**Στέργιος Τσάνταλης (iis21125)**

1. α)Για το S1:

q1: P = 30/(30 + 20) = 30/50 = 3/5

R = 30/(30 + 20) = 30/50 = 3/5

F = 2\*(1/(1/(3/5))+(1/(3/5))) = 3/5 = 0.6

q2: P = 3/(3+47) = 3/50

R = 3/(3+17) = 3/20

F = 2\*(1/(1/(3/50))+(1/(3/20)))= 3/35 = 0.08

q3: P = 50/(50+0) = 1

R = 50/(50+150) = 50/200 = 1/4

F = 2\*(1/(1/1)+(1/(1/4))) = 2/5 = 0.4

|  |  |  |
| --- | --- | --- |
| Ερωτήματα | F1-Score | Macro-Averaged Values |
| q1 | 0.6 | (0.6+0.08+0.4)/3  = 0.36 |
| q2 | 0.08 |
| q3 | 0.4 |

Για το S2:

q1: P = 50/(50+0) = 1

R = 50/(50+0) = 1

F = 2\*(1/(1/1)+(1/1)) = 1

q2: P = 20/(20+30) = 20/50 = 2/5

R = 20/(20+0) = 1

F = 2\*(1/(1/(2/5))+(1/1)) = 4/7 = 0.57

q3: P = 150/(150+50) = 150/200 = 15/20

R = 150/(150+50) = 150/200 = 15/20

F = 2\*(1/(1/(15/20))+(1/(15/20))) = 3/4 = 0.75

|  |  |  |
| --- | --- | --- |
| Ερωτήματα | F1-Score | Macro-Averaged Values |
| q1 | 1 | (1+0.57+0.75)/3  = 0.77 |
| q2 | 0.57 |
| q3 | 0.75 |

Παρατηρούμε ότι το σύστημα S2 είναι καλύτερο και επομένως θα το προτιμούσαμε, καθώς έχει μεγαλύτερες τιμές στα precision, recall και F-measure με macro-averaging σε σχέση με το σύστημα S1.

β) Για το S1:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Ερωτήματα | TP | FP | FN | Micro-Averaged Values |
| q1 | 30 | 20 | 20 | P=83/(83+217)=0.27  R=83/(83+187)=0.3  F1Score=83/(83+(1/2\*(217+187)))=0.29 |
| q2 | 3 | 47 | 17 |
| q3 | 50 | 150 | 150 |
| Συνολικά | 83 | 217 | 187 |

Για το S2:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Ερωτήματα | TP | FP | FN | Micro-Averaged Values |
| q1 | 50 | 0 | 0 | P=220/(220+80)=0.73  R=220/(220+50)=0.81  F1Score=220/(220+(1/2\*(80+50)))=0.77 |
| q2 | 20 | 30 | 0 |
| q3 | 150 | 50 | 50 |
| Συνολικά | 220 | 80 | 50 |

Παρατηρούμε ότι το σύστημα S2 είναι καλύτερο και επομένως θα το προτιμούσαμε, καθώς έχει μεγαλύτερες τιμές στα precision, recall και F-measure με micro-averaging σε σχέση με το σύστημα S1.

1. (α) Για το S1: P = 5/(5+8) = 5/13

R = 5/(5+0) = 1

F1 = 2\*(1/(1/(5/13))+(1/1)) = 5/9 = 0.55

Για το S2: P = 5/(5+7) = 5/12

R = 5/(5+0) = 1

F1 = 2\*(1/(1/(5/12))+(1/1)) = 10/17 = 0.58

Για το S3: P = 4/(4+3) = 4/7

R = 4/(4+1) = 4/5

F1 = 2\*(1/(1/(4/7))+(1/(4/5))) = 2/3 = 0.66

Παρατηρούμε ότι το σύστημα S3 έχει μεγαλύτερες τιμές στα precision και F-measure αλλά μικρότερη τιμή στο recall σε σχέση με τα άλλα δύο συστήματα. Επομένως θα προτιμούσαμε το σύστημα S3.

(β) Για το S1: Precision@5 = 3/5

Recall@5 = 3/5

Για το S2: Precision@5 = 1/5

Recall@5 = 1/5

Για το S3: Precision@5 = 4/5

Recall@5 = 4/5

Παρατηρούμε ότι το σύστημα S3 έχει τις μεγαλύτερες τιμές για τα μέτρα Precision@5 και Recall@5 ενώ το σύστημα S2 τις μικρότερες τιμές. Επομένως θα προτιμούσαμε το σύστημα S3.

