

Learning to Rank journey at GetYourGuide: Our Logbook

Felipe Besson • 11 Jun 2018

About GetYourGuide

GYG is the leading global marketplace for tours and activities

- ❑ 33k+ activities
- ❑ 7K+ destinations
- ❑ 415+ employees
 - ❑ 20 in the Data team
- ❑ Based in Berlin and Zurich



Full-text Search

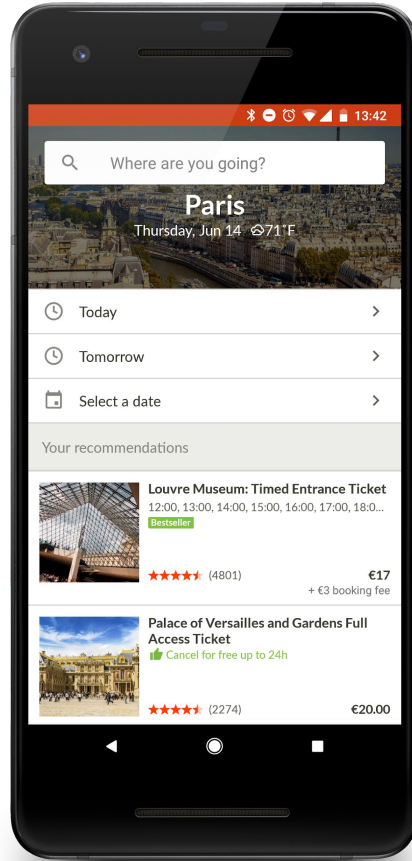
- Location driven
- Discovery

Rank

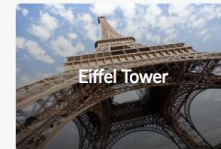
Business metrics

+

Text Relevance



Top attractions



Location Pages (LPs)

- ❑ Location driven
- ❑ Dates are **very** important
- ❑ **High-intent customers**
- ❑ **Paid traffic**

Rank

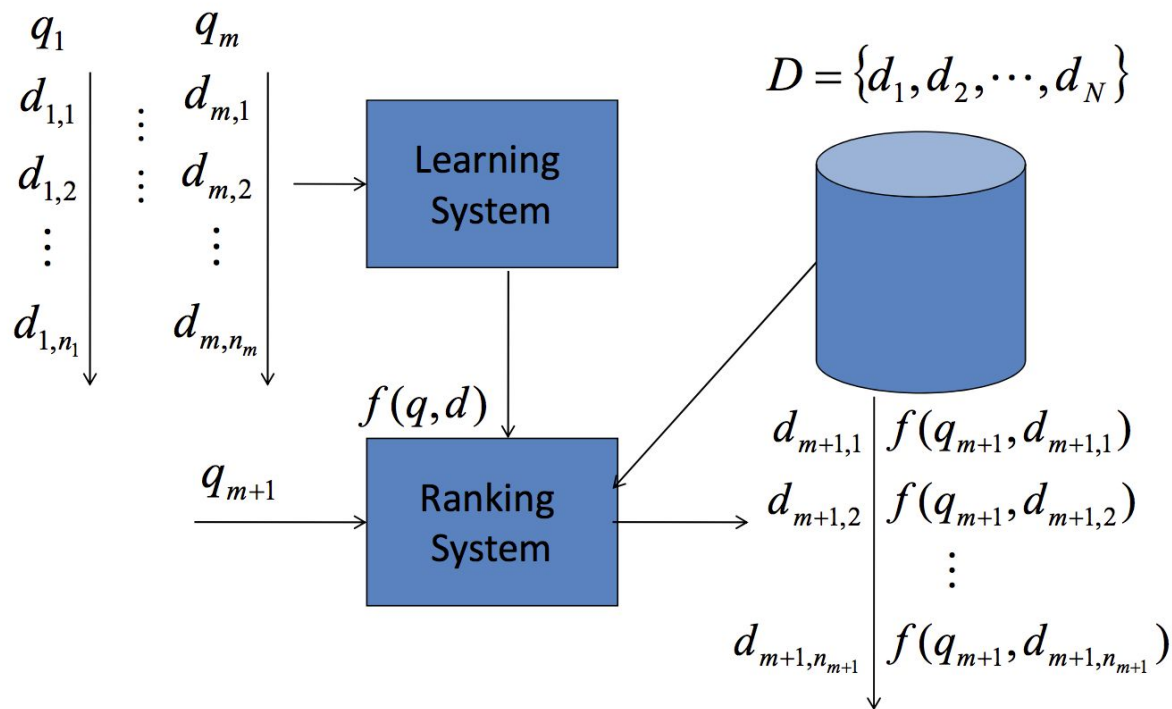
Business metrics

The screenshot shows the Paris location page on the Get Your Guide website. At the top, there's a search bar with the text "Where are you going?" and a magnifying glass icon. To the right, there are links for "English", "EUR (€)", "Cart", "Help", "Log in", and "Sign up". The main heading is "Paris" in a large, white font, with the subtitle "Must-see sights and things to do". Below this, there's a search filter for "Enter your dates to find available activities" with "from" and "to" input fields and a "Check availability" button. A banner below the search filter lists three benefits: "GetYourGuide gives you:", "The best selection More than 34,520 things to do", "The lowest prices We guarantee it!", and "Fast & easy booking Book online to lock in your tickets". The "Top attractions in Paris" section features a grid of attraction cards. The first card is for "Eiffel Tower Walk to the 2nd Floor & Optional Summit Access", priced from €32, with 402 reviews and a "LIKELY TO SELL OUT" badge. The second card is for "Louvre Museum: Timed Entrance Ticket", priced from €17 (+ €3 booking fee), with 4681 reviews and a "BESTSELLER" badge. Other cards include "Paris Illuminations 1-Hour Cruise" (€15, 2419 reviews), "Palace of Versailles and Gardens Full Access Ticket" (€20, 2233 reviews), "Paris: Skip-the-Line Arc de Triomphe Rooftop Tickets", "Paris: 1, 2, or 3-Day Hop-on Hop-off Ticket", "Eiffel Tower: Second Floor Priority Access", and "Disneyland Paris 1-Day Ticket" (10 hours - 1 day).

