KRAMP
It's that easy.

Relevance metrics

April 28, 2023

Measuring relevance

Measuring relevance scores for search will allow us to measure the effect of changes we will make to improve search rankings and recall.

Our other key metrics for search (Search Success Rate and Time to Success) could be unaffected or even temporarily worsen while we do improve rankings.
This is especially relevant for Kramp, where we know we deal with learned behaviour in combination with repeated searches. An example of this is a user who searches for something quite generic like ‘M8’, scrolling down to position #20 and clicking that result. All in such a fast pace that it is clear this is learned behaviour.
Imagine we improve (the users) search results in such a way that the clicked result now is on position 1. It could very well be that Time to Success will be increased since the user won’t find the desired result at the expected position. However, we can reasonably expect in general that it is better to place the most relevant results high in the rankings.

The challenge is how to measure this properly.

Using DCG and NDCG

DCG is a metric that will represent the quality/relevance of the search results for a search query.

It will add up relevance scores for the results while giving more priority to the higher ranked results. Basically assuming that having the highest relevant result at the highest rank is what we want to achieve.

The main challenge of DCG is that it doesn't allow for comparisons between search queries, since the number of results will affect the range of possible DCG values.

A solution for this is to normalize the DCG values by calculating the ideal DCG (iDCG) (ie the maximum possible value for DCG for this search query) and dividing the DCG by the iDCG.
This results in NDCG values in a range of 0 to 1.

Using click data to calculate DCG and NDCG means that we should interpret this as the relevance of the search query *for that user at that moment in time*

Current definition of DCG

Currently we calculate the DCG by the following method:

On a search event level, we look at interactions within the search results.

- We give a relevance score of 2 to a ‘success action’ (NPC, ATC, ATP, ATF)
- We give a relevance score of 2 to a product_list_click that resulted in a success action
- We give a relevance score of 1 to a product_list_click that didn’t result in a success action
- We give a relevance score of 1 to a quick_view click
- We give a relevance score of 0 to all other results
- We only count 1 click per search result per search event.
    - We consider successful clicks before product list clicks
- We currently use all click positions.

We then calculate DCG in two ways:

DCG = ∑[i=1 to p](rel_i / log2(i+1)) $= \sum_{i-1}^{p} \frac{rel_i}{\log_2(i+1)}$

adjusted_DCG = ∑[i=1 to p]((2^rel_i - 1) / log2(i+1)) $= \sum_{i-1}^{p} \frac{2^{rel_i} - 1}{\log_2(i+1)}$

The 2nd option gives more emphasis on the relevance score. Since we only use a 3 point scale score for relevance at this moment, this will give very similar results as the first formula.
Therefor we’ll be using the first version for now.

Current definition of iDCG and NDCG

Currently we calculate the iDCG by the same method as calculating DCG. The only difference is that we don't use the actual clicked positions, but rather assume it would be ideal to have the clicked results ranked at the top. Therefore we do the following:

- Count the distinct clicked results for success clicks -> max_success_rank
- Count the distinct clicked results for product_list_clicks that aren't in success clicks -> max_productclick_rank
- Count the distinct clicked results for quickview clicks that arent in the previous 2 click arrays -> max_quickview_rank
- Give relevance value of 2 for results 1 up to max_success_rank
- Give relevance value of 1 for results max_success_rank up to max_success_rank + max_productclick_rank
- Give relevance value of 1 for results max_success_rank + max_productclick_rank up to max_quickview_rank

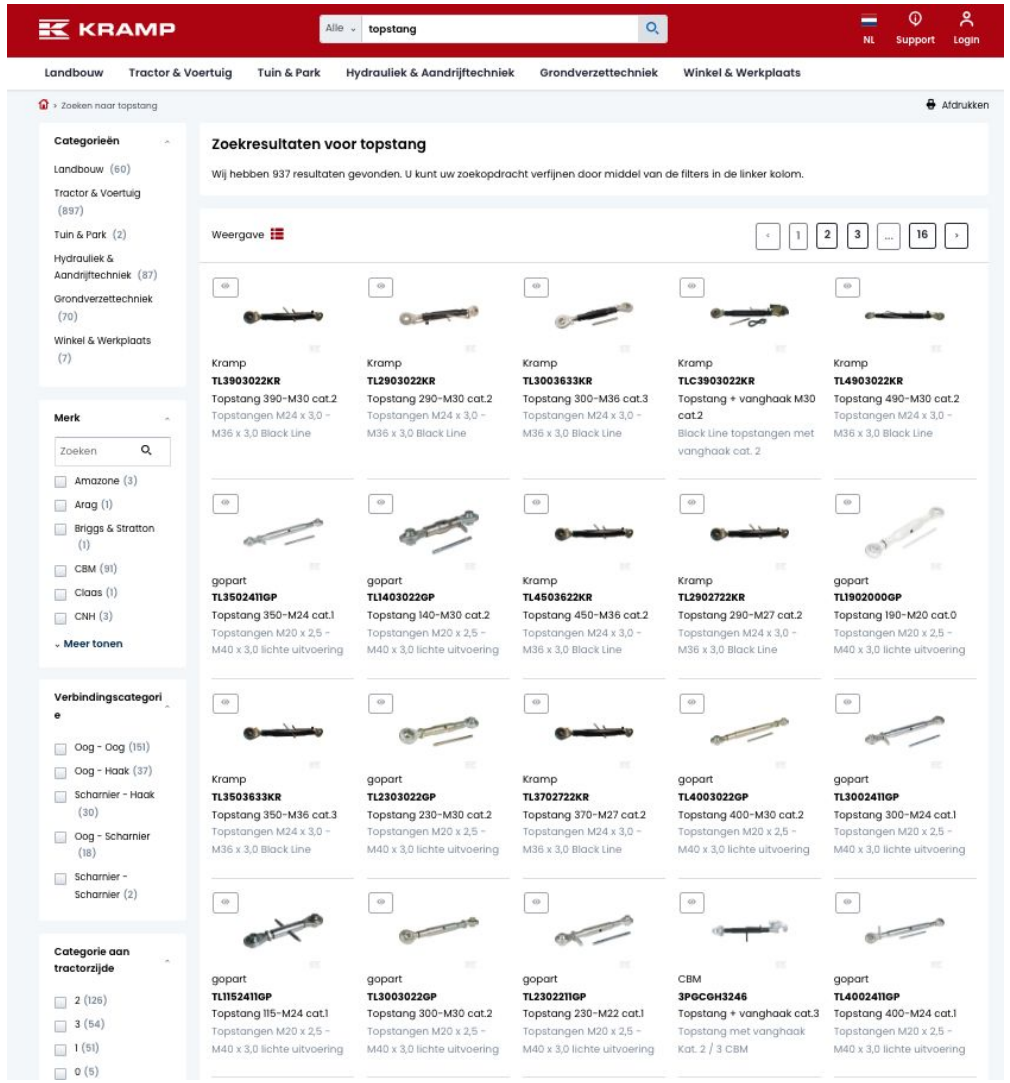
We then calculate iDCG in the same ways as DCG