(γ) Για το S1: R-Precision = 3/5

Για το S2: R-Precision = 1/5

Για το S3: R-Precision = 4/5

Παρατηρούμε ότι το σύστημα S3 έχει την μεγαλύτερη τιμή για το μέτρο R-Precision και ενώ το σύστημα S2 την μικρότερή τιμή. Επομένως θα προτιμούσαμε το σύστημα S3.

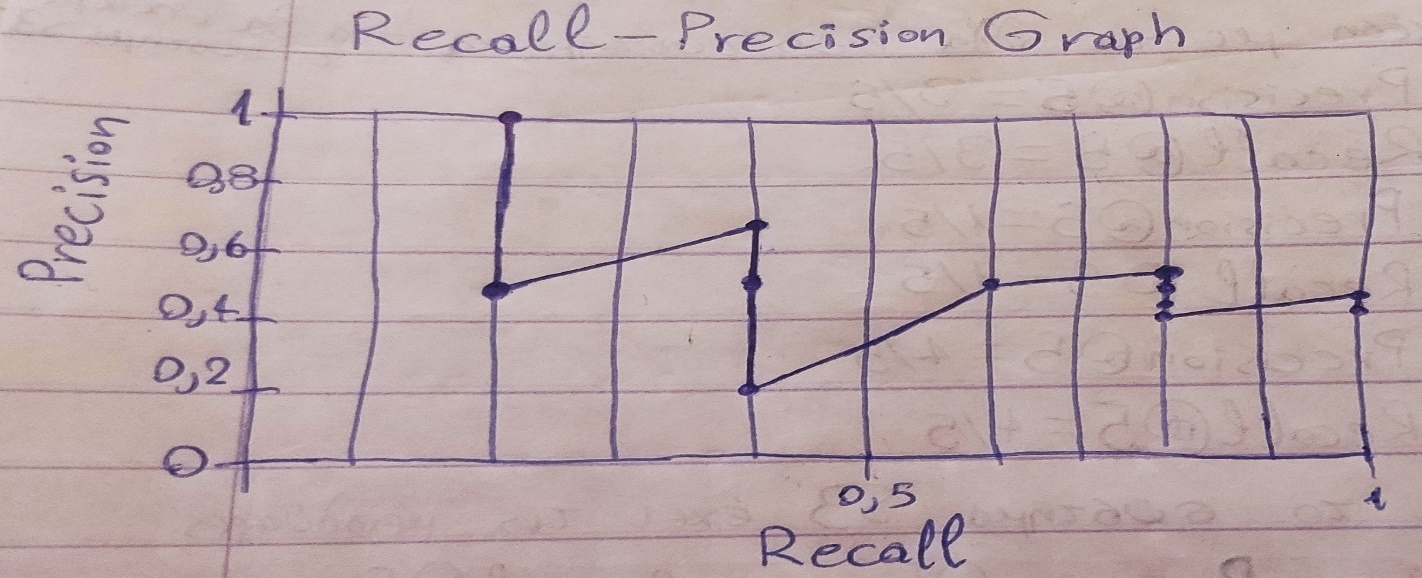
(δ) Για το S1: MAP = (1/3)\*((1/1)+(2/3)+(3/5)) = 34/45 = 0.75

