# STELLA Infrastructure and Tech Details

Philipp Schaer, Timo Breuer, Leyla Jael Castro, Benjamin Wolff,
Johann Schaible and Narges Tavakolpoursaleh

Living Labs for Academic Search (LiLAS)


CLEF 2021; September, 21-24, 2021; online event (from Bucharest - Romania).

# Why do we need another infrastructure?

# STELLA complements existing infrastructures

**Table 1** Evaluation infrastructures with respect to online/offline evaluations, domain specificity, and reproducibility

| Infrastructure | Description | Evaluation | | Domains | Reproducibility | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Online | Offline | | P | R | I | M | A | D |
| BioASQ | BioASQ is an initiative for semantic indexing and question answering [27] | ◐ | ● | Biomedical text indexation and Question Answering | ○ | ● | ○ | ● | ● | ● |
| TIRA | Underlying platform of the PAN lab dedicated to digital text forensics [23] | ○ | ● | Multi-domain, e.g., text forensics (author profiling), clickbait detection | ● | ● | ● | ● | ◐ | ● |
| LL4IR | Living labs infrastructure for ad-hoc search using interleaving [17] | ● | ○ | Multi-domain, e.g., online shop, academic repositories and search systems | ○ | ● | ○ | ● | ● | ● |
| NewsREEL | Living lab infrastructure to evaluate news article recommenders [15] | ◐ | ● | Commercial news articles recommendation | ◐ | ● | ○ | ◐ | ◐ | ● |

○ requirement not fulfilled, ◐ requirement partially fulfilled, ● requirement fulfilled

# STELLA complements existing infrastructures

**Table 1** Evaluation infrastructures with respect to online/offline evaluations, domain specificity, and reproducibility

| Infrastructure | Description | Evaluation | | Domains | Reproducibility | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Online | Offline | | P | R | I | M | A | D |
| BioASQ | BioASQ is an initiative for semantic indexing and question answering [27] | ◐ | ● | Biomedical text indexation and Question Answering | ○ | ● | ○ | ● | ● | ● |
| TIRA | Underlying platform of the PAN lab dedicated to digital text forensics [23] | ○ | ● | Multi-domain, e.g., text forensics (author profiling), clickbait detection | ● | ● | ● | ● | ◐ | ● |
| LL4IR | Living labs infrastructure for ad-hoc search using interleaving [17] | ● | ○ | Multi-domain, e.g., online shop, academic repositories and search systems | ○ | ● | ○ | ● | ● | ● |
| NewsREEL | Living lab infrastructure to evaluate news article recommenders [15] | ◐ | ● | Commercial news articles recommendation | ◐ | ● | ○ | ◐ | ◐ | ● |

○ requirement not fulfilled, ◐ requirement partially fulfilled, ● requirement fulfilled

# STELLA complements existing infrastructures

**Table 1** Evaluation infrastructures with respect to online/offline evaluations, domain specificity, and reproducibility

| Infrastructure | Description | Evaluation | | Domains | Reproducibility | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Online | Offline | | P | R | I | M | A | D |
| BioASQ | BioASQ is an initiative for semantic indexing and question answering [27] | ◐ | ● | Biomedical text indexation and Question Answering | ○ | ● | ○ | ● | ● | ● |
| TIRA | Underlying platform of the PAN lab dedicated to digital text forensics [23] | ○ | ● | Multi-domain, e.g., text forensics (author profiling), clickbait detection | ● | ● | ● | ● | ◐ | ● |
| LL4IR | Living labs infrastructure for ad-hoc search using interleaving [17] | ● | ○ | Multi-domain, e.g., online shop, academic repositories and search systems | ○ | ● | ○ | ● | ● | ● |
| NewsREEL | Living lab infrastructure to evaluate news article recommenders [15] | ◐ | ● | Commercial news articles recommendation | ◐ | ● | ○ | ◐ | ◐ | ● |

○ requirement not fulfilled, ◐ requirement partially fulfilled, ● requirement fulfilled