Problems with LP Ranking

- ❑ Focus on business metrics
- ❑ Customer intentions (search keywords)
 - ❑ "Eiffel Tower ticket" = "Eiffel Tower restaurant"
- ❑ Difficult to introduce new and diverse products
- ❑ **We needed to learn how to rank activities in LPs!**

Learning to Rank (LTR)



Extracted from ACML 2009 Tutorial Nov. 2, 2009 Nanjing

Our Logbook

- ☐ Scope
- ☐ Decisions
- ☐ Tools and data pipelines
- ☐ **Lessons learned**

Not covered by this talk

- ☐ Heavy statistical modeling
- ☐ Evaluation of models



A three-masted sailing ship with white sails is shown on a dark sea under a blue sky. The ship is viewed from a side profile, moving towards the right. The sails are partially unfurled. The ship's hull is dark, and the deck is visible. The text "First iteration (V1) Scope and decisions" is overlaid in large white font in the center of the image.

First iteration (V1) Scope and decisions

Learning to Rank (LTR) at GYG

Apply Machine Learning to introduce **relevance** factors into our ranking formula

Use our user intention data to have a dynamic LP ranking

V1 Focus

- ❑ Vertical: Points of Interest
 - ❑ Ticket, Tour, Museum, Historic site, park, ...
- ❑ Only in English (we have 22 languages)
- ❑ Location pages have no explicit user query
 - ❑ Search Keywords:

"Statue of Liberty boat tour"



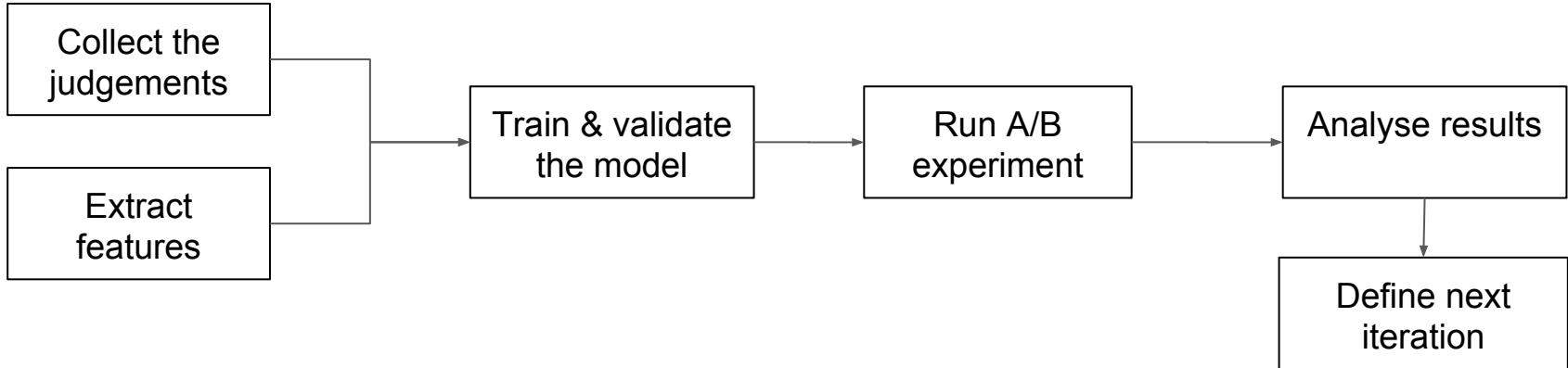
location



intention

MVP mindset

Follow the standard steps of a LTR solution



We started the journey!



Judgement List

- ❑ Ground truth set (Golden set)
- ❑ Highest traffic keywords

keyword	location id	doc id	judgement
eiffel tower summit access	2600 #Eiffel tower	42 #Optional Summit Access	?
eiffel tower summit access	2600 #Eiffel tower	24 #Paris cooking class	?
audio guide	2600 #Eiffel tower	41 #Audio Guided Eiffel Tower	?

query (qid)

Human labeling judgement list

- ❑ Judgements were collected from **Domain Experts**
 - ❑ Internal stakeholders of GYG
- ❑ Judgement scale from **0 - 3**
- ❑ ~ 30k judgements
- ❑ Pre analysis of current rank
 - ❑ $NDCG@7 = \mathbf{0.55}$

Human labeling judgement list



- ✓ Good approach when data is incomplete/inconsistent
- ✓ When what is a relevant result is still unclear
- ✓ No need to normalize queries deeply



- ✗ Relevance is subjective from user to user
- ✗ Hard to scale
- ✗ Crowdsourcing is expensive