Για το S2: MAP = (1/1)\*(1/5) = 1/5 = 0.2

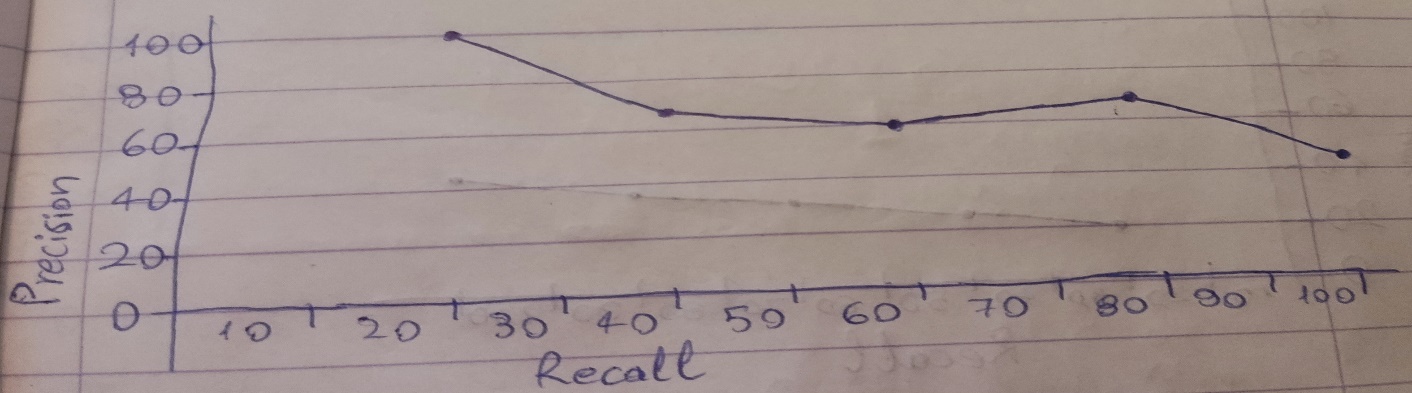
Για το S3: MAP = (1/4)\*((1/1)+(2/2)+(3/3)+(4/4)) = 1

Παρατηρούμε ότι το σύστημα S3 έχει την μεγαλύτερη τιμή για το μέτρο MAP και ενώ το σύστημα S2 την μικρότερή τιμή. Επομένως θα προτιμούσαμε το σύστημα S3.