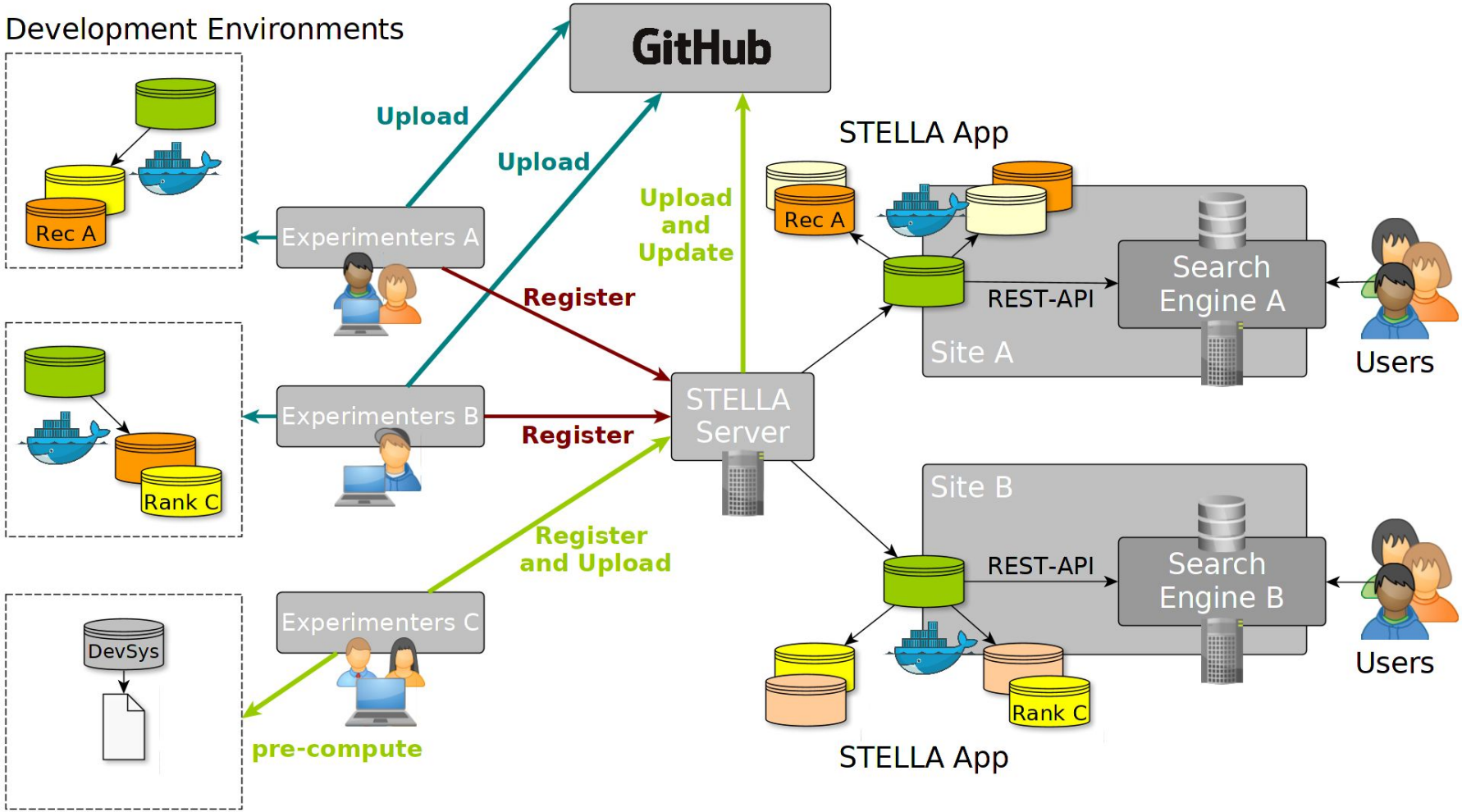
# STELLA complements existing infrastructures