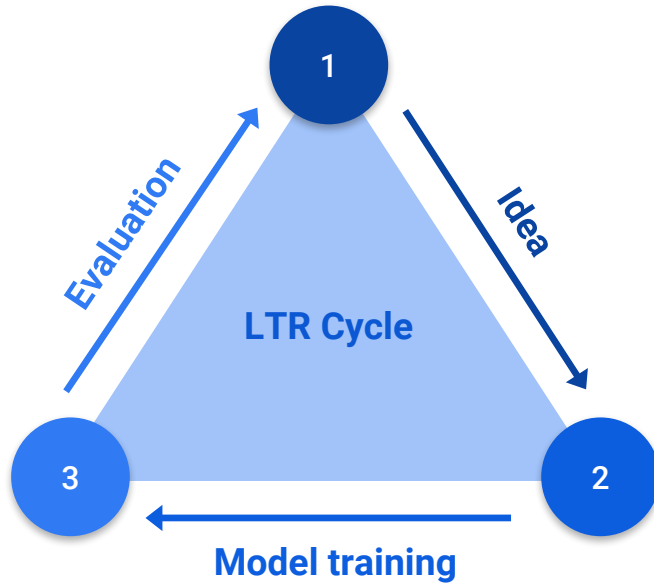


Enriching Judgements with features

Feature Engineering

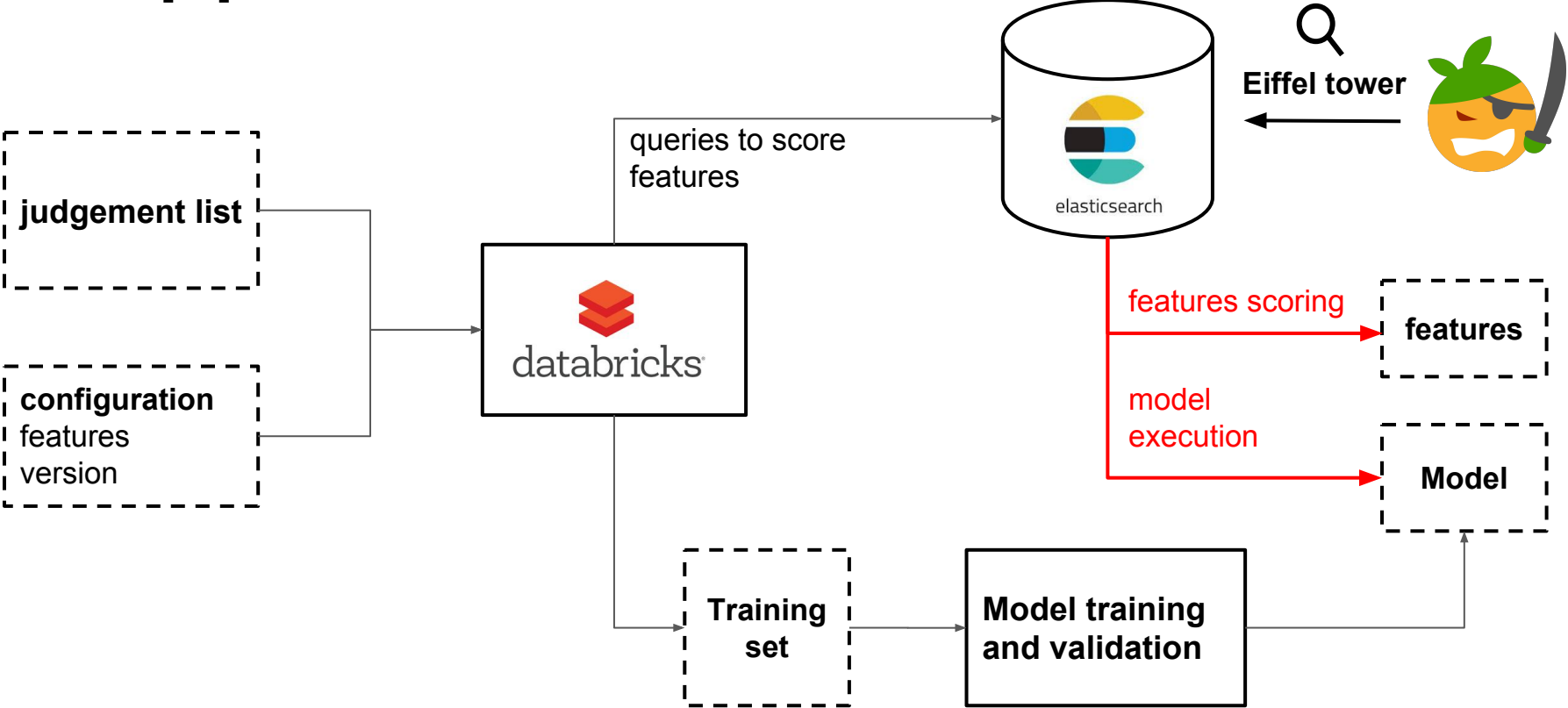
Query document	Business metrics	Document
<ul style="list-style-type: none">● BM25 of single text fields● Multi-match combinations	<ul style="list-style-type: none">● Raw metrics: clicks, bookings, impressions● Rates: CTR, CR	<ul style="list-style-type: none">● Activity attributes: price, duration, # reviews● Categories (Dummy variables)

Why to develop a data pipelines ?