(ε) Για το S1: R=0.2 P=1, R=0.4 P=0.66, R=0.6 P=0.5, R=0.8 P=0.57, R=1 P=0.45

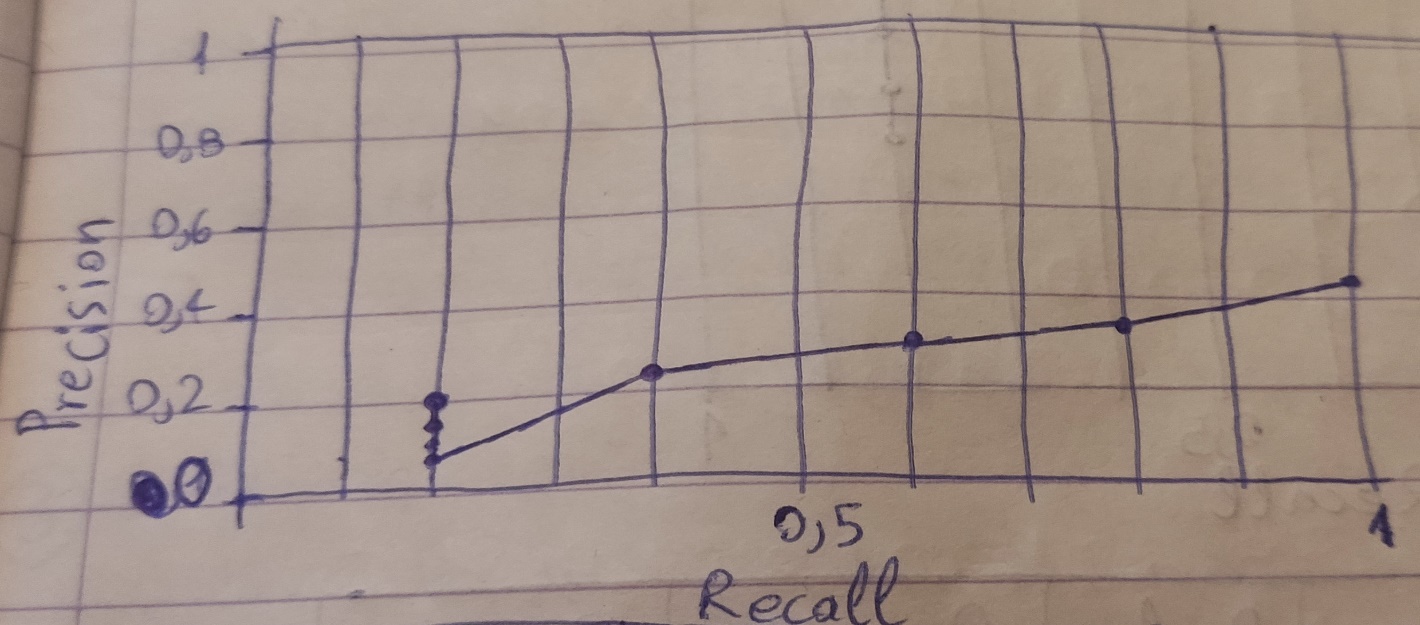


|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Rank | Doc | Rel | Recall | Precision |
| 0 |  |  | 0% | 0% |
| 1 | d11 | Yes | 20% | 100% |
| 2 | d4 |  | 20% | 50% |
| 3 | d18 | Yes | 40% | 67% |
| 4 | d30 |  | 40% | 50% |
| 5 | d21 | Yes | 60% | 60% |
| 6 | d33 | Yes | 80% | 67% |
| 7 | d9 |  | 80% | 57% |
| 8 | d7 |  | 80% | 50% |
| 9 | d8 |  | 80% | 44% |
| 10 | d6 |  | 80% | 40% |
| 11 | d1 | Yes | 100% | 45% |
| 12 | d5 |  | 100% | 42% |