iDCG = ∑[i=1 to p](rel_i / log2(i+1)) $= \sum_{i-1}^{p} \frac{rel_i}{\log_2(i+1)}$

adjusted_iDCG = ∑[i=1 to p]((2^rel_i - 1) / log2(i+1)) $= \sum_{i-1}^{p} \frac{2^{rel_i} - 1}{\log_2(i+1)}$

NDCG then simply is DCG / iDCG

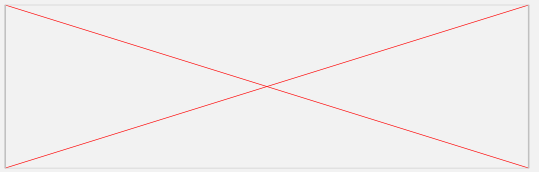
Qualitative versus click modeling



Qualitatively we would probably assign a NDCG score of 1 to this search result, since all results are ‘topstang’ products that exactly match to the search query.

However, customers will only be searching for one specific topstang and therefor interact with a view results, leading to vastly different NDCG scores.

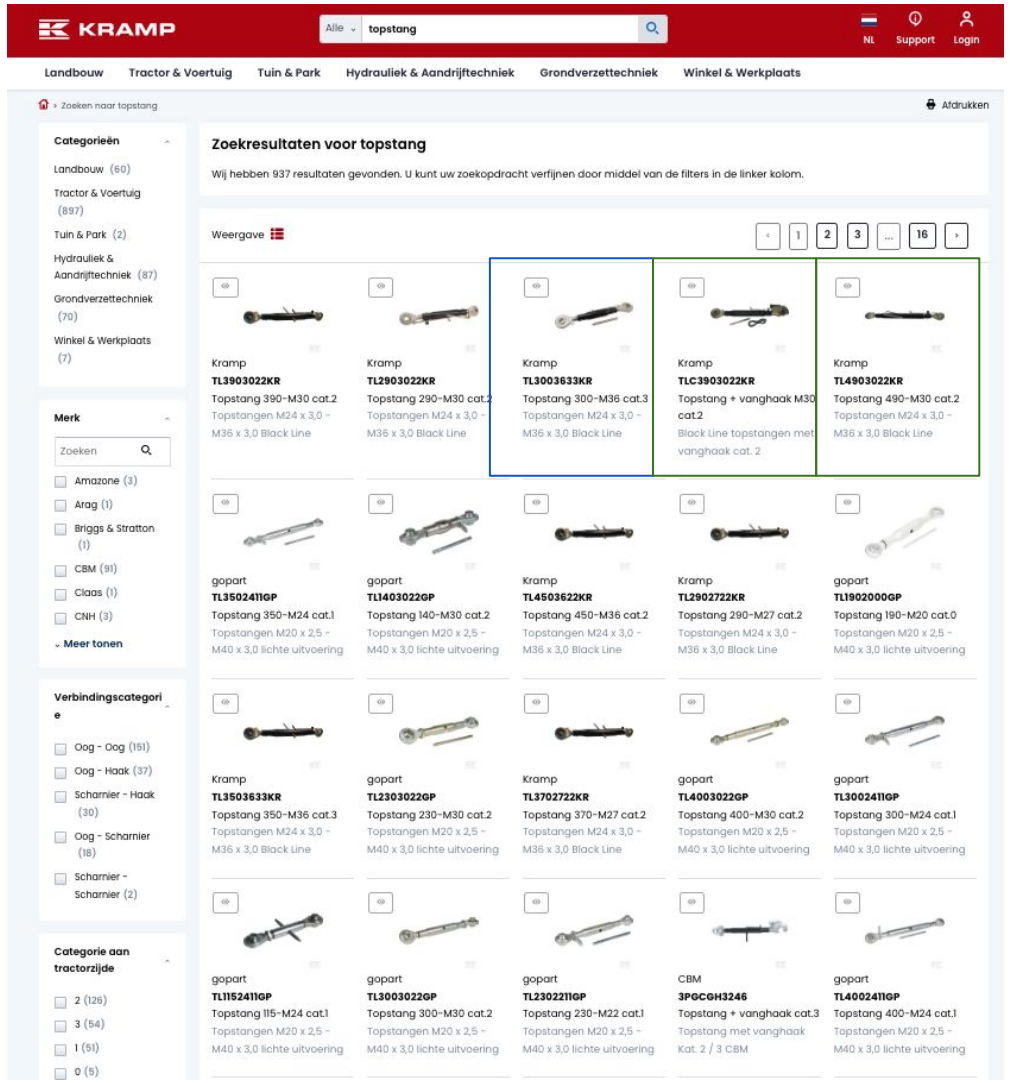
This makes our NDCG based on click modeling useful for determining how useful the result was for the user at a certain moment in time, rather than a general ‘judgement’ of the quality of search results. (basically assuming there is no such thing in general)



