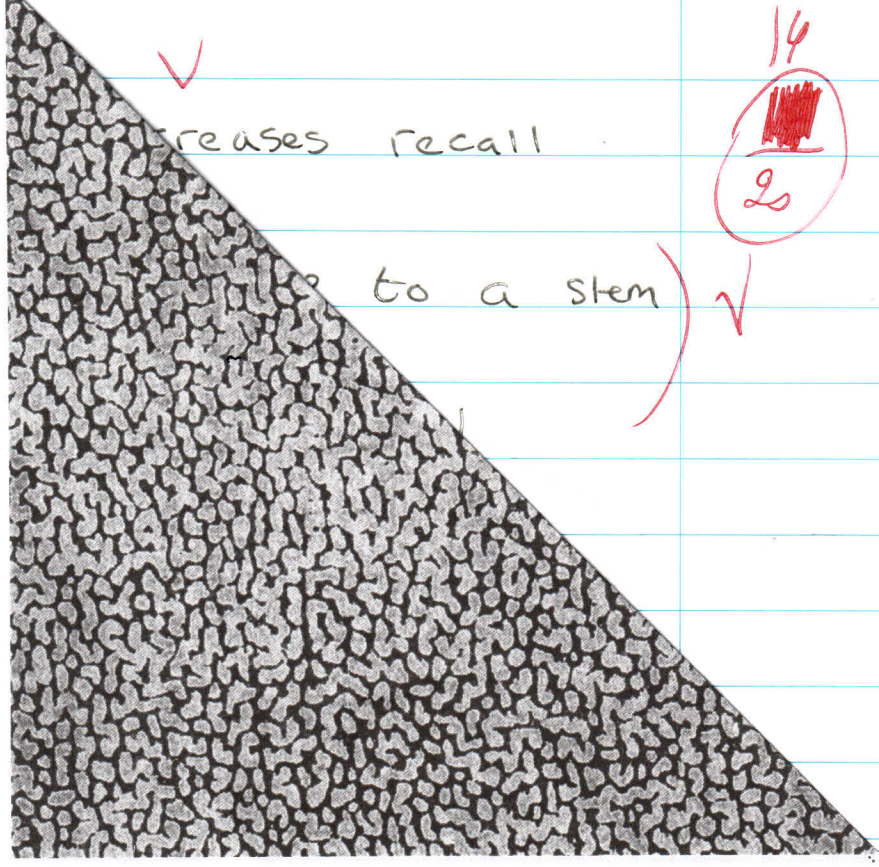
55 ~~57~~



Coláiste na hOllscoile
Corcaigh
University College Cork

Uimhir Scrúdaithe
Examination Number 9 1 7 1 6 3

Module Code CS4611

Paper No. _____

Mír
Section _____

Calculator, Please state:	No. of Books submitted
Name	2
Model	

Note: If there are different sections on this paper, a separate Answer Book MUST be used for each section.

Do na Scrúdaitheoirí amháin **For Examiner's use only**	
1	14/20
2	16/20
3	20/20
4	20/20
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
Iomlán **Total**	70/80

14
18
20

Q1 A Stemming typically increases recall and reduces percision. because more words map to a stem making more words equivalent Example WA WB WC → Stem W looking for either WA WB WC will get docs for WA WB WC will get docs for all

4/4

Q1B Databases typically use boolean retrieval and can occasionally support soundex Their data is structured into rows and columns and expert users typically create queries which involve the conjunction and disjunction of columns.

information retrieval data consists of documents with typically little structure (can have headings and titles)

