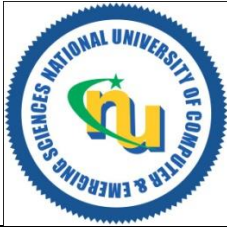


**National University of Computer and Emerging Sciences, Lahore Campus**

<b>Course:</b>	Information Retrieval and Text Mining	<b>Course Code:</b>	CS567
<b>Program:</b>	MS(Computer Science)	<b>Semester:</b>	Fall 2018
<b>Duration:</b>	20 Minutes	<b>Total Marks:</b>	10
<b>Paper Date:</b>	11-Oct-18	<b>Weight</b>	3.3%
<b>Section:</b>	CS	<b>Page(s):</b>	2
<b>Exam:</b>	Quiz 2	<b>Roll No:</b>	

**Q1)** Given the three-document corpus and a stop word list below, answer the following question AFTER removing stopwords.

<b>Document 1</b>	information retrieval is process of index search retrieval
<b>Document 2</b>	retrieval is used for evaluation of search results retrieval retrieval
<b>Document 3</b>	evaluation in information in evaluation process search
<b>Query</b>	information retrieval
<b>Stopwords</b>	is , of, in, for, to

Calculate cosine similarity between document 1 and query. Use tf.idf weight for query vector and normalized tf weight for document vector. [6 points]

**Q 2)** Suppose one document has 200 words and another document has 10,000 words. According to Laplace smoothing, which document's smoothed language model will give more weight to unseen query words ? Why ? [4 Points]