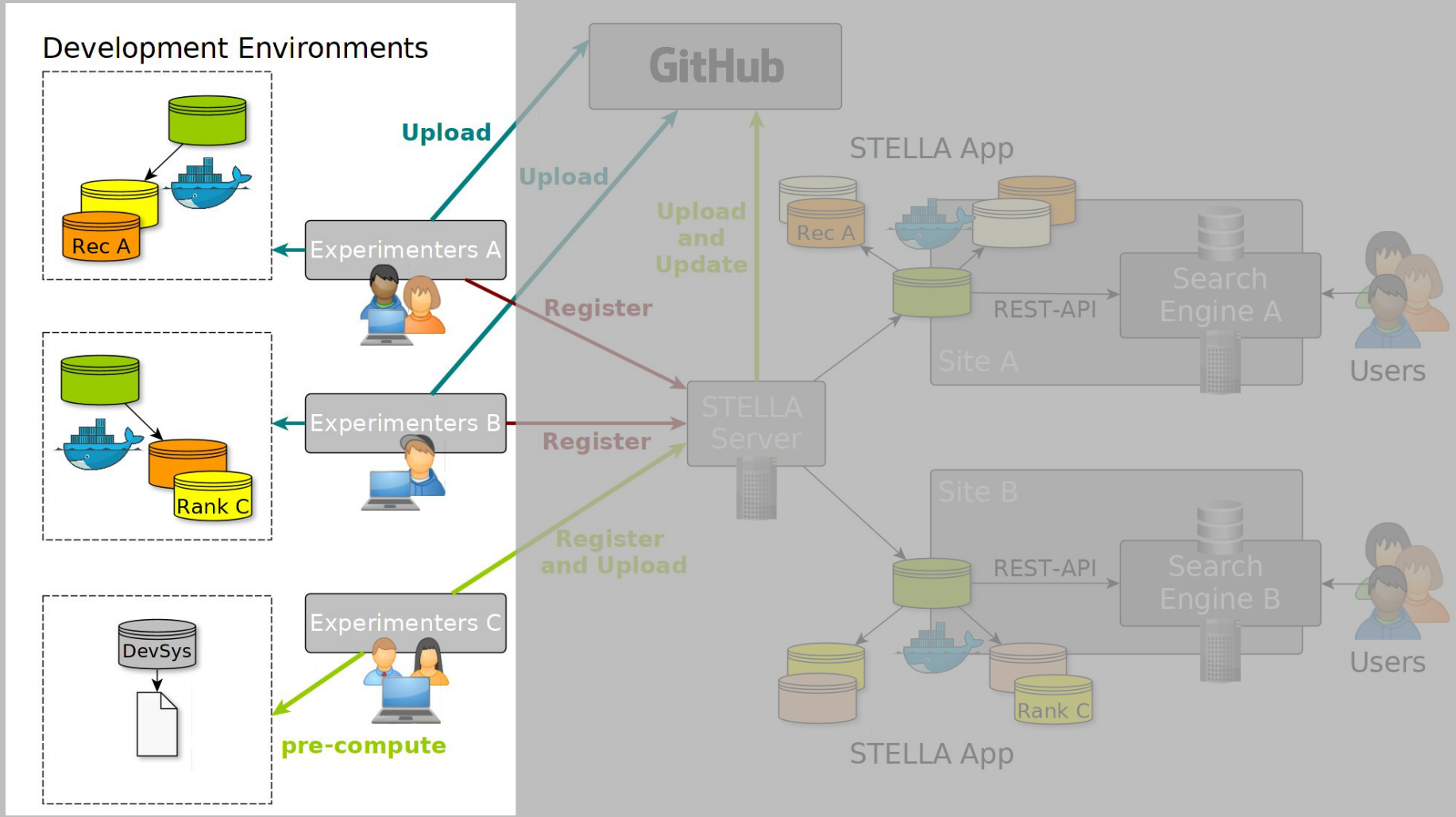
**Table 1** Evaluation infrastructures with respect to online/offline evaluations, domain specificity, and reproducibility