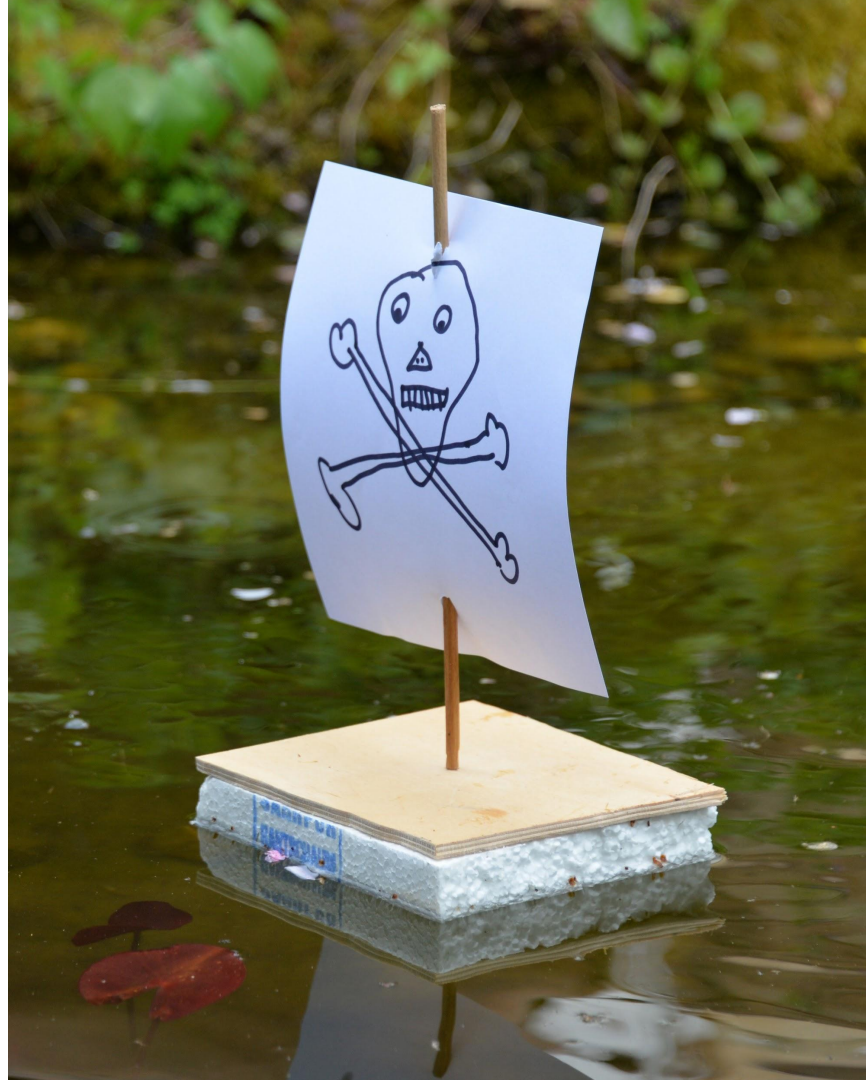
- ❑ Reliable
- ❑ Reproducible
- ❑ Fast Iterations

First pipeline to collect features



Problems

- ❑ Managing requests to ES
 - ❑ 300 features x 30k rows = **9mi queries**
- ❑ Versioning features and training sets
- ❑ How to execute complex models on query time ?

