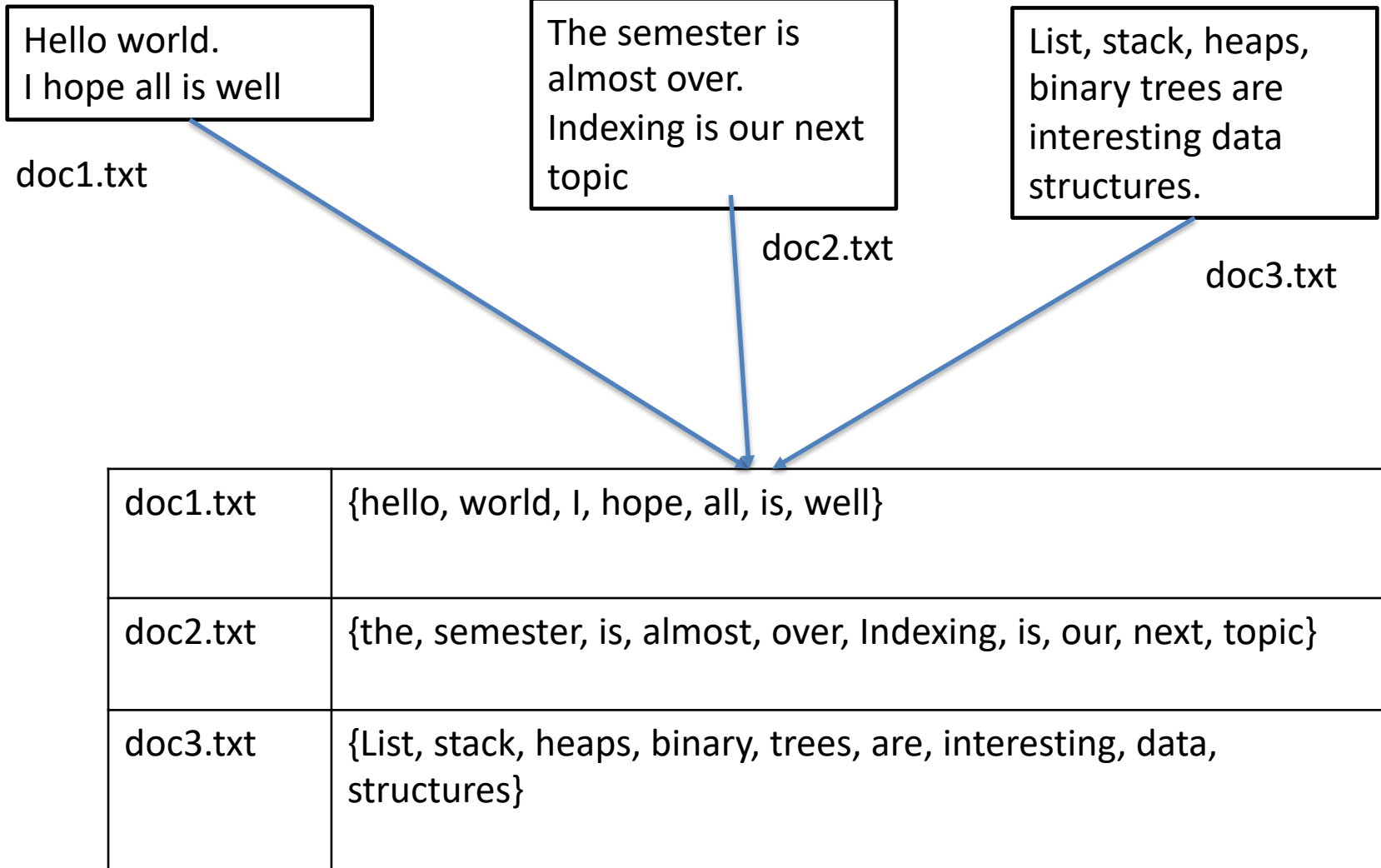


News Aggregator

CIT594

Task 2: Parsing documents



Task 3: Forward index

doc1.txt	{hello, world, I, hope, all, is, well}
doc2.txt	{the, semester, is, almost, over, Indexing, is, our, next, topic}
doc3.txt	{List, stack, heaps, binary, trees, are, interesting, data, structures}



**The forward
index**

doc1.txt	{<all,0.156>,<hello,0.156>,<hope,0.156>,<i,0.156>,<is,0.057>,<well,0.156>,<world,0.156>}
doc2.txt	{<almost,0.109>,<indexing,0.109>,<is,0.081>,<next,0.109>,<our,0.109>,<over,0.109>,<semester,0.109>,<the,0.109>,<topic,0.109>}
doc3.txt	{<are,0.122>,<binary,0.122>,<data,0.122>,<heaps,0.122>,<interesting,0.122>,<list,0.122>,<stack,0.122>,<structures,0.122>,<trees,0.122>}

Task 3: Inverted index

Forward index



next	doc2	0.10986122886681099>		
over	doc2	0.10986122886681099>		
all	doc1	0.15694461266687282>		
stack	doc3	0.12206803207423442>		The inverted index
indexing	doc2	0.10986122886681099>		
data	doc3	0.12206803207423442>		
almost	doc2	0.10986122886681099>		
interesting	doc3	0.12206803207423442>		
structures	doc3	0.12206803207423442>		
i	doc1	0.15694461266687282>		
is	doc2	0.08109302162163289>	doc1	0.05792358687259491>
hope	doc1	0.15694461266687282>		
list	doc3	0.12206803207423442>		
our	doc2	0.10986122886681099>		
trees	doc3	0.12206803207423442>		
heaps	doc3	0.12206803207423442>		
the	doc2	0.10986122886681099>		
world	doc1	0.15694461266687282>		
are	doc3	0.12206803207423442>		
binary	doc3	0.12206803207423442>		
topic	doc2	0.10986122886681099>		
well	doc1	0.15694461266687282>		
semester	doc2	0.10986122886681099>		
hello	doc1	0.15694461266687282>		

Task 5: Homepage

Inverted index



```
{<world,[doc1.txt]>,<trees,[doc3.txt]>,  
<topic,[doc2.txt]>,<structures,[doc3.txt]>,  
<stack,[doc2.txt]>,<semester,[doc2.txt]>,  
<list,[doc3.txt]>,<interesting,[doc3.txt]>,  
<indexing,[doc2.txt]>,<hope,[doc1.txt]>,  
<hello,[doc1.txt]>,<heaps,[doc3.txt]>,  
<data,[doc.txt3]>,<binary,[doc3.txt]>}
```

Stop words are removed