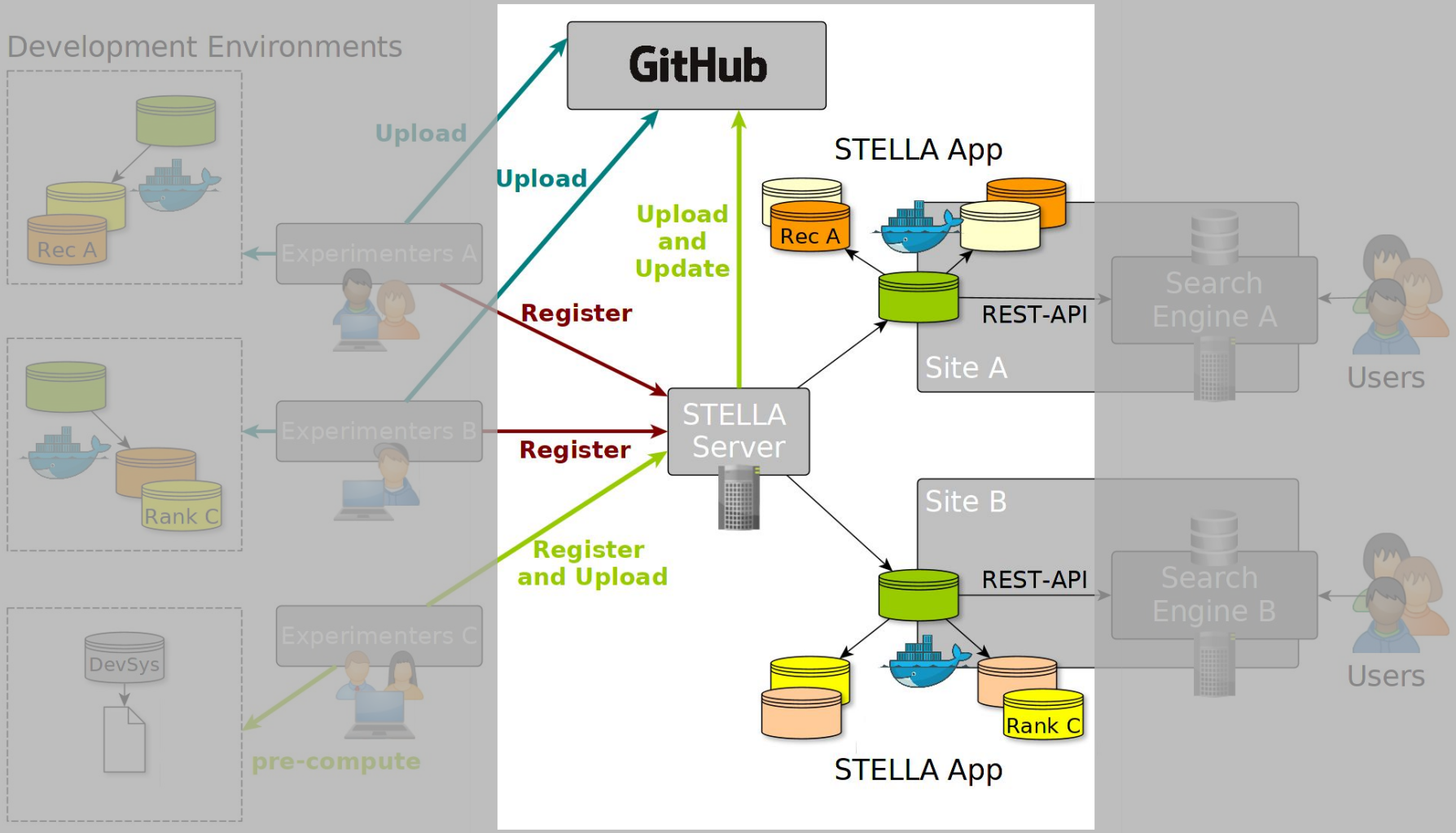
| Infrastructure | Description | Evaluation | | Domains | Reproducibility | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Online | Offline | | P | R | I | M | A | D |
| BioASQ | BioASQ is an initiative for semantic indexing and question answering [27] | ◐ | ● | Biomedical text indexation and Question Answering | ○ | ● | ○ | ● | ● | ● |
| TIRA | Underlying platform of the PAN lab dedicated to digital text forensics [23] | ○ | ● | Multi-domain, e.g., text forensics (author profiling), clickbait detection | ● | ● | ● | ● | ◐ | ● |
| LL4IR | Living labs infrastructure for ad-hoc search using interleaving [17] | ● | ○ | Multi-domain, e.g., online shop, academic repositories and search systems | ○ | ● | ○ | ● | ● | ● |
| NewsREEL | Living lab infrastructure to evaluate news article recommenders [15] | ◐ | ● | Commercial news articles recommendation | ◐ | ● | ○ | ◐ | ◐ | ● |