Example1

Current definition DCG and NDCG

Example1



User_1 searches for ‘topstang’,

Clicks towards the product detail page on #3,

Goes back and clicks ATP on #4 and #5

Does ATC for #5

We now calculate DCG for this search as:
DCG = 1/log2(1+3) + 2/log2(1+4) + 2/log2(1+5)

DCG = 2.14

Please not we assume all the other results not to be relevant in this calculation

For iDCG we basically assume the successful clicks should be at the top positions, followed by the product_list_clicks

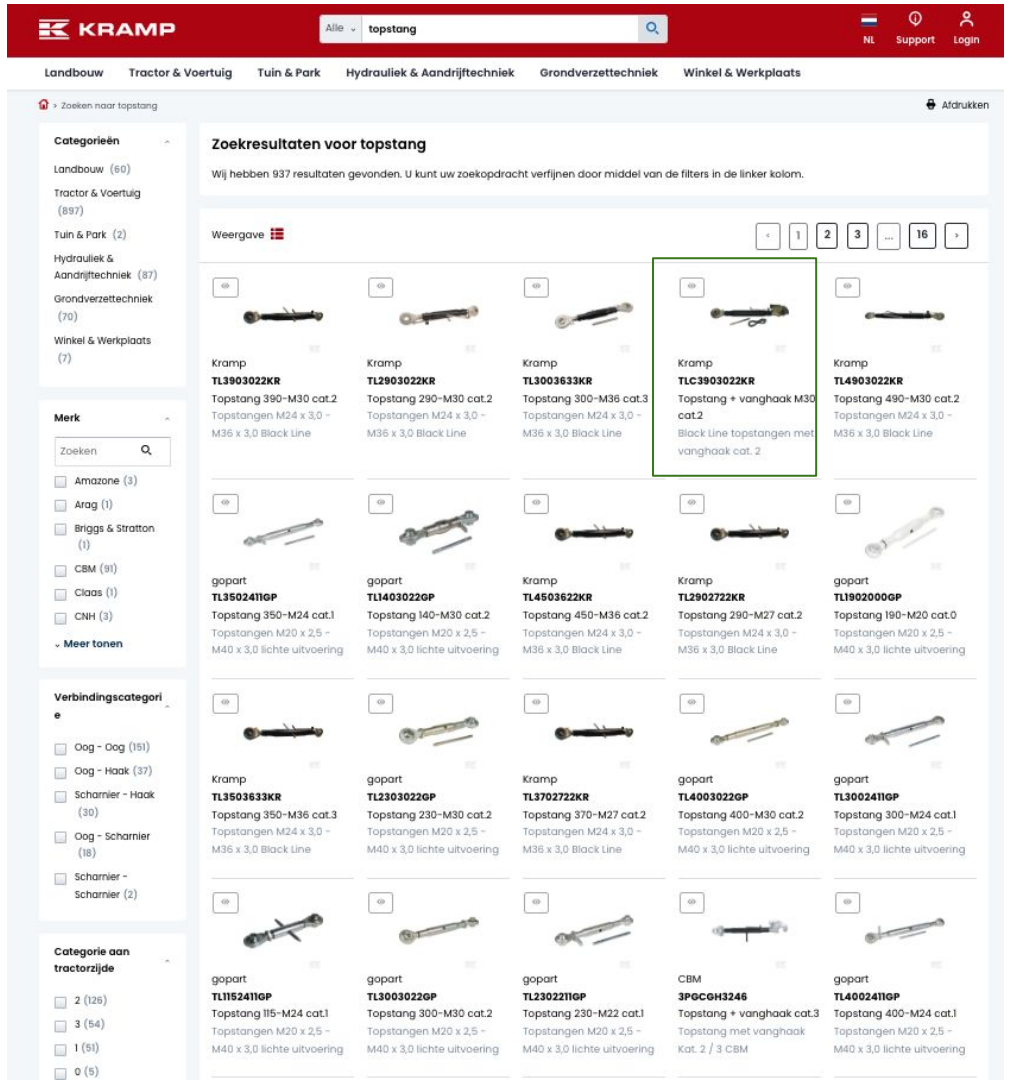
iDCG = 2/log2(1+2) + 2/log2(1+2) + 1/log2(1+3)

iDCG = 3.02

NDCG = DCG/iDCG
NDCG = 2.14 / 3.02 = 0.71

Current definition DCG and NDCG

Example2



User_2 searches for ‘topstang’,

Clicks ATC on #4

We now calculate DCG for this search as:
DCG = 1/log2(1+4)