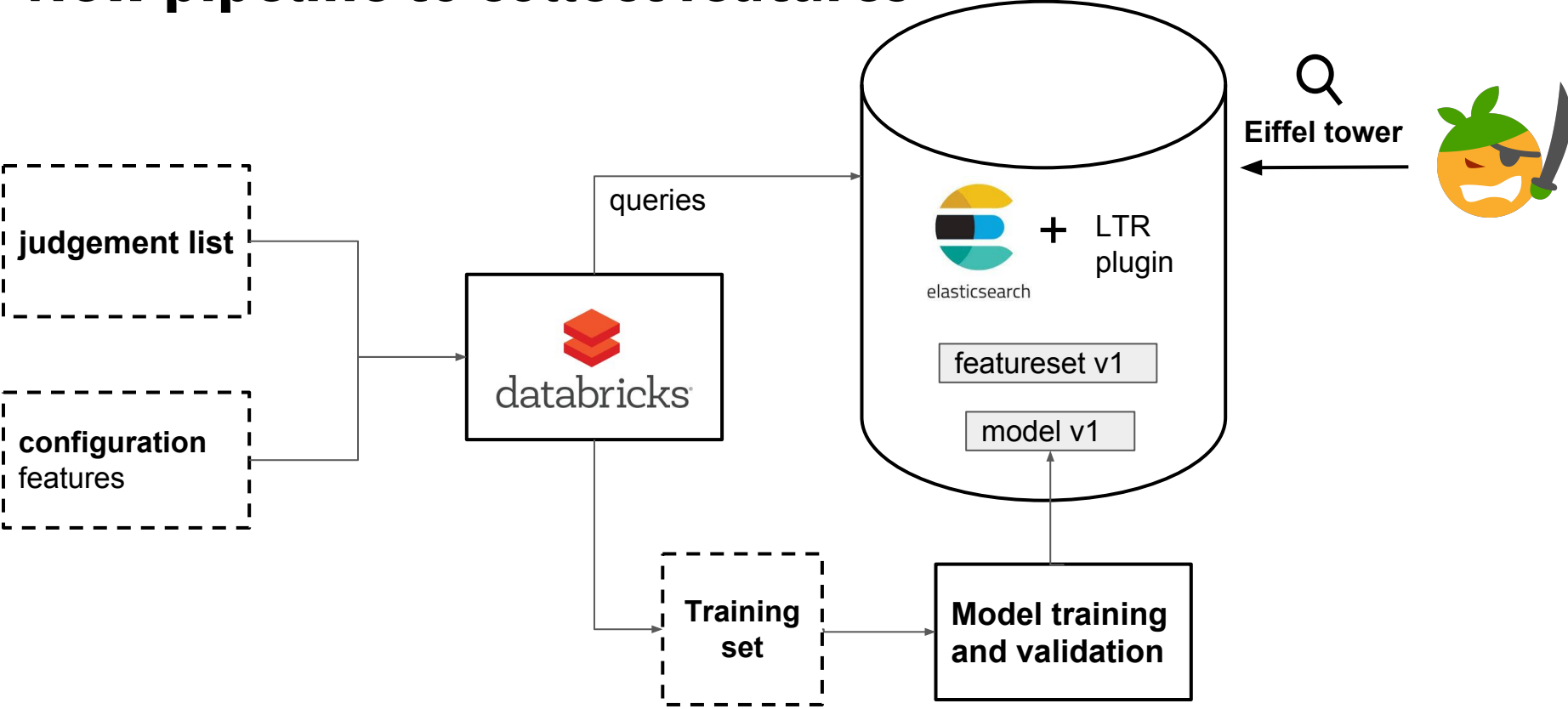


Elasticsearch LTR plugin

- ❑ Developed by OpenSource Connections
- ❑ **RankLib and XGBoost**
- ❑ Features and models storage (versioning)
- ❑ Endpoint for scoring features
 - ❑ Fewer queries to ES (one per judgement)

repo: <https://github.com/o19s/elasticsearch-learning-to-rank>

New pipeline to collect features



Our pipeline



- ✓ Iterate faster to get new datasets
- ✓ Versioning of models and feature sets
- ✓ Simplicity to run models



- ✗ Not automated from end to end (but good enough)

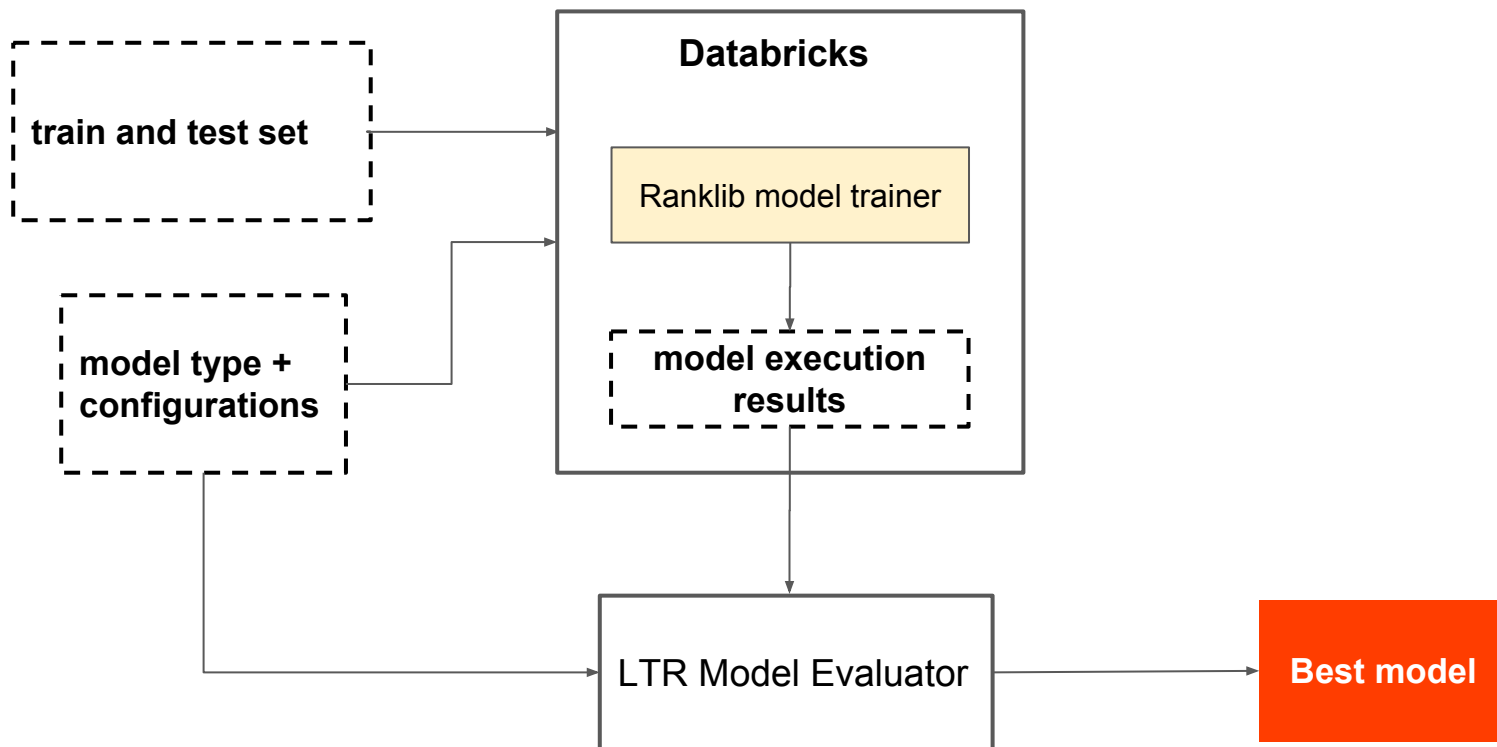
A three-masted sailing ship with white sails is shown on a dark sea under a blue sky. The ship is the central focus, with its sails partially unfurled. The text 'Training and validating Models' is overlaid in large white font across the center of the image.