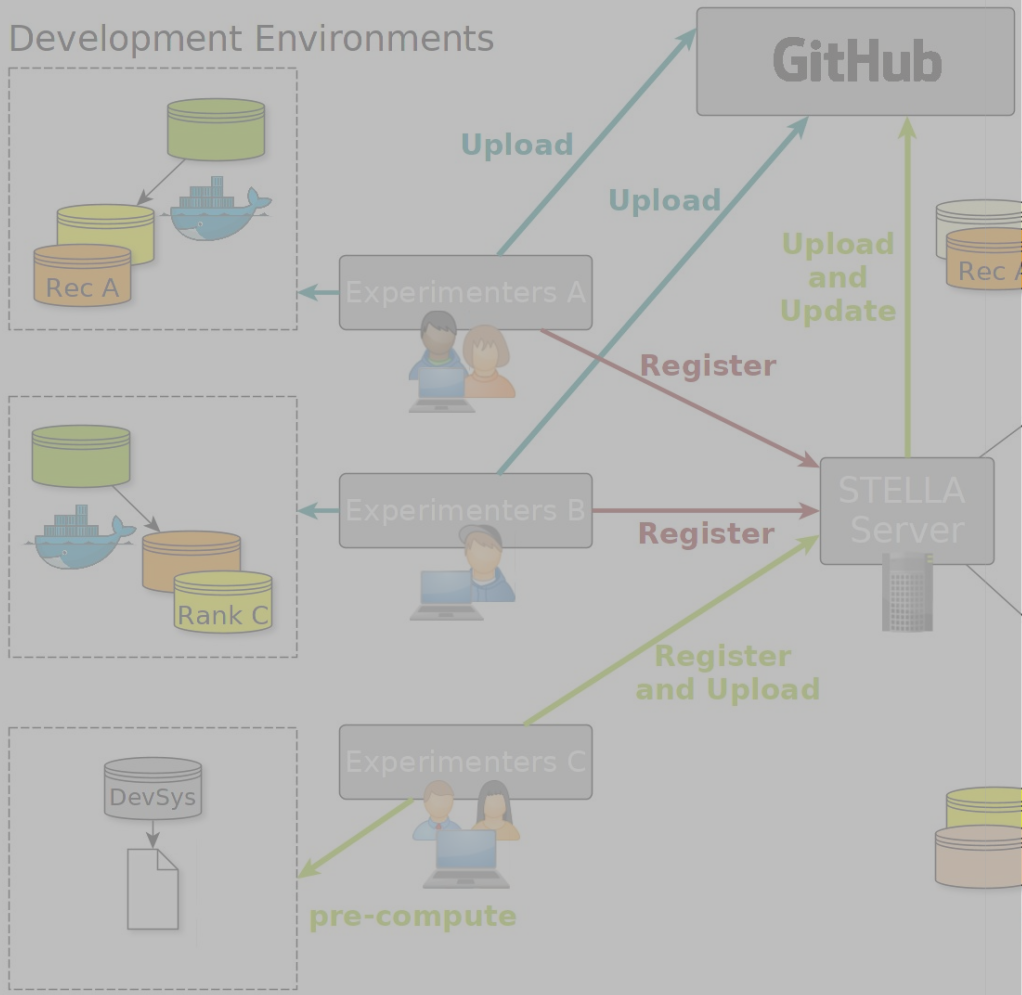
○ requirement not fulfilled, ◐ requirement partially fulfilled, ● requirement fulfilled