DCG = 0.86

Please not we assume all the other results not to be relevant in this calculation

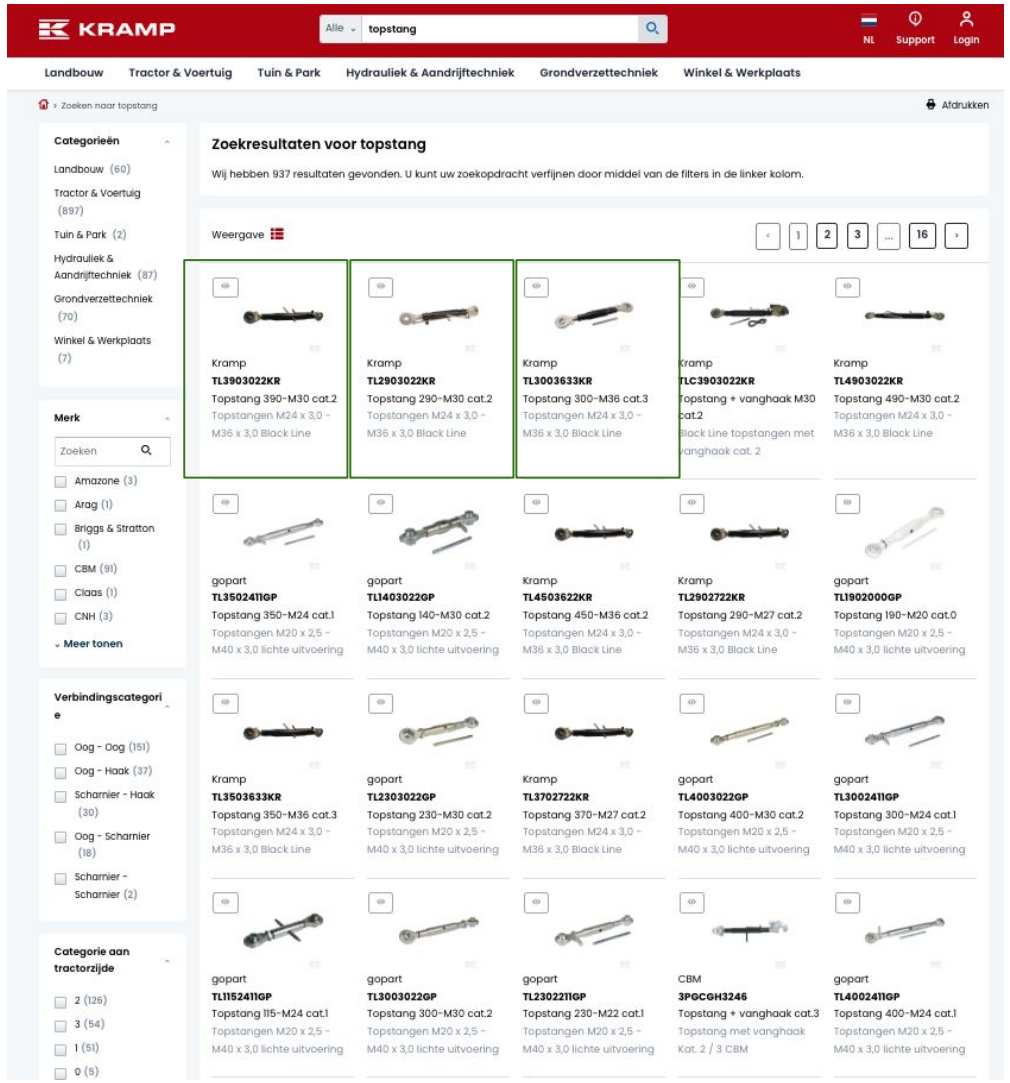
iDCG = 2/log2(1+1)

iDCG = 2

NDCG = DCG/iDCG
NDCG = 0.86/2 = 0.43

Current definition DCG and NDCG

Example3



User_3 searches for ‘topstang’,

Clicks NPC on #1 #2 and #3

We now calculate DCG for this search as:
DCG = 2/log2(1+1) + 2/log2(1+2) + 2/log2(1+3)

DCG = 4.26

Please not we assume all the other results not to be relevant in this calculation