Training and validating Models

Goals

- ❑ Have a reasonable model for location pages
 - ❑ relevance + business metrics
- ❑ Evaluation metric: **NDCG@10**
- ❑ Success (business): **↑ CTR** (Click-Through Rate)
- ❑ Constraints
 - ❑ Do not include user features

Pipeline to train and test Ranklib models



Model training pipeline



- ✓ Iterate faster to try new models
- ✓ Automate models evaluation



- ✗ Initial overhead to distribute Ranklib, we could use XGBoost
- ✗ Analyze feature importance with Ranklib

Best V1 Model

❑ LambdaMart ❑ $\text{NDCG@10} = 0.9282$










Query document	Business metrics	Document
<ul style="list-style-type: none">● Title● Highlight● Description● Best field multi-match	<ul style="list-style-type: none">● Clicks● Bookings● Impressions● CR	<ul style="list-style-type: none">● #Reviews● Review rating● Deal price● Best seller



Best V1 model didn't work










🔍 "Eiffel tower skip-the-line ticket"

CURRENT RANK

	Eiffel Tower: Second Floor Priority Access or Summit Ticket ★★★★★ 4918 Duration: 2 hours LIKELY TO SELL OUT	From US\$51.53
	Eiffel Tower Walk to the 2nd Floor & Optional Summit Access ★★★★★ 130 Duration: 1.5 hours LIKELY TO SELL OUT	From US\$40.22
	Eiffel Tower Guided Climb with Optional Summit Access ★★★★★ 240 Duration: 2 hours LIKELY TO SELL OUT	From US\$42.73
	Paris Eiffel Tower Skip-the-Line Tour and Summit Access ★★★★★ 3753 Duration: 90 minutes	From US\$86.72
	Paris: Eiffel Tower by Night Skip-the-Line Guided Tour ★★★★★ 126 Duration: 1.5 hours	From US\$45.25 <small>US\$56.56</small>
	Skip-the-Line Eiffel Tower Ticket and 2nd Level Tour ★★★★★ 48 Duration: 90 minutes	From US\$57.81
	Eiffel Tower: Skip-the-Line Summit Tour ★★★★★ 6 Duration: 90 minutes	From US\$75.28
	Eiffel Tower Priority Access to the Second Floor ★★★★★ 127 Duration: 2 hours	From US\$61.58
	Skip-the-Line Eiffel Tower Ticket and Small Group Tour ★★★★★ 70 Duration: 1.5 hours	From US\$50.90 <small>US\$56.56</small>

Get Your Guide

MODEL RANK


	Skip The Line: 2-Hour Guided Eiffel Tower Tour ★★★★★ 198 Duration: 2 hours LIKELY TO SELL OUT	From US\$86
	Eiffel Tower Dinner With River Cruise ★★★★★ 21 Duration: 4 hours LIKELY TO SELL OUT	From US\$209
	Guided Tour to Eiffel Tower 2nd Floor with Priority Access <i>No reviews yet</i> Duration: 90 minutes NEW ACTIVITY	<i>Intro Special</i> From US\$55
	Paris: City Tour and Skip-the-Line Eiffel Tower Ticket Duration: 2 hours	From US\$273
	Louvre, Notre Dame, Seine Cruise & Eiffel Tower Lunch ★★★★★ 45 Duration: 7.5 hours	From US\$215
	Eiffel Tower, Hop on Hop Off & Option for Cruise Ticket ★★★★★ 1 Duration: 1 day	From US\$110
	Paris: Eiffel Tower Dinner & Seine River Cruise ★★★★★ 20 Duration: 4 hours	From US\$219
	Eiffel Tower: Skip-the-Line Summit Tour ★★★★★ 6 Duration: 90 minutes	From US\$75.28
	Priority Eiffel Tower Tickets & Audio Guided Paris Bus Tour ★★★★★ 403 Duration: 3 hours	From US\$86

Main lessons learned

- ❑ Relevance of results for LP
- ❑ Judgement list extraction
- ❑ Quality of our queries
- ❑ Distribution of judgements

What is relevance for our business ?

- ❑ Our use case: Location pages
 - ❑ First point of contact of many visitors
 - ❑ Few rank positions to change
 - ❑ **Business metrics matter (e.g., revenue)**

- ❑ Domain Experts labeling
 - ❑ This document is relevant for this query ? **0 - 3**
 - ❑ This document is a potential conversion ? 

Another approach

- ❑ Data approach for e-commerces
 - ❑ Perceived utility of:
 - ❑ search results (Click through rate)
 - ❑ product page (Add-to-cart)
 - ❑ Overall user satisfaction (Conversion)
 - ❑ Business value (Revenue)

- ❑ Experts could refine judgements collected from data

Reference: [On Application of Learning to Rank for E-Commerce Search](#) by Santu, Sondhi and Zhai (2017)

Quality of our queries

- ❑ Didn't consider real user query but the keyword search engine matches
- ❑ Location part is not relevant for scoring some queries

"Statue of Liberty boat tour"