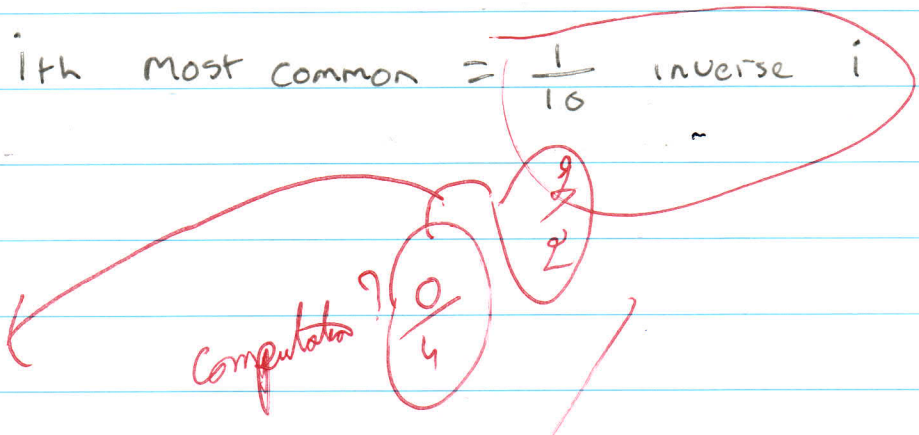
4/4

Queries are constructed so they satisfy an INFO need. They are typically consisting of words that the doc should contain. good IR systems rank data in order of how likely it is to be matching the information need

Q1 c

Zipfs Law $C_i = \frac{1}{i}(C_{i-1})$

ith most common = $\frac{1}{\text{inverse } i}$



Q1 d

rb * rb

doesn't exhibit 2

[rbArbA would match

even though rbarb[a] not rb * rb

14/6

Q2

16/20

A

	Retrieved		Not Retrieved
1	Relevant	1	Not relevant
2	Relevant	2	relevant
3	Not Relevant	3	Not relevant
4	Not relevant	4	Not relevant

$$\text{Precision} \quad \frac{\#\text{relevent Items Retrieved}}{\#\text{Retrieved Items}}$$

$$\text{Precision} = \frac{2}{4} = \frac{1}{2}$$

$$\text{Recall} = \frac{\#\text{relevant Items Retrieved}}{\#\text{Relevant items}}$$

$$\text{Recall} = \frac{2}{3}$$

4/4

Q2 B $$F = \frac{1}{\alpha\left(\frac{1}{P}\right) + (1-\alpha)\left(\frac{1}{R}\right)}$$

For Precision to be twice as important as recall $\alpha = \frac{2}{3}$ ✓

$$F = \frac{1}{\frac{2}{3}\left(\frac{1}{\left(\frac{1}{2}\right)}\right) + \frac{1}{3}\left(\frac{1}{\left(\frac{2}{3}\right)}\right)} \Rightarrow \frac{1}{\frac{4}{3} + \frac{1}{2}} = \frac{6}{11}$$

4/4

Q 2

c $$k = \frac{P(A) - P(E)}{1 - P(E)}$$

	Judge b. yes	Judge b. no	Total
Judge a Yes	2	1	3
Judge a NO	0	1	1
Total	2	2	4

0/1 3/3

$$P(A) = \frac{2+1}{4} = \frac{3}{4}$$ 1/1

$$P(E) = \frac{2+2+3+1}{4+4+4+4} = \frac{1}{2}$$ 0/2

values P(rule), P(non-rule) 0/2

$$k = \frac{\frac{3}{4} - \frac{1}{2}}{1 - \frac{1}{2}} = \frac{1}{2}$$ 3/3

1/1 NO there is not

A good agreement requires k to be between $\frac{2}{3}$ and 1. $\frac{1}{2} < \frac{2}{3}$

Q3

A

20/20

Document 1 Precision is very very high

Document 2 high Precision is very very very important

Term Freaency	D1	D2
Precision	$1+\log_{10}(1)$	$1+\log_{10}(1)$
very	$1+\log_{10}(2)$	$1+\log_{10}(3)$
high	$1+\log_{10}(1)$	$1+\log_{10}(1)$
Important	0	$1+\log_{10}(1)$

8/10

10/10

Term Freaency	D1	D2
Precision	1	1
very	1.30	1.48
high	1	1
Important	0	1

d/d

$$\text{Cosine similarity} = \frac{\sum_{i=1}^{|V|} d1_i \, d2_i}{\sqrt{\sum_{i=1}^{|V|} d1_i^2} \cdot \sqrt{\sum_{i=1}^{|V|} d2_i^2}}$$

