

Ранжирование

INVENTORY

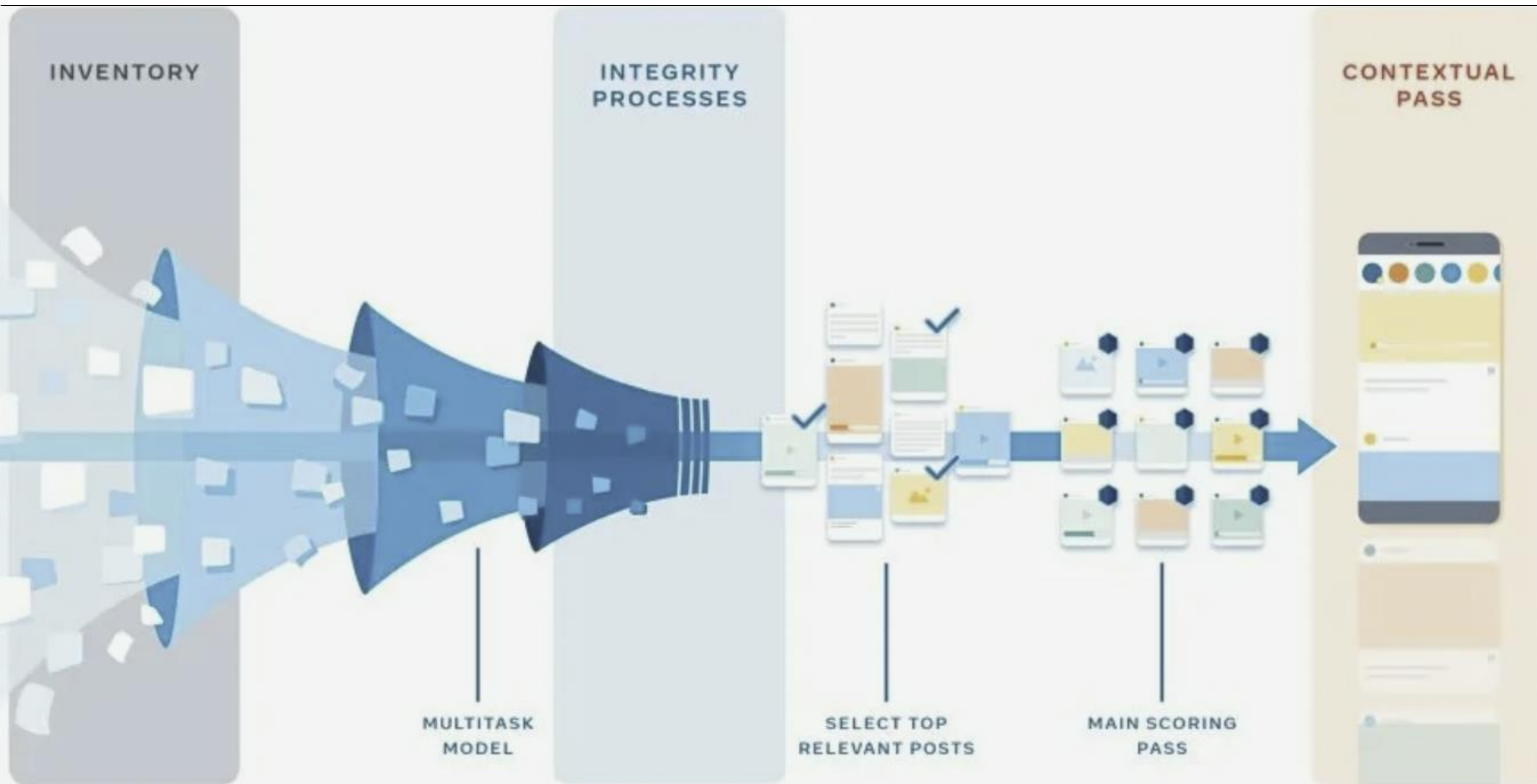
INTEGRITY
PROCESSES

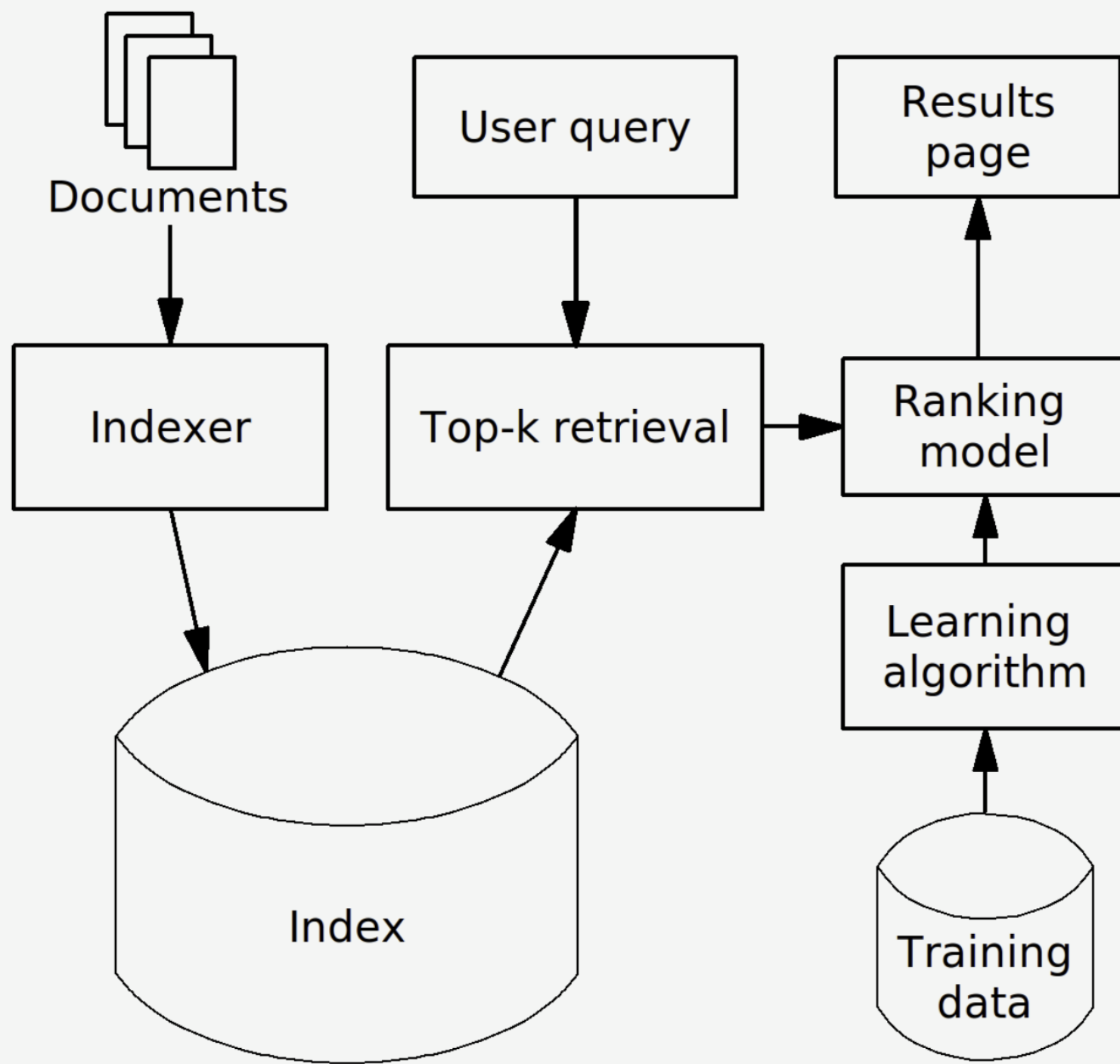
CONTEXTUAL
PASS

MULTITASK
MODEL

SELECT TOP
RELEVANT POSTS

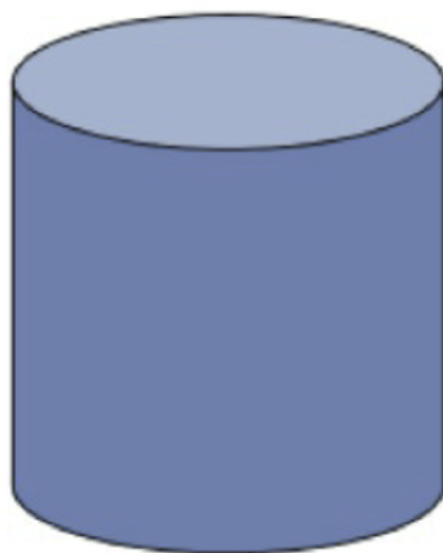
MAIN SCORING
PASS





documents