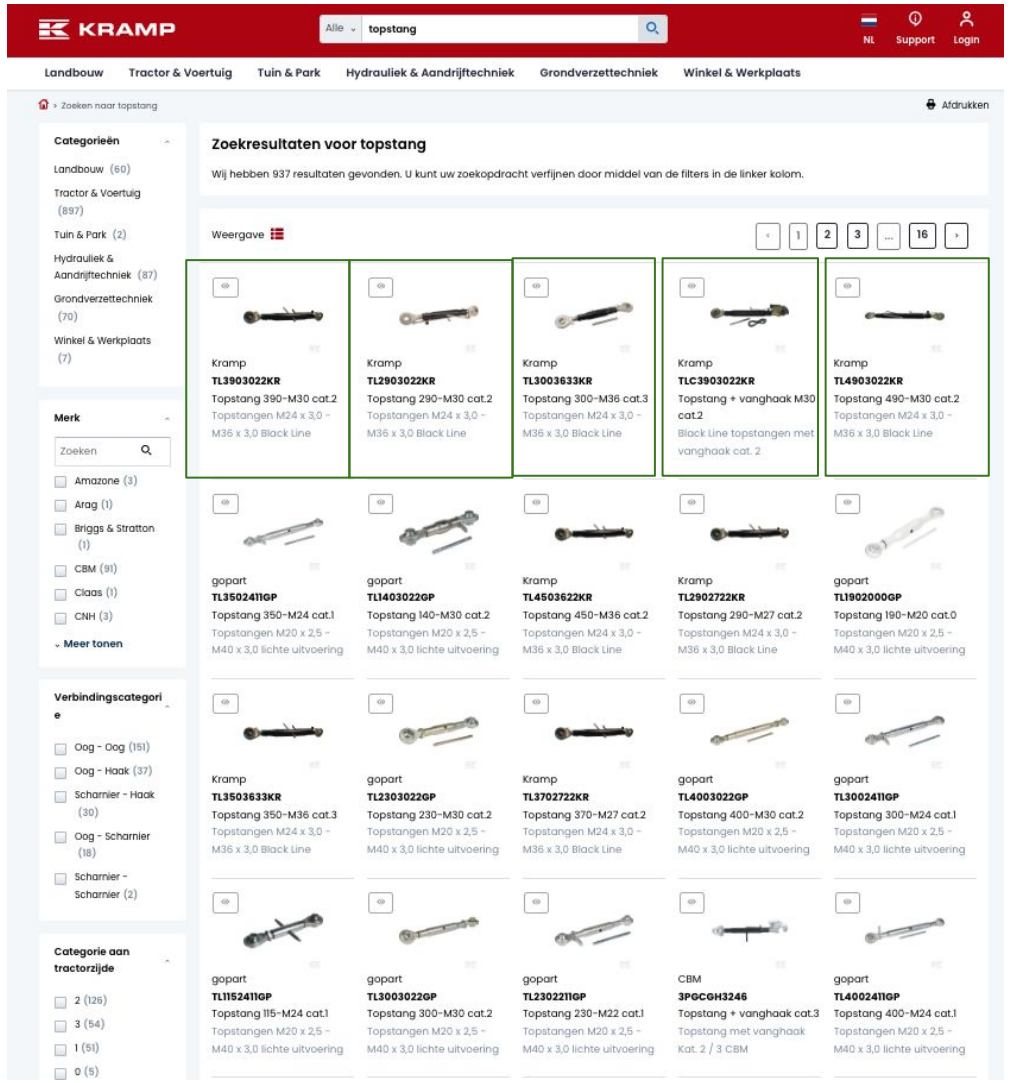
iDCG = 2/log2(1+1) + 2/log2(1+2) + 2/log2(1+3)

iDCG = 4.26

NDCG = DCG/iDCG
NDCG = 4.26/4.26= 1

Using search_query level NDCG

Combining all searches



If we combine the data of all 3 queries and add query level idcg we get:

Average DCG = (2.14+0.86+4.26) / 3

Average DCG = 2.42

Average NDCG = (0.71+0.43+1)/3

Average NDCG = 0.71

DCG, iDCG and NDCG on query level:

Success clicks: 4,5,4,1,2,3

Product list clicks: 3

Use frequency to determine relevance by counting every click position once per user.

DCG = 2/log2(1+4)+2/log2(1+5)+2/log2(1+4)+2/log2(1+1)+2/log2(1+2)+2/log2(1+3)

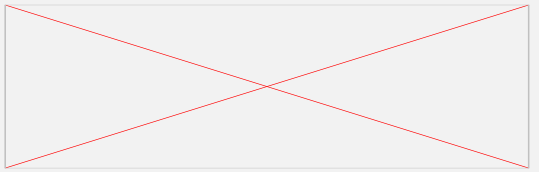
DCG = 6.76

iDCG = 2/log2(1+1) +2/log2(1+1) + 2/log2(1+2) + 2/log2(1+3) + 2/log2(1+4) + 2/log2(1+5)

