

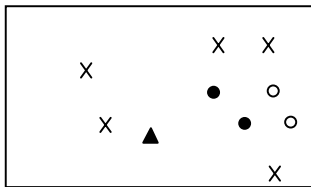
Name: _____ Student ID: _____

1. [30] Given the following three document spaces, fill in i), ii), or iii) in the space provided under each document space the method that might enable the system to return more relevant documents, or iv) if you think there is no way to improve the result. The four choices are:

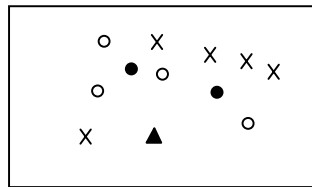
- i) reformulate original query into a new query based on relevant documents identified by user
- ii) document space modification
- iii) query splitting
- iv) no way to get more relevant document by relevance feedback

● documents identified as relevant by the user X non-relevant documents

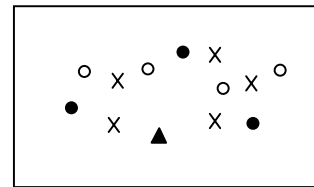
○ relevant documents not yet retrieved ▲ original query



method ____i____



method ____ii____



method iii(iv is acceptable)

2. [20] A corpus contains the following words of some unknown language:

doxp u v i n g
doxp u v e r
doxp u v e d
doxx o n
dox i n g
doxp l e x e n v

Give the successor variety of each prefix (d, do, dox, ..., etc) of the word "doxp u v e d" and then segment the word using peak and plateau method.

Prefix	Successor variety	Letters
d	1	o
do	1	x
dox	3	p, x, i
doxp	2	u, l
doxpu	1	v
doxpuv	2	e, i
doxpudev	2	r, d

Segment: dox | pudev

3. [30] After indexing a set of document, and you find that the documents are very close together according to the similarity measure you use. Explain the effect of having documents too close to each other on precision and recall.

Answer: precision and recall will both be dropped

What is the effect on the average distance between the documents in the collection if you:

- i) remove a common word from the index; give a brief explanation on your answer.

Answer: remove a common word will increase the average distance

- ii) remove a very rare word from the index; give a brief explanation on your answer.

Answer: remove a rarely word will increase the average distance very little;
we may regard the distance to be unchanged

4. [20] Circle the correct answer(s) below:

	improves	
stemming	<u>recall</u>	precision
thesaurus	<u>recall</u>	precision
term phrase formation	recall	<u>precision</u>
relevance feedback	<u>recall</u>	<u>precision</u>