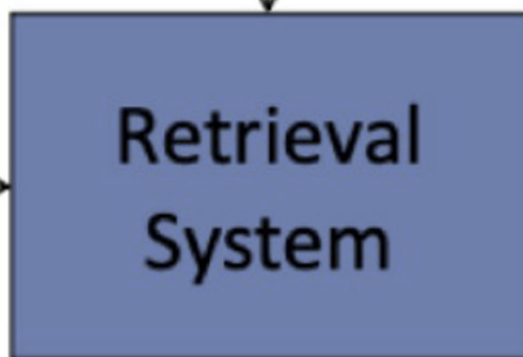
$$D = \{d_1, d_2, \dots, d_N\}$$



ranking based on
relevance

query

q



$d_{q,1}$
 $d_{q,2}$
 \vdots
 d_{q,n_q}

ranking of documents



Матчинг

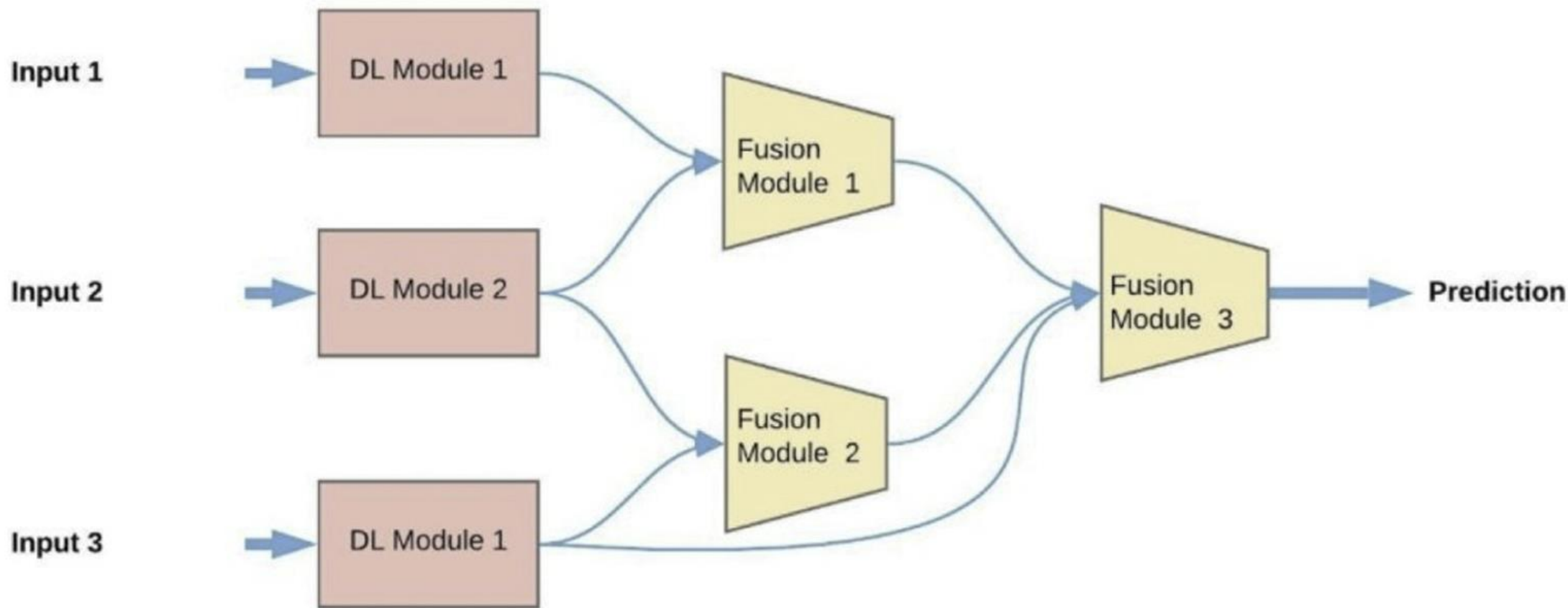
Магазин 1



Магазин 2



**Введение
понятия SKU**