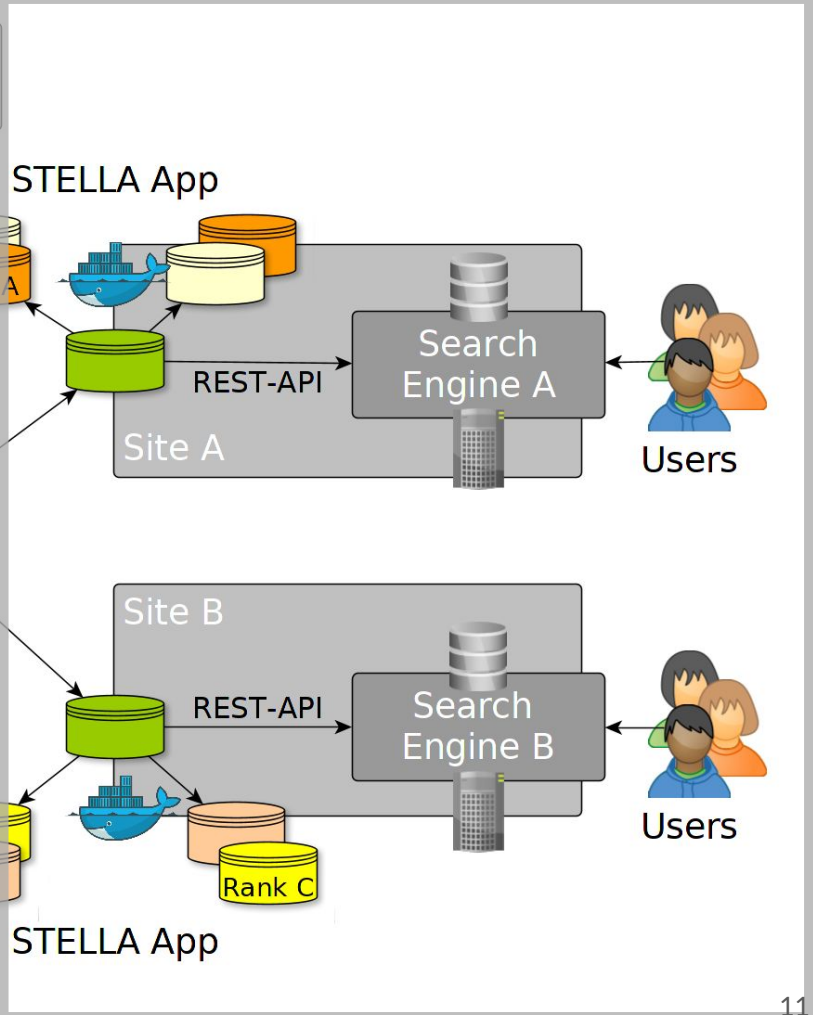
# The Big Picture


Development Environments
GitHub
Upload
Upload
Upload and Update
Register
Register
Register and Upload
pre-compute
Experimenters A
Experimenters B
Experimenters C
Rec A
Rank C
DevSys
STELLA Server
STELLA App
Rec A
REST-API
Search Engine A
Site A
Users
Site B
REST-API
Search Engine B
Rank C
Users
STELLA App


Development Environments
Rec A
Rank C
DevSys
Experimenters A
Experimenters B
Experimenters C
Upload
Upload
Register
Register
Upload and Update
Register and Upload
pre-compute
GitHub
STELLA Server
STELLA App
Rec A
REST-API
Search Engine A
Site A
Users
Site B
REST-API
Search Engine B
Rank C
Users
STELLA App


Development Environments
Rec A
Rank C
DevSys
Experimenters A
Experimenters B
Experimenters C
Upload
Upload
Upload and Update
Register
Register
Register and Upload
pre-compute
GitHub
STELLA Server
STELLA App
Rec A
REST-API
Site A
Search Engine A
Users
Site B
REST-API
Rank C
Search Engine B
Users
STELLA App


Development Environments
GitHub
Upload
Upload
Upload and Update
Register
Register
Register and Upload
pre-compute
Rec A
Rank C
DevSys
Experimenters A
Experimenters B
Experimenters C
STELLA Server