iDCG = 5.9

NDCG = 6.76/7.9

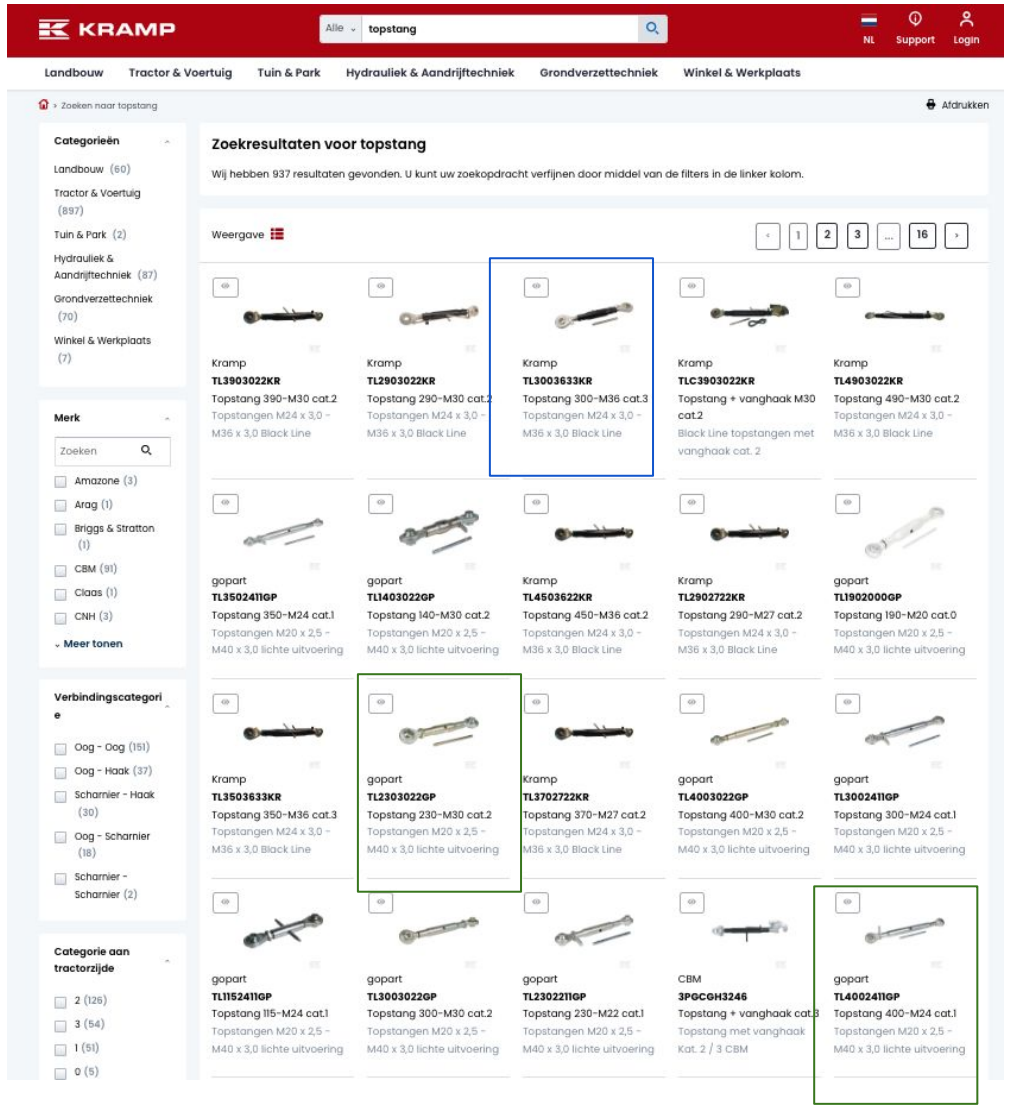
NDCG = 0.86



Example2

Current definition DCG and NDCG

Example1



User_1 searches for ‘topstang’,
Clicks towards the product detail page on #3,
Goes back and clicks NPC on #12 and #20
Does ATC for #12

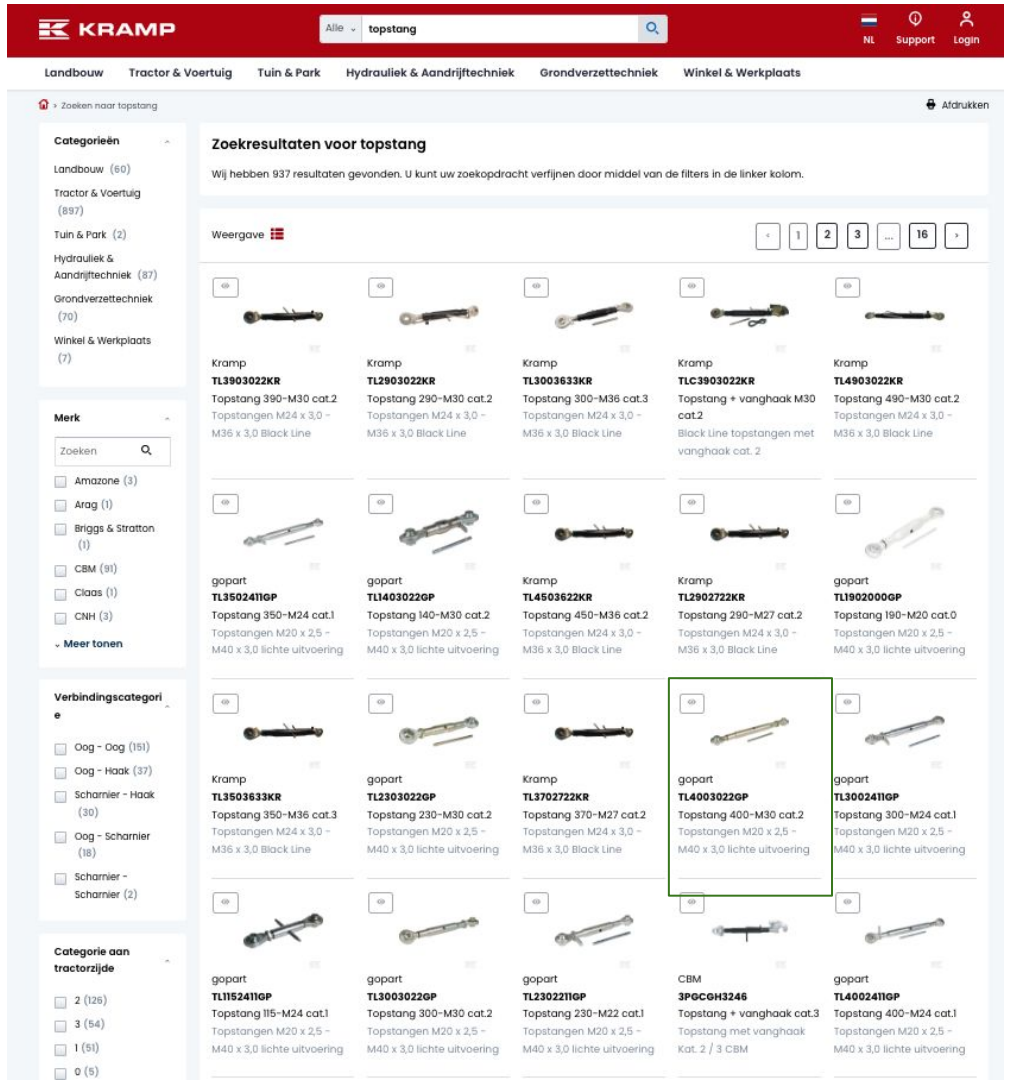
We now calculate DCG for this search as:
DCG = 1/log2(1+3) + 2/log2(1+12) + 2/log2(1+20)
DCG = 1.50
Please not we assume all the other results not to be relevant in this calculation