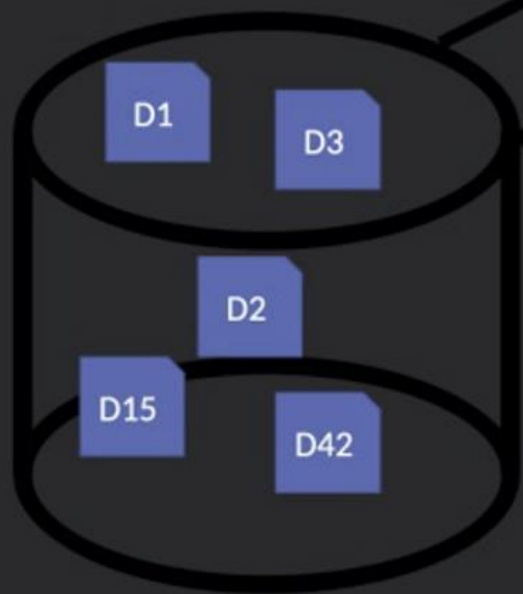


Минусы и приложения матчинга



Метрики ранжирования

Запросы



Документы

Всего релевантных
10

Модель 1

Модель 2

Для Q1:

D1 + Precision: 2/3
D2 + Recall: 2/10
D3 -

D1 + Precision: 3/6
D3 - Recall: 3/10
D4 +
D5 -
D6 +
D7 -

Оценки
релевантности