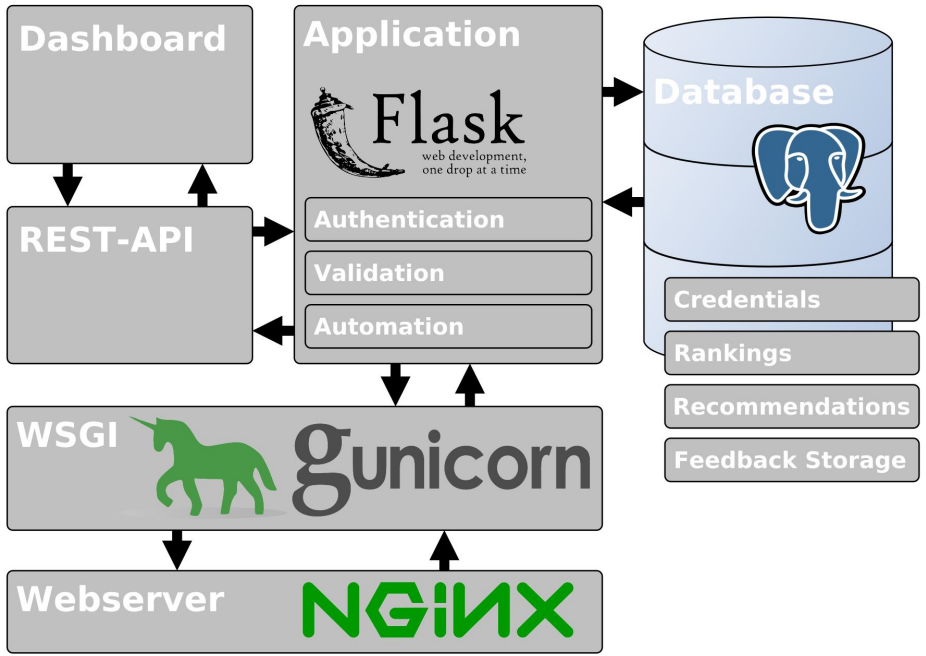

STELLA App
Rec A
REST-API
Search Engine A
Site A
Users
Site B
REST-API
Search Engine B
Users
Rank C
STELLA App

# STELLA Server

# STELLA Server

Functionalities:

- User administration
- Dashboard service
- Updating STELLA app → docker-compose.yml
- Data storage (user feedback)
  for data analysis, training, etc.

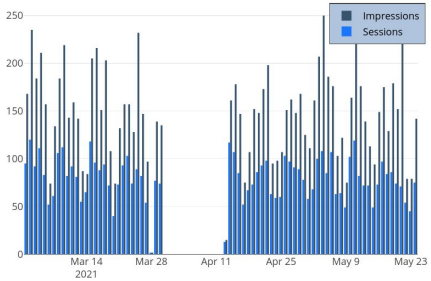


# Dashboard

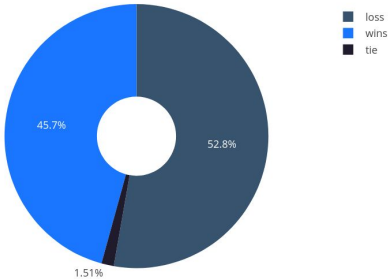
gesis_rec_pyserini@GESIS

Show results

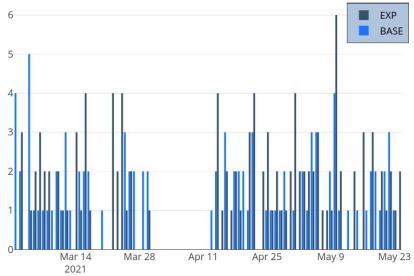
Sessions & Impressions




Wins, Losses and Ties




Clicks




| Metric | Value | Explanation |
|---|---|---|
| Win | 91 | A system 'wins' if it has more clicks on results assigned to it by the interleaving than clicks on results by the baseline system. |
| Loss | 105 | Opposite of 'Win'. Number of times when the system has less clicks on results than the baseline system. |
| Tie | 3 | Equal number of clicks for your system and the baseline. Only results having at least two clicks are included. |
| Outcome | 0.4643 | #Wins / (#Wins + #Loss) |
| Sessions | 5723 | Total number of sessions for which your system was used. |
| Impressions | 10482 | Total number of results for which your system was used. |
| Clicks | 94 | Total number of clicks your system received. |
| CTR | 0.009 | Click-through rate |

# Dashboard

| Status | Name | Submission date | Site | Task | Type | Repository | Activate | Deactivate | Delete | Feedback Data |
|---|---|---|---|---|---|---|---|---|---|---|
| running | gesis rec pyserini | 2019-06-10 | GESIS | Recommendation | Docker Container | | | | | |
| running | gesis rec pyterrier | 2019-06-10 | GESIS | Recommendation | Docker Container | | | | | |
| running | livivo base | 2019-06-10 | LIVIVO | Ranking | Docker Container | | | | | |
| submitted | livivo rank pyterrier | 2019-06-10 | