



# 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:

$$\text{DCG} = \sum_{i=1}^p (\text{rel}_i / \log_2(i+1))$$

$$\text{adjusted\_DCG} = \sum_{i=1}^p ((2^{\text{rel}_i} - 1) / \log_2(i+1))$$

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 aren't 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

$$\begin{aligned} \text{iDCG} &= \sum_{i=1}^p (\text{rel}_i / \log_2(i+1)) \\ \text{adjusted_iDCG} &= \sum_{i=1}^p ((2^{\text{rel}_i} - 1) / \log_2(i+1)) \end{aligned}$$

NDCG then simply is DCG / iDCG

# Qualitative versus click modeling

The screenshot shows a search results page for 'topstang' on the KRAMP website. The search bar at the top contains the term 'topstang'. Below the search bar, there are several navigation links: Landbouw, Tractor & Voertuig, Tuin & Park, Hydrauliek & Aandrijftechniek, Grondverzettechniek, Winkel & Werkplaats. On the left side, there are three vertical columns of filters: Categorieën, Merk, and Verbindingscategorie. The Categorieën column includes filters for Landbouw (60), Tractor & Voertuig (897), Tuin & Park (2), Hydrauliek & Aandrijftechniek (87), Grondverzettechniek (70), and Winkel & Werkplaats (7). The Merk column includes filters for Amazon (3), Arag (1), Briggs & Stratton (1), CBM (91), Claas (1), CNH (3), and Meer tonen. The Verbindingscategorie column includes filters for Oog - Oog (15), Oog - Haak (37), Scharnier - Haak (30), Oog - Scharnier (18), and Scharnier - Scharnier (2). The main content area displays a grid of search results for 'topstang' products. Each result includes a small image of the product, its brand name, model number, and a brief description. For example, the first result is 'Kramp TL3903022KR Topstang 390-M30 cat.2 Topstangen M24 x 3,0 - M36 x 3,0 Black Line'. There are 937 results in total.

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 therefore 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

The screenshot shows a search results page for 'topstang' on the KRAMP website. The search bar at the top contains 'topstang'. Below it, there are several filters: 'Alle' (All), 'topstang', a magnifying glass icon, and language buttons for NL, EN, and DE, along with 'Support' and 'Login' links. The main content area is titled 'Zoekresultaten voor topstang' and displays 937 results. It includes a sidebar with category filters for Landbouw, Tractor & Voertuig, Tuin & Park, Hydrauliek & Aandrijftechniek, Grondverzettechniek, Winkel & Werkplaats, and Merk (Brand). There are also filters for Verbindingscategorie (Connection category) and Categorie aan tractorzijde (Category for tractor side). The results are presented in a grid format with product images, names, and descriptions. One result, 'Kramp TL3003633KR Topstang 350-M36 cat.3 Topstangen M24 x 3,0 - M36 x 3,0 Black Line', is highlighted with a blue border.

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:  
 $DCG = 1/\log_2(1+3) + 2/\log_2(1+4) + 2/\log_2(1+5)$