For iDCG we basically assume the successful clicks should be at the top positions, followed by the product_list_clicks
iDCG = 2/log2(1+2) + 2/log2(1+2) + 1/log2(1+3)
iDCG = 3.02

NDCG = DCG/iDCG
NDCG = 1.50 / 3.02 = 0.50

Current definition DCG and NDCG

Example2



User_2 searches for ‘topstang’,

Clicks ATC on #14

We now calculate DCG for this search as:
DCG = 1/log2(1+14)

DCG = 0.51

Please not we assume all the other results not to be relevant in this calculation

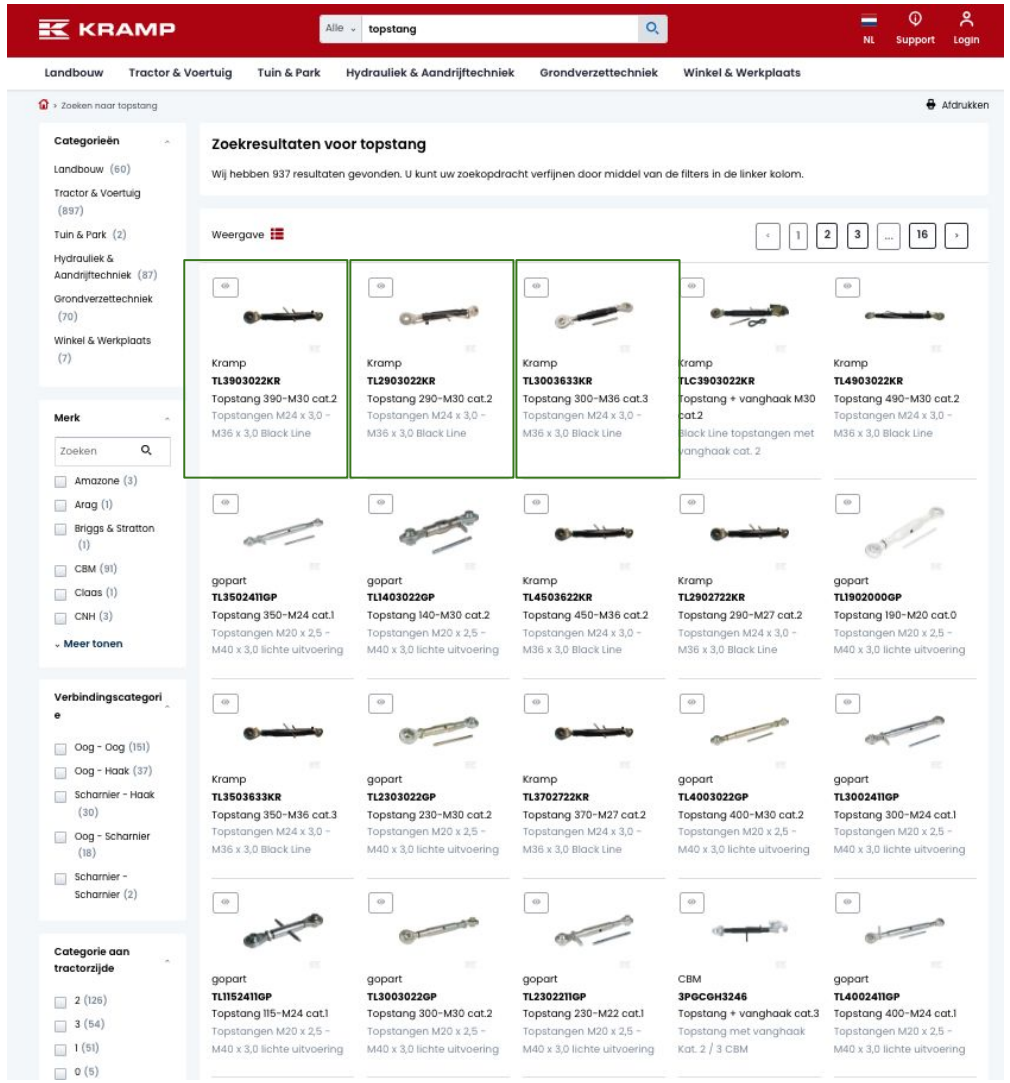
iDCG = 2/log2(1+1)

iDCG = 2

NDCG = DCG/iDCG
NDCG = 0.51/2 = 0.26

Current definition DCG and NDCG

Example3



User_3 searches for ‘topstang’,

Clicks NPC on #1 #2 and #3

We now calculate DCG for this search as:
DCG = 2/log2(1+1) + 2/log2(1+2) + 2/log2(1+3)