normalises len normalises len

= use this to Find similrity

subbing in TF For D_1 and D_2

b

3 grams

gram ⇒ gr, gra, ram, am

spam ⇒ sp, spa, pam, am

cram ⇒ cr, cra, ram, am

scram ⇒ sc, scr, cra, ram, am

am$ → cram, gram, scram, spam

cra → cram, scram

gra → gram

pam → spam

ram → cram gram scram

scr → scram

spa → SPAM

$cr → cram

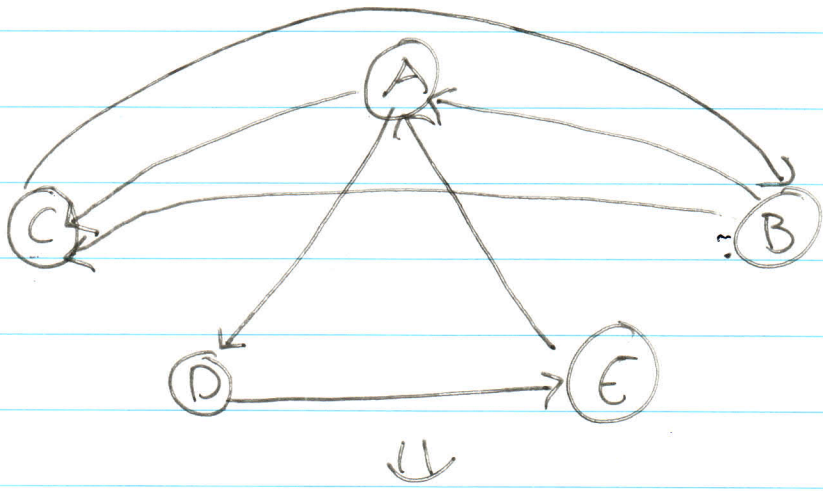
$gr → gram

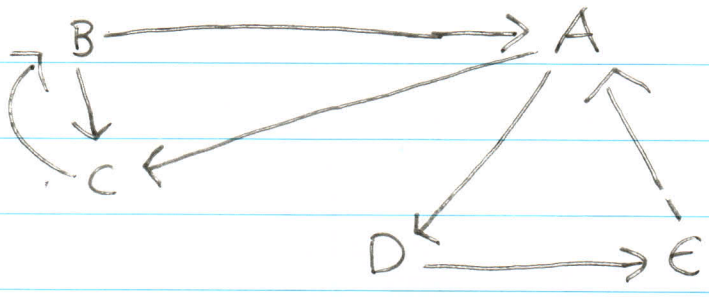
$sc → scram

$sp → spam

10/10

Question 4





Q4

A

Adjacency Matrix

	A	B	C	D	E
A	0	0	1	1	0
B	1	0	1	0	0
C	0	1	0	0	0
D	0	0	0	0	1
E	1	0	0	0	0

3/3

20/20

Adjacency matrix Probability no teleport

	A	B	C	D	E
A	0	0	0.5	0.5	0
B	0.5	0	0.5	0	0
C	0	1	0	0	0
D	0	0	0	0	1
E	1	0	0	0	0

2/2

Adjacency matrix Probality with teleporting

	A	B	C	D	E
A	0.1	0.1	0.25 + 0.1	0.25 + 0.1	0.1
B	0.25 + 0.1	0.1	0.25 + 0.1	0.1	0.1
C	0.1	0.5 + 0.1	0.1	0.1	0.1
D	0.1	0.1	0.1	0.1	0.5 + 0.1
E	0.5 + 0.1	0.1	0.1	0.1	0.1

Where Chance of teleport = 0.5

Q4 B

	A	B	C	D	E
A	0.1	0.1	0.35	0.35	0.1
B	0.35	0.1	0.35	0.1	0.1
C	0.1	0.6	0.1	0.1	0.1
D	0.1	0.1	0.1	0.1	0.6
E	0.6	0.1	0.1	0.1	0.1

Q4C (0 0 1 0 0)