Q1 D1 +
Q1 D2 +
Q1 D3 -
Q1 D4 -

...

Q2 D1 -
Q2 D2 +

...

Q40 D42 +

Метрики ранжирования

Точность, полнота, F-мера, PR кивая

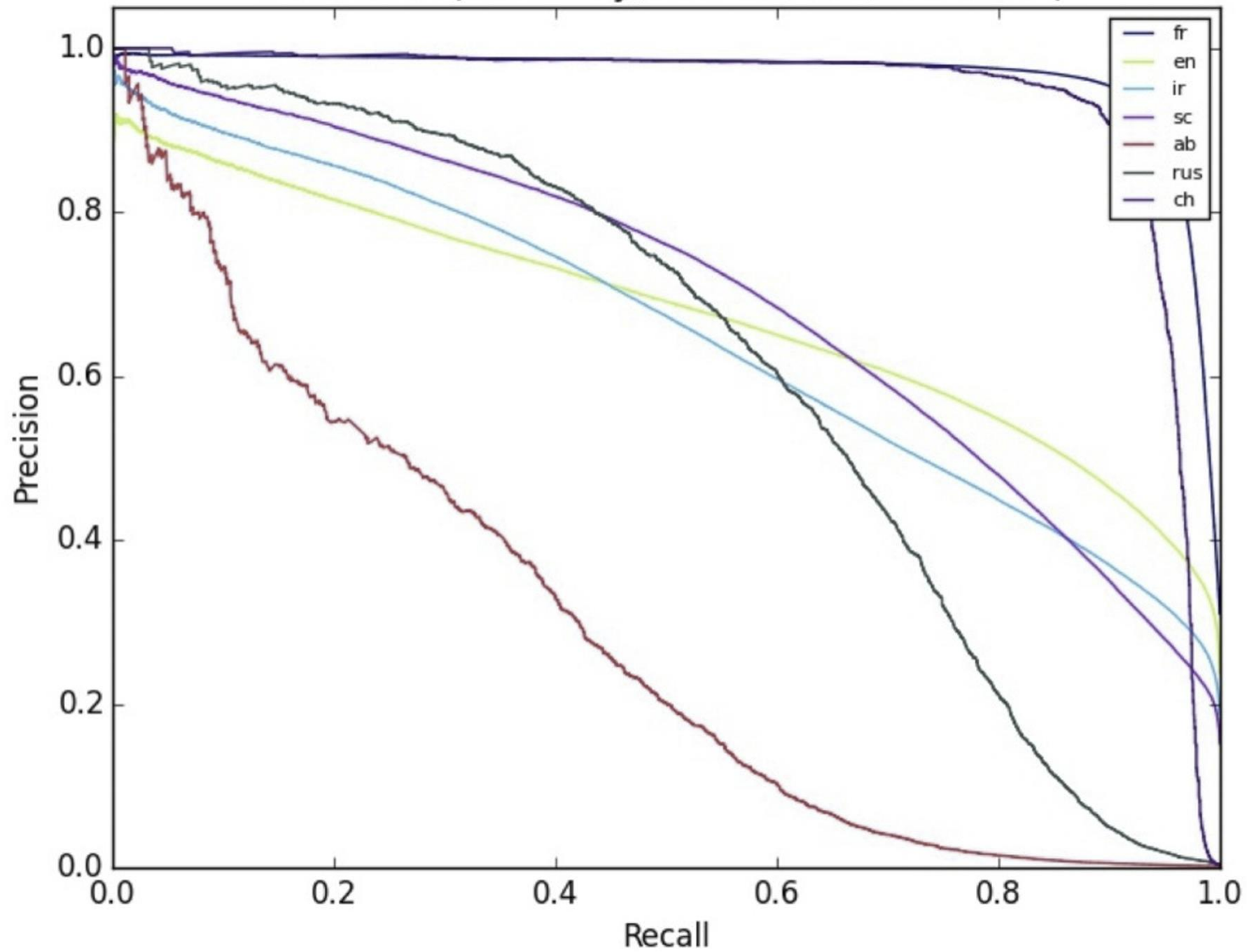
$$precision = \frac{|\{relevant\ documents\}| \cap |\{retrieved\ documents\}|}{|\{retrieved\ documents\}|}$$