$DCG = 2.14$

*Please note 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

$iDCG = 2/\log_2(1+2) + 2/\log_2(1+2) + 1/\log_2(1+3)$

$iDCG = 3.02$

$NDCG = DCG/iDCG$

$NDCG = 2.14 / 3.02 = 0.71$

# Current definition DCG and NDCG

## Example2

The screenshot shows a search results page for 'topstang' on the KRAMP website. The search bar at the top contains 'topstang'. Below the search bar, there are several filters: 'Alle' (All), 'topstang', a magnifying glass icon, 'NL' (Dutch), 'Support', and 'Login'. The main content area is titled 'Zoekresultaten voor topstang' (Search results for topstang) and states 'Wij hebben 937 resultaten gevonden. U kunt uw zoekopdracht verfijnen door middel van de filters in de linker kolom.' (We have found 937 results. You can refine your search by using the filters in the left column.). The results are displayed in a grid format. The fourth result in the first row is highlighted with a green border. This result is a Kramp topstang (TL3903022KR) with a length of 390 mm, M30 thread, and a thickness of 3.0 mm. It is described as having a black line and a vanghaak (clip). The other results in the grid are also Kramp products with various dimensions and descriptions. The left sidebar contains filters for categories like Landbouw, Tractor & Voertuig, Tuin & Park, Hydrauliek & Aandrijftechniek, Grondverzettechniek, Winkel & Werkplaats, and Merk (Brand). There are also filters for Verbindingscategorie (Connection category) such as Oog - Oog, Oog - Haak, Scharnier - Haak, Oog - Scharnier, and Scharnier - Scharnier. The bottom sidebar shows filters for Categorie aan tractorzijde (Category for tractor side) with options 2, 3, 1, and 0.

User\_2 searches for 'topstang',  
Clicks ATC on #4

We now calculate DCG for this search as:  
 $DCG = 1/\log_2(1+4)$

$$DCG = 0.86$$

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

$$iDCG = 2/\log_2(1+1)$$

$$iDCG = 2$$

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

# Current definition DCG and NDCG

## Example3

The screenshot shows a search results page for 'topstang' on the KRAMP website. The search bar at the top contains 'topstang'. The results are displayed in a grid format with 16 items per page. The first three items are highlighted in green, indicating they are the top-ranked results. Each item has a thumbnail image, the brand name (Kramp or gopart), the model number, and a brief description. The left sidebar contains filters for categories like Landbouw, Tractor & Voertuig, Tuin & Park, Hydrauliek & Aandrijftechniek, Grondverzettechniek, Winkel & Werkplaats, Categorieën, Merk, Verbindingscategorie, and Categorie aan tractorzijde.

Rank	Model Number	Description
1	TL3903022KR	Kramp Topstang 390-M30 cat.2 Topstangen M24 x 3,0 - M36 x 3,0 Black Line
2	TL2903022KR	Kramp Topstang 290-M30 cat.2 Topstangen M24 x 3,0 - M36 x 3,0 Black Line
3	TL3003633KR	Kramp Topstang 300-M36 cat.3 Topstangen M24 x 3,0 - M36 x 3,0 Black Line
4	TL3903022KR	Kramp Topstang + vanghaak M30 cat.2
5	TL4903022KR	Kramp Topstang 490-M30 cat.2 Topstangen M24 x 3,0 - M36 x 3,0 Black Line
6	TL3502411GP	gopart Topstang 350-M24 cat.1 Topstangen M20 x 2,5 - M40 x 3,0 lichte uitvoering
7	TL1403022GP	gopart Topstang 140-M30 cat.2 Topstangen M20 x 2,5 - M40 x 3,0 lichte uitvoering
8	TL4503622KR	Kramp Topstang 450-M36 cat.2 Topstangen M24 x 3,0 - M36 x 3,0 Black Line
9	TL2902722KR	Kramp Topstang 290-M27 cat.2 Topstangen M24 x 3,0 - M36 x 3,0 Black Line
10	TL1902000GP	gopart Topstang 190-M20 cat.0 Topstangen M20 x 2,5 - M40 x 3,0 lichte uitvoering
11	TL3503633KR	Kramp Topstang 350-M36 cat.3 Topstangen M24 x 3,0 - M36 x 3,0 Black Line
12	TL2303022GP	gopart Topstang 230-M30 cat.2 Topstangen M20 x 2,5 - M40 x 3,0 lichte uitvoering
13	TL3702722KR	Kramp Topstang 370-M27 cat.2 Topstangen M24 x 3,0 - M36 x 3,0 Black Line
14	TL3003022GP	gopart Topstang 400-M30 cat.2 Topstangen M20 x 2,5 - M40 x 3,0 lichte uitvoering
15	TL3002411GP	gopart Topstang 300-M24 cat.1 Topstangen M20 x 2,5 - M40 x 3,0 lichte uitvoering
16	TL1152411GP	gopart Topstang 115-M24 cat.1 Topstangen M20 x 2,5 - M40 x 3,0 lichte uitvoering
17	TL3003022GP	gopart Topstang 300-M30 cat.2 Topstangen M20 x 2,5 - M40 x 3,0 lichte uitvoering
18	TL1202211GP	gopart Topstang 220-M22 cat.1 Topstangen M20 x 2,5 - M40 x 3,0 lichte uitvoering
19	3P0C6H3246	CBM Topstang + vanghaak cat.3 Topstang met vanghaak Kat. 2 / 3 CBM
20	TL4002411GP	gopart Topstang 400-M24 cat.1 Topstangen M20 x 2,5 - M40 x 3,0 lichte uitvoering

User\_3 searches for 'topstang',  
Clicks NPC on #1 #2 and #3

We now calculate DCG for this search as:  
 $DCG = 2/\log_2(1+1) + 2/\log_2(1+2) + 2/\log_2(1+3)$

$DCG = 4.26$

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

$iDCG = 2/\log_2(1+1) + 2/\log_2(1+2) + 2/\log_2(1+3)$

$iDCG = 4.26$

$NDCG = DCG/iDCG$

$NDCG = 4.26/4.26 = 1$

# Using search\_query level NDCG

Combining all searches

The screenshot shows the KRAMP website's search results page for the query "topstang". The top navigation bar includes the KRAMP logo, a search bar with the term "topstang", and language and user options. Below the search bar, a breadcrumb trail shows the current search term. The main content area is titled "Zoekresultaten voor topstang" and displays 937 results. The results are presented in a grid format with five columns. Each result card includes a small image of the product, its name, and a brief description. The products listed include various types of topstangs from brands like Kramp, gopart, and CBM, ranging from M20 to M36 sizes.

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

$$\text{Average DCG} = (2.14+0.86+4.26) / 3$$

$$\text{Average DCG} = 2.42$$

$$\text{Average NDCG} = (0.71+0.43+1)/3$$

$$\text{Average NDCG} = 0.71$$

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 =

$$2/\log_2(1+4)+2/\log_2(1+5)+2/\log_2(1+4)+2/\log_2(1+1)+2/\log_2(1+2)+2/\log_2(1+3)$$

$$\text{DCG} = 6.76$$

$$\text{iDCG} = 2/\log_2(1+1) + 2/\log_2(1+1) + 2/\log_2(1+2) + 2/\log_2(1+3) + 2/\log_2(1+4) + 2/\log_2(1+5)$$

$$\text{iDCG} = 5.9$$

$$\text{NDCG} = 6.76/7.9$$

$$\text{NDCG} = 0.86$$



## Example2

# Current definition DCG and NDCG

## Example1

The screenshot shows a search results page for 'topstang' on the KRAMP website. The search bar at the top contains 'topstang'. The results are displayed in a grid format. Several products are highlighted with blue or green boxes:

- Top row, third product: Kramp TL3003633KR Topstang 300-M36 cat.3
- Second row, first product: gopart TL150241IGP Topstang 350-M24 cat.1
- Second row, second product: Kramp TL1403022GP Topstang 140-M30 cat.2
- Third row, first product: Kramp TL3003633KR Topstang 350-M36 cat.3
- Third row, second product: gopart TL12303022GP Topstang 230-M30 cat.2
- Fourth row, third product: Kramp TL1702722KR Topstang 290-M27 cat.2
- Fourth row, fourth product: Kramp TL14003022GP Topstang 400-M30 cat.2
- Fourth row, fifth product: gopart TL19020000GP Topstang 190-M20 cat.0
- Fifth row, first product: Kramp TL115241IGP Topstang 115-M24 cat.1
- Fifth row, second product: gopart TL13003022GP Topstang 300-M30 cat.2
- Fifth row, third product: Kramp TL120221IGP Topstang 230-M22 cat.1
- Fifth row, fourth product: CBM 3PCGCH3246 Topstang + vanghaak cat.3
- Fifth row, fifth product: gopart TL1400241IGP Topstang 400-M24 cat.1

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:  
 $DCG = 1/\log_2(1+3) + 2/\log_2(1+12) + 2/\log_2(1+20)$