DCG = 4.26

Please not we assume all the other results not to be relevant in this calculation

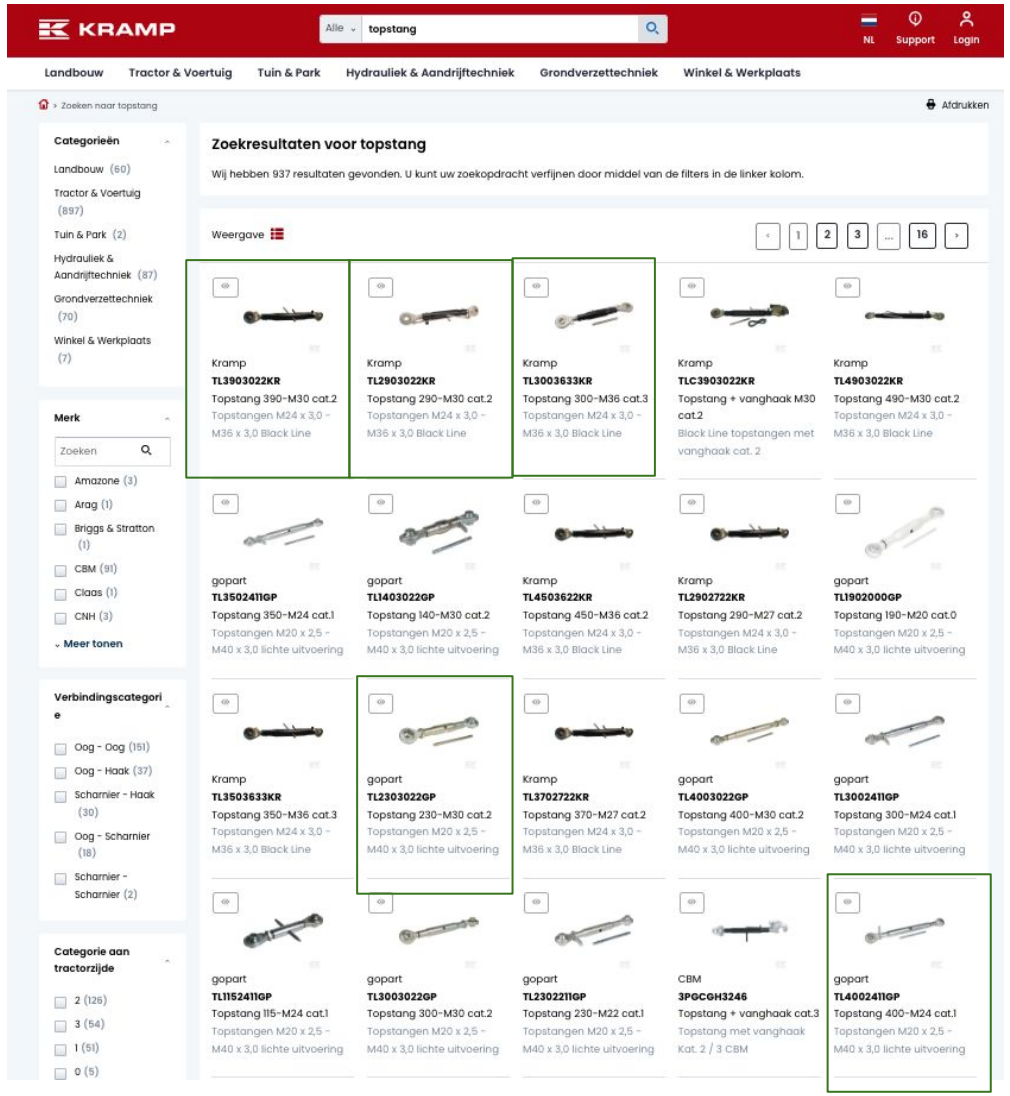
iDCG = 2/log2(1+1) + 2/log2(1+2) + 2/log2(1+3)

iDCG = 4.26

NDCG = DCG/iDCG
NDCG = 4.26/4.26= 1

Using search_query level NDCG

Combining all searches



If we combine the data of all 3 queries and add query level idcg we get:

Average DCG = (1.5+0.51+4.26) / 3
Average DCG = 2.09

Average NDCG = (0.5+0.26+1)/3
Average NDCG = 0.59

DCG, iDCG and NDCG on query level:
Success clicks: 12,20,14,1,2,3
Product list clicks: 3
Use frequency to determine relevance by counting every click position once per user.
DCG = 5.77
iDCG = 6.61

NDCG =5.77/6.61
NDCG = 0.87

NDCG for different click positions

success_clicks	product_list_clicks	NDCG	NDCG adjusted
1,2,5		0.95	0.95
2	1,2	0.86	0.8
1,5		0.85	0.85
4	1,5	0.72	0.65
1,6,10,45		0.71	0.71
2,4,5		0.68	0.68
	3,4	0.57	0.57
2,9,12,17,37		0.55	0.55
4		0.43	0.43
6	7	0.4	0.39
9,10		0.36	0.36
11	5	0.36	0.34
	7	0.33	0.33
	20,21	0.28	0.28
12		0.27	0.27
	18	0.24	0.24
26	33	0.23	0.23
	21	0.22	0.22
37		0.19	0.19