$$recall = \frac{|\{relevant\ documents\}| \cap |\{retrieved\ documents\}|}{|\{relevant\ documents\}|}$$

PR-кривая

a01	1	6.4	1
a01	3	0.7	0
b02	2	0.6	1
c03	2	-0.8	0

P-R curve (LR-binary, 14 features, N=2000k)



Метрики ранжирования

Average precision, mean average precision
CG, Normalized DCG, PFound
MRR, Kendall's tau

$$AP = \sum_K (Recall@k - Recall@[k - 1]) \cdot Precision@k$$

(Кол-во корректных предсказаний) / k

k	Document ID	Predicted Relevance	Actual Relevance
1	06	0.90	Relevant (1.0)
2	03	0.85	Not Relevant (0.0)
3	05	0.71	Relevant (1.0)
4	00	0.63	Relevant (1.0)
5	04	0.47	Not Relevant (0.0)
6	02	0.36	Relevant (1.0)
7	01	0.24	Not Relevant (0.0)
8	07	0.16	Not Relevant (0.0)

Всего релевантных нашли

Скользящая сумма

1

$0 + 1/1 = 1$

1

1

2

$1 + 2/3 = 1.67$

3

$1.67 + 3/4 = 2.42$

3

2.42

4

$2.42 + 4/6 = 3.08$

4

3.08

4

3.08

$3.08 / 4 = 0.77$

MAP — среднее AP по всем запросам Q

$$MAP = \frac{\sum_{q=1}^Q AP(q)}{Q}$$

“Gain”

Cumulative Gain

D1 3

3

D2 2

3+2

D3 1

3+2+1

D4 1

3+2+1+1

D5 3

3+2+1+1+3

D6 1

3+2+1+1+3+1

D7 2

3+2+1+1+3+1+2

“Gain”		Discounted Cumulative Gain
D1	3	3
D2	2	$3 + 2/\log(3)$
D3	1	$3 + 2/\log(3) + 1/\log(4)$
D4	1	$3 + 2/\log(3) + 1/\log(4) + 1/\log(5)$
D5	3	...
D6	1	$DCG@7 = 3 + 2/\log(3) + \dots + 2/\log(8) \sim 7.38$
D7	2	
		$IdealDCG@7 = 3 + 3/\log(3) + \dots + 1/\log(8) \sim 7.83$

$$nDCG@k = \frac{DCG@k}{IdealDCG@k}$$

$$nDCG \in [0, 1]$$

$$pfound = \sum_{i=1}^n pLook[i] \cdot pRel[i]$$

$$pLook[i] = pLook[i - 1] \cdot (1 - pRel[i - 1]) \cdot (1 - pBreak)$$

$pBreak$ — вероятность прекращения просмотра выдачи

$$MRR = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{rank_i}$$

$$\tau = \frac{(\text{number of concordant pairs}) - (\text{number of discordant pairs})}{\binom{n}{2}}$$

$$\tau \in [-1, 1]$$