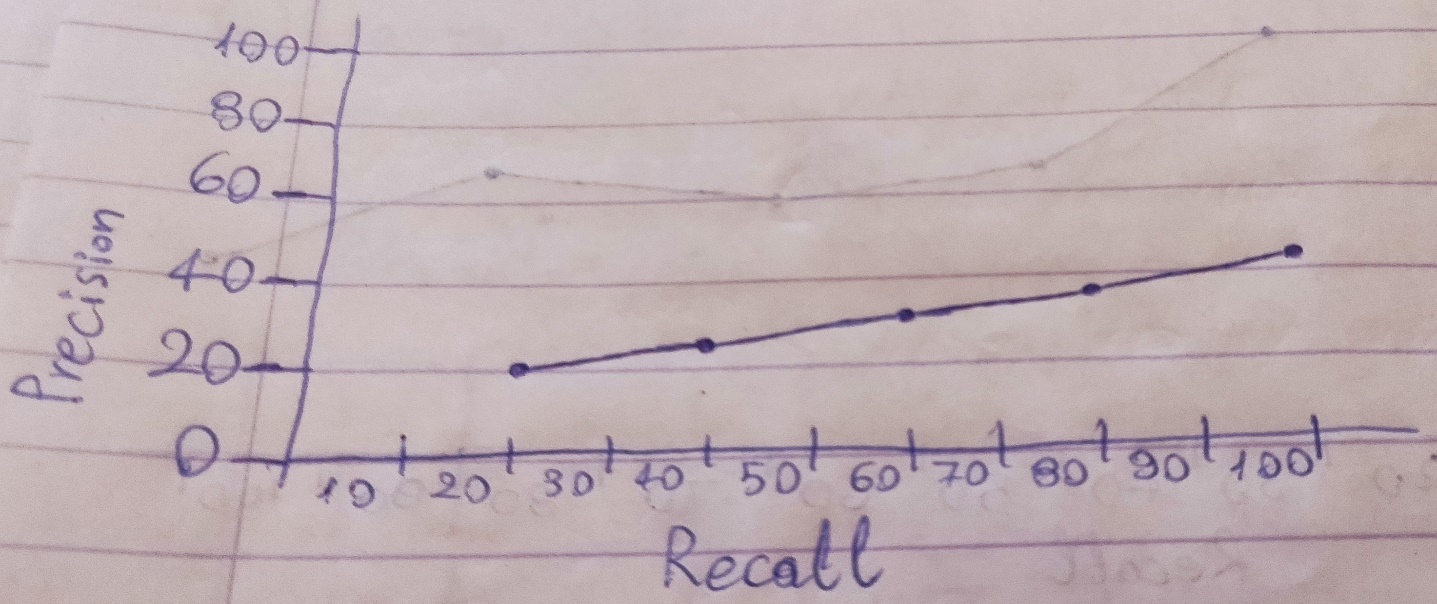


Για το S2: R=0.2 P=0.2, R=0.4 P=0.22, R=0.6 P=0.3, R=0.8 P=0.36, R=1 P=0.42

**Recall-Precision Graph**

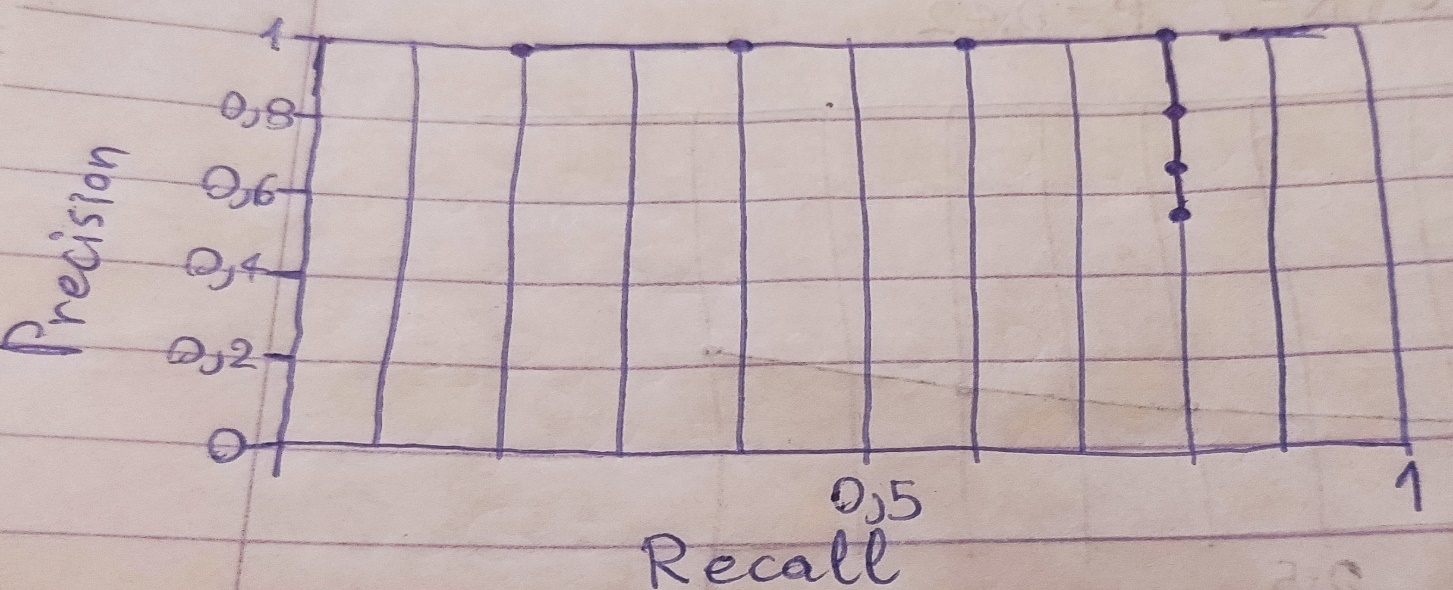
****

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Rank | Doc | Rel | Recall | Precision |
| 0 |  |  | 0% | 0% |
| 1 | d9 |  | 0% | 0% |
| 2 | d7 |  | 0% | 0% |
| 3 | d5 |  | 0% | 0% |
| 4 | d6 |  | 0% | 0% |
| 5 | d11 | Yes | 20% | 20% |
| 6 | d4 |  | 20% | 17% |
| 7 | d8 |  | 20% | 14% |
| 8 | d30 |  | 20% | 12% |
| 9 | d1 | Yes | 40% | 22% |
| 10 | d33 | Yes | 60% | 30% |
| 11 | d18 | Yes | 80% | 36% |
| 12 | d21 | Yes | 100% | 42% |

