrieve its postings.

- “Merge” the two postings (intersect the

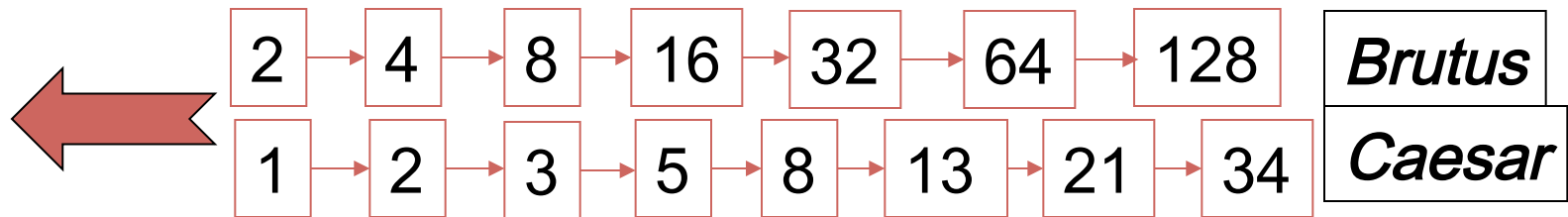


Brutus

Caesar

The merge

- Walk through the two postings simultaneously, in time linear in the total number of postings entries



If the list lengths are x and y , the merge takes $O(x+y)$ operations.

Crucial: postings sorted by docID.

Intersecting two postings lists (a “merge” algorithm)

```
INTERSECT( $p_1, p_2$ )  
  1   $answer \leftarrow \langle \rangle$   
  2  while  $p_1 \neq \text{NIL}$  and  $p_2 \neq \text{NIL}$   
  3  do if  $docID(p_1) = docID(p_2)$   
  4      then  $\text{ADD}(answer, docID(p_1))$   
  5           $p_1 \leftarrow next(p_1)$   
  6           $p_2 \leftarrow next(p_2)$   
  7      else if  $docID(p_1) < docID(p_2)$   
  8          then  $p_1 \leftarrow next(p_1)$   
  9          else  $p_2 \leftarrow next(p_2)$   
 10 return  $answer$ 
```

Introduction to **Information Retrieval**

Query processing with an inverted index