$DCG = 1.50$

*Please note 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

$iDCG = 2/\log_2(1+2) + 2/\log_2(1+2) + 1/\log_2(1+3)$

$iDCG = 3.02$

$NDCG = DCG/iDCG$

$NDCG = 1.50 / 3.02 = 0.50$

# Current definition DCG and NDCG

## Example2

The screenshot shows a search results page for 'topstang' on the KRAMP website. The search bar at the top contains the term 'topstang'. Below the search bar, there are several filters on the left side, including categories like Landbouw, Tractor & Voertuig, Tuin & Park, Hydrauliek & Aandrijftechniek, Grondverzettechniek, Winkel & Werkplaats, and more detailed filters for Merk (Brand) and Verbindingscategorie (Connection category). The main area displays a grid of search results, each showing a product image, the brand name, model number, and a brief description. The results are paginated at the top right. A specific result, 'TL3002411GP' by gopart, is highlighted with a green border.

User\_2 searches for 'topstang',  
Clicks ATC on #14

We now calculate DCG for this search as:  
 $DCG = 1/\log_2(1+14)$

$DCG = 0.51$

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

$iDCG = 2/\log_2(1+1)$

$iDCG = 2$

$NDCG = DCG/iDCG$

$NDCG = 0.51/2 = 0.26$

# Current definition DCG and NDCG

## Example3

The screenshot shows a search results page for 'topstang' on the KRAMP website. The search bar at the top contains 'topstang'. Below the search bar, there are filters for categories like Landbouw, Tractor & Voertuig, Tuin & Park, Hydrauliek & Aandrijftechniek, Grondverzettechniek, and Winkel & Werkplaats. On the left, there are detailed filters for Categorieën (Landbouw, Tractor & Voertuig, Tuin & Park, Hydrauliek & Aandrijftechniek, Grondverzettechniek, Winkel & Werkplaats), Merk (Amazona, Arag, Briggs & Stratton, CBM, Claas, CNH, etc.), Verbindingscategorie (Oog - Oog, Oog - Haak, Scharnier - Haak, Oog - Scharnier, Scharnier - Scharnier), and Categorie aan tractorzijde (2, 3, 1, 0). The main area displays a grid of search results with images, product names, and descriptions. The results are paginated from 1 to 16.

User\_3 searches for 'topstang',  
Clicks NPC on #1 #2 and #3

We now calculate DCG for this search as:  
 $DCG = 2/\log_2(1+1) + 2/\log_2(1+2) + 2/\log_2(1+3)$

$DCG = 4.26$

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

$iDCG = 2/\log_2(1+1) + 2/\log_2(1+2) + 2/\log_2(1+3)$

$iDCG = 4.26$

$NDCG = DCG/iDCG$

$NDCG = 4.26/4.26 = 1$

# Using search\_query level NDCG

Combining all searches

The screenshot shows a search results page for 'topstang' on the KRAMP website. The search bar at the top contains 'Alle topstang'. The results are displayed in a grid format under the heading 'Zoekresultaten voor topstang'. There are 937 results found. The results are categorized by brand (Kramp, gopart) and part number. Some results are highlighted with green boxes, specifically the first three rows of the first two columns. The left sidebar contains filters for categories like Landbouw, Tractor & Voertuig, Tuin & Park, Hydrauliek & Aandrijftechniek, Grondverzettechniek, Winkel & Werkplaats, and various sub-categories like 'Topstangen' and 'Verbindingscategorie'. The bottom sidebar shows filters for 'Categorie aan tractorzijde'.

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

$$\text{Average DCG} = (1.5+0.51+4.26) / 3$$

$$\text{Average DCG} = 2.09$$

$$\text{Average NDCG} = (0.5+0.26+1)/3$$

$$\text{Average NDCG} = 0.59$$

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.

$$\text{DCG} = 5.77$$

$$\text{iDCG} = 6.61$$

$$\text{NDCG} = 5.77/6.61$$

$$\text{NDCG} = 0.87$$

# 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