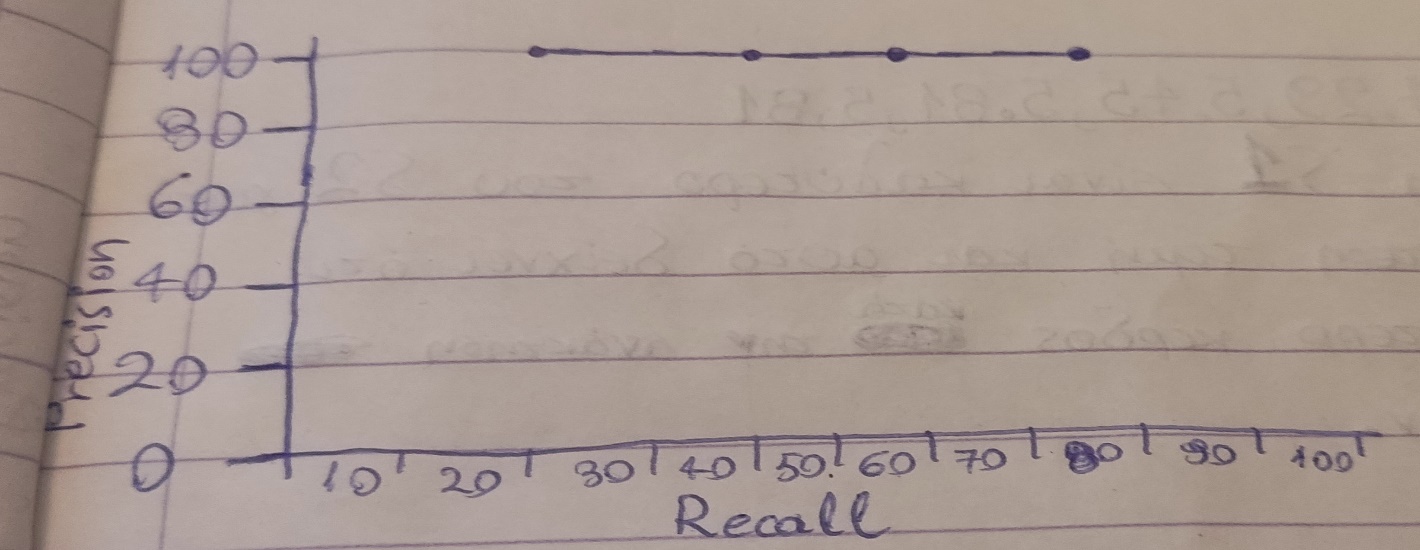
****

Για το S3: R=0.2 P=1, R=0.4 P=1, R=0.6 P=1, R=0.8 P=1

**Recall-Precision Graph**



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Rank | Doc | Rel | Recall | Precision |
| 0 |  |  | 0% | 0% |
| 1 | d18 | Yes | 20% | 100% |
| 2 | d33 | Yes | 40% | 100% |
| 3 | d11 | Yes | 60% | 100% |
| 4 | d1 | Yes | 80% | 100% |
| 5 | d12 |  | 80% | 80% |
| 6 | d2 |  | 80% | 67% |
| 7 | d30 |  | 80% | 57% |



Το σύστημα S3 παρόλο που έχει υψηλό Precision χάνει ένα συναφές έγγραφο το οποίο δεν ανακτά ποτέ. Το Precision του συστήματος S2 αυξάνεται παράλληλα με την αύξηση του Recall. Το Precision του συστήματος S1 αυξομειώνεται όσο αυξάνεται το Recall. Επομένως, θα επιλέγαμε το σύστημα S2 με το αυξανόμενο Precision παρόλο που το σύστημα S3 έχει άριστο Precision δεν ανακτά όμως ένα έγγραφο.

1. (α) Για το S1:

1,3,0,2,0,3,0,1

Discounted gain:

1,3/1,0,2/2,0,3/2.59,0,1/3

DSG:

1,4,4,5,5,6.16,6.16,6.49

Για το S2:

2,1,0,0,3,3,1,0

Discounted gain:

2,1/1,0,0,3/2.32,3/2.59,1/2.81,0

DSG:

2,3,3,3,4.29,5.45,5.81,5.81

Το σύστημα S1 είναι καλύτερο του S2 γιατί έχει μεγαλύτερη τιμή και αυτό δείχνει ότι έχει μεγαλύτερο κέρδος κατά την ανάκτηση των εγγράφων.

(β)Κάνουμε κανονικοποίηση των δύο συστημάτων:

Για το S1:

3,3,2,1,1,0,0,0

Discounted gain:

3,3/1,2/1.56,1/2,1/2.32,0,0,0

NDSG:

3,6,7.28,7.78,8.21,8.21,8.21,8.21

Για το S2:

3,3,2,1,1,0,0,0

Discounted gain:

3,3/1,2/1.56,1/2,1/2.32,0,0,0

NDSG:

3,6,7.28,7.78,8.21,8.21,8.21,8.21

Παρατηρούμε ότι κανονικοποιημένα τα δύο συστήματα έχουν την ίδια τιμή NDSG. Αυτό μας δείχνει ότι έχουν ίδιο κέρδος κατά την ανάκτηση των εγγράφων.