

## INFORMATION RETRIEVAL – SHORT EXERCISES VI – INDEX CONSTRUCTION AND COMPRESSION

I. Consider the following fragment of a term-based positional index in the format:

**term:** doc1: <position1,position2,...>; doc2: <position1,...>; etc.

**Gates:** 1: <3>; 2: <6>; 3: <2,17>; 4: <1>;

**IBM:** 4: <3>; 7: <14>;

**Microsoft:** 1: <1>; 2: <1,21>; 3: <3>; 5: <16,22,51>;

The **/k** operator, **word1 /k word2** finds occurrences of word1 within k words of word2 (on either side), where **k** is a positive integer argument. Which document(s) satisfy the query “**Gates /2 Microsoft**”?

Answer:

II. Build a suffix array for “couscous\$” using the *qsufsort* algorithm.

	i	1	2	3	4	5	6	7	8	9
h	$x_i$	c	o	u	s	c	o	u	s	\$
	A[i]	9	1	5	2	6	4	8	3	7
	V[A[i]]	1	3	3	5	5	7	7	9	9
1	V[A[i]+h]		5	5	9	9	3	1	7	7
	A[i]	9	1	5	2	6	8	4	3	7
	V[A[i]]	1	3	3	5	5	6	7	9	9
2	V[A[i]+h]		9	9	7	6			3	1
	A[i]									
	V[A[i]]									
4	V[A[i]+h]									
	A[i]									

III. Encode 15 in  $\gamma$ . Answer:

IV. Decode 00111000001 written in the  $\delta$ -code. Answer: