

Cumulative Gain

- With graded relevance judgments, we can compute the *gain* at each rank.
- **Cumulative Gain** at rank n :

$$CG_n = \sum_{i=1}^n rel_i$$

(Where rel_i is the graded relevance of the document at position i)

n	doc #	relevance	CG _n
		(gain)	
1	588	1.0	1.0
2	589	0.6	1.6
3	576	0.0	1.6
4	590	0.8	2.4
5	986	0.0	2.4
6	592	1.0	3.4
7	984	0.0	3.4
8	988	0.0	3.4
9	578	0.0	3.4
10	985	0.0	3.4
11	103	0.0	3.4
12	591	0.0	3.4
13	772	0.2	3.6
14	990	0.0	3.6

Discounting Based on Position

- Users care more about high-ranked documents, so we **discount** results by $1/\log_2(rank)$

- Discounted Cumulative Gain:**

$$DCG_n = rel_1 + \sum_{i=2}^n \frac{rel_i}{\log_2 i}$$

n	doc #	rel (gain)	CG _n	log _n	DCG _n
1	588	1.0	1.0	-	1.00
2	589	0.6	1.6	1.00	1.60
3	576	0.0	1.6	1.58	1.60
4	590	0.8	2.4	2.00	2.00
5	986	0.0	2.4	2.32	2.00
6	592	1.0	3.4	2.58	2.39
7	984	0.0	3.4	2.81	2.39
8	988	0.0	3.4	3.00	2.39
9	578	0.0	3.4	3.17	2.39
10	985	0.0	3.4	3.32	2.39
11	103	0.0	3.4	3.46	2.39
12	591	0.0	3.4	3.58	2.39
13	772	0.2	3.6	3.70	2.44
14	990	0.0	3.6	3.81	2.44

Normalized Discounted Cumulative Gain (NDCG)

- To compare DCGs, normalize values so that a *ideal ranking* would have a **Normalized DCG** of 1.0
- Normalized Cumulative Gain (NDCG) at rank n
 - Normalize DCG at rank n by the DCG value at rank n of the ideal ranking
 - The ideal ranking would first return the documents with the highest relevance level, then the next highest relevance level, etc
 - Compute the precision (at rank) where each (new) relevant document is retrieved => $p(1), \dots, p(k)$, if we have k rel. docs

Normalized Discounted Cumulative Gain (NDCG)

- To compare DCGs, normalize values so that a *ideal ranking* would have a **Normalized DCG** of 1.0
- Ideal ranking:

n	doc #	rel (gain)	CG _n	log _n	DCG _n
1	588	1.0	1.0	0.00	1.00
2	589	0.6	1.6	1.00	1.60
3	576	0.0	1.6	1.58	1.60
4	590	0.8	2.4	2.00	2.00
5	986	0.0	2.4	2.32	2.00
6	592	1.0	3.4	2.58	2.39
7	984	0.0	3.4	2.81	2.39
8	988	0.0	3.4	3.00	2.39
9	578	0.0	3.4	3.17	2.39
10	985	0.0	3.4	3.32	2.39
11	103	0.0	3.4	3.46	2.39
12	591	0.0	3.4	3.58	2.39
13	772	0.2	3.6	3.70	2.44
14	990	0.0	3.6	3.81	2.44



n	doc #	rel (gain)	CG _n	log _n	IDCG _n
1	588	1.0	1.0	0.00	1.00
2	592	1.0	2.0	1.00	2.00
3	590	0.8	2.8	1.58	2.50
4	589	0.6	3.4	2.00	2.80
5	772	0.2	3.6	2.32	2.89
6	576	0.0	3.6	2.58	2.89
7	986	0.0	3.6	2.81	2.89
8	984	0.0	3.6	3.00	2.89
9	988	0.0	3.6	3.17	2.89
10	578	0.0	3.6	3.32	2.89
11	985	0.0	3.6	3.46	2.89
12	103	0.0	3.6	3.58	2.89
13	591	0.0	3.6	3.70	2.89
14	990	0.0	3.6	3.81	2.89

Normalized Discounted Cumulative Gain (NDCG)

- Normalize by DCG of the ideal ranking:

$$NDCG_n = \frac{DCG_n}{IDCG_n}$$

- $NDCG \leq 1$ at all ranks
- NDCG is comparable across different queries
- NDCG is now quite popular in evaluating Web search

n	doc #	rel			
		(gain)	DCG_n	$IDCG_n$	$NDCG_n$
1	588	1.0	1.00	1.00	1.00
2	589	0.6	1.60	2.00	0.80
3	576	0.0	1.60	2.50	0.64
4	590	0.8	2.00	2.80	0.71
5	986	0.0	2.00	2.89	0.69
6	592	1.0	2.39	2.89	0.83
7	984	0.0	2.39	2.89	0.83
8	988	0.0	2.39	2.89	0.83
9	578	0.0	2.39	2.89	0.83
10	985	0.0	2.39	2.89	0.83
11	103	0.0	2.39	2.89	0.83
12	591	0.0	2.39	2.89	0.83
13	772	0.2	2.44	2.89	0.84
14	990	0.0	2.44	2.89	0.84