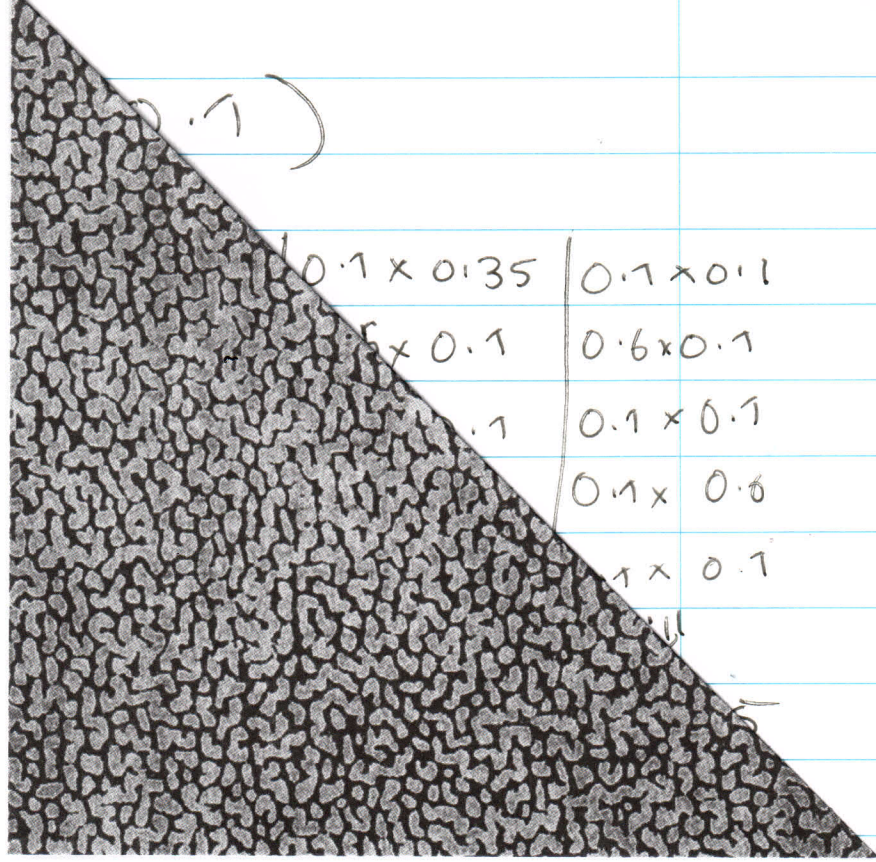
Q4D

From matrix multiplication

0 x 0.1 + 0 x 0.35 + 1 x 0.1 + 0 x 0.1 + 0 x 0.6	0 x 0.1 + 0 x 0.1 + 1 x 0.6 + 0 x 0.1 + 0 x 0.1	0 x 0.35 + 0 x 0.35 + 1 x 0.1 + 0 x 0.1 + 0 x 0.1	0 x 0.35 + 0 x 0.1 + 1 x 0.1 + 0 x 0.1 + 0 x 0.1	0 x 0.1 + 0 x 0.1 + 1 x 0.1 + 0 x 0.6 + 0 x 0.1

(0.1, 0.6, 0.1, 0.1, 0.1)

55 ~~54~~



Coláiste na hOllscoile
Corcaigh
University College Cork

0.1 x 0.35 | 0.1 x 0.1
x 0.1 | 0.6 x 0.1
.1 | 0.1 x 0.1
0.1 x 0.6
1 x 0.1

Uimhir Scrúdaithe
Examination Number 9 1 7 1 6 3

Module Code CS4611

Paper No. _____

Mír
Section _____

Do na Scrúdaitheóirí amháin **For Examiner's use only**	
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
Iomlán **Total**	

Calculator, Please state:	No. of Books submitted
Name ~~FX~~ Casio	2
Model Fx9lles	

Note: If there are different sections on this paper, a separate Answer Book MUST be used for each section.

Q4 D continued

(0.1 0.6 0.1 0.1 0.1)

0.1 × 0.1	0.1 × 0.1	0.1 × 0.35	0.1 × 0.35	0.1 × 0.1
0.6 × 0.35	0.6 × 0.1	0.6 × 0.35	0.6 × 0.1	0.6 × 0.1
0.1 × 0.1	0.1 × 0.6	0.1 × 0.1	0.1 × 0.1	0.1 × 0.1
0.1 × 0.1	0.1 × 0.1	0.1 × 0.1	0.1 × 0.1	0.1 × 0.6
0.1 × 0.6	0.1 × 0.1	0.1 × 0.1	0.1 × 0.1	0.1 × 0.1
↓	↓	↓	↓	↓
0.3	0.15	0.275	0.125	0.15
A	B	C	D	E

15/15