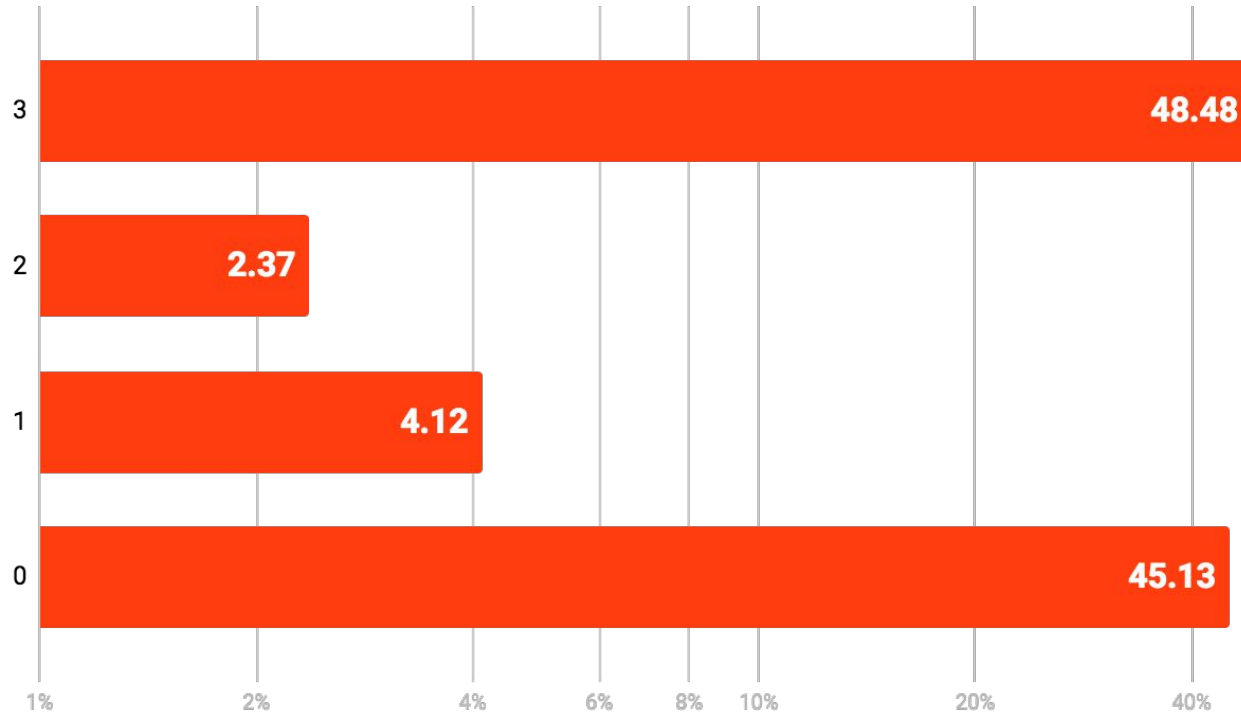


**All results
contain this location**

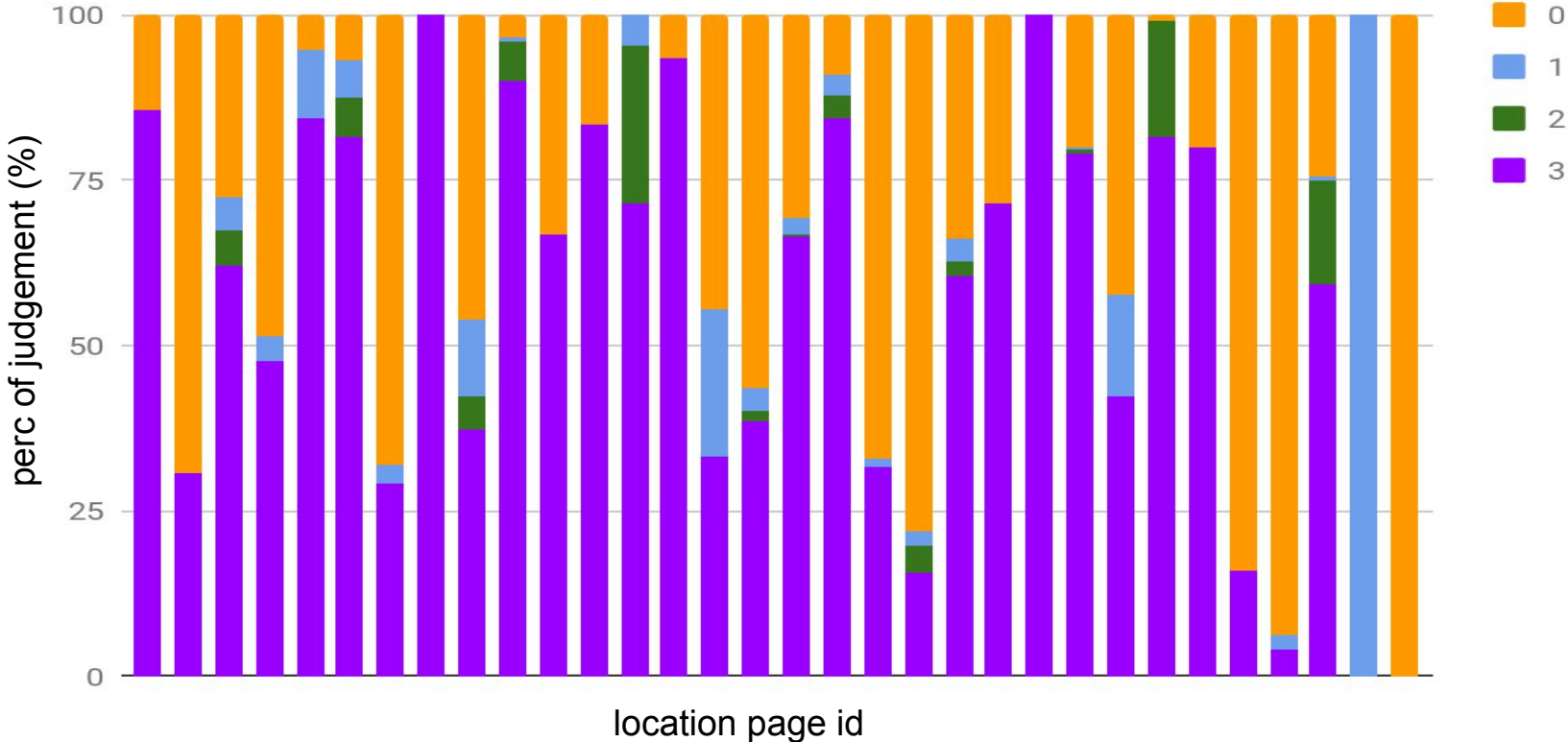


good!

Distribution of our Judgements



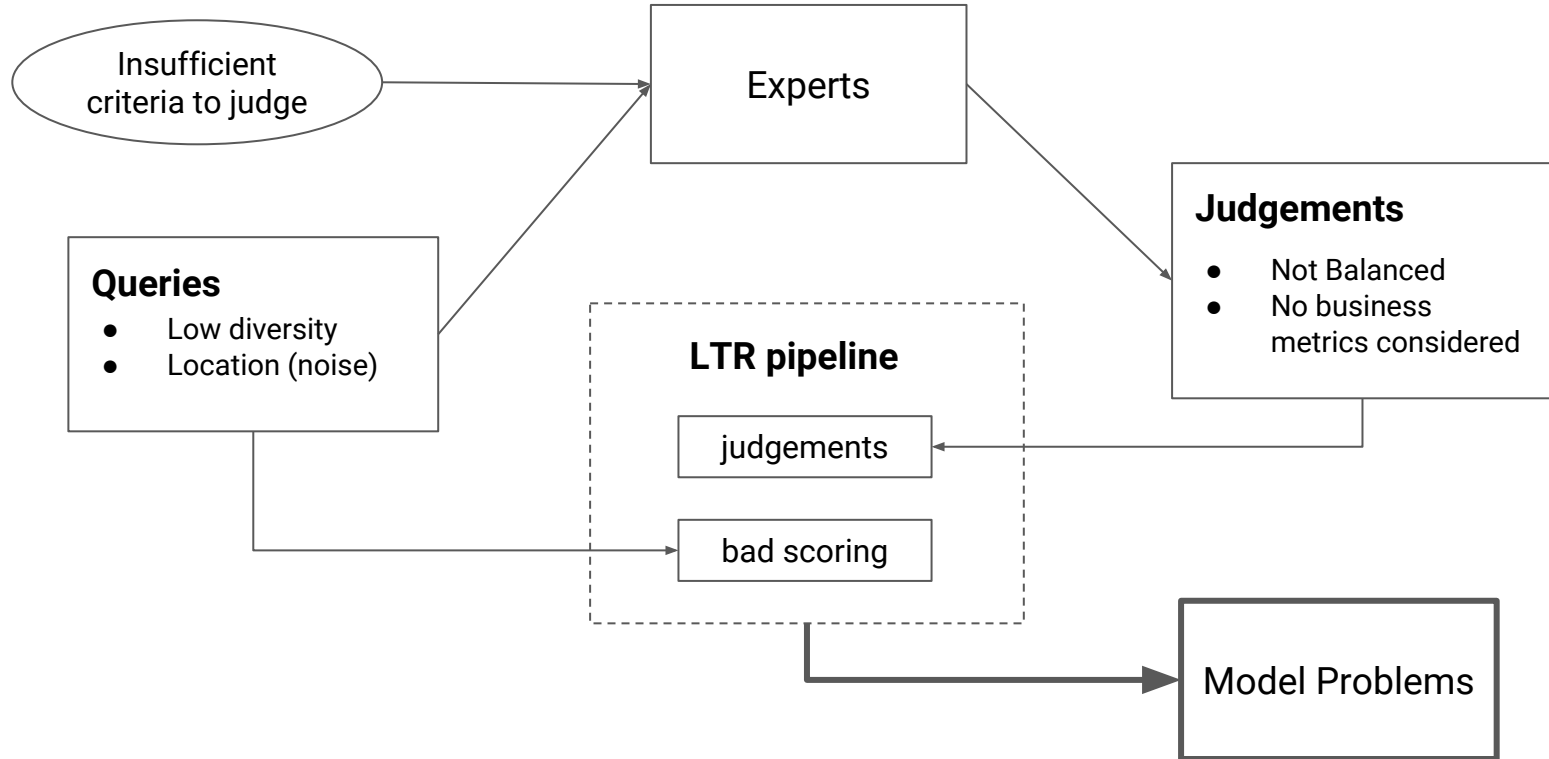
Distribution of our Judgements per page



We have very similar items in LP!

- ❑ **In general:** judgements didn't reflect those differences (binary relevance)
- ❑ **In each page:** very different and not balanced judgement distribution

Everything is connected



Next steps for V2

- ❑ Collect judgements from data
- ❑ Redefine our criteria for measuring relevance
- ❑ Apply LTR in another GYG search features
- ❑ Extract the intentions from the keywords
 - ❑ Query understanding might help
- ❑ Judge the judgements very often

We need to keep sailing!



A three-masted sailing ship with white sails is sailing on a dark sea under a blue sky. The ship has a red hull and white upper sections. The name "JEAN DE LA LOUÏE" is visible on the side of the hull. The ship is positioned in the center of the frame, with the text "Thank you!" and "Questions ?" overlaid on it.

Thank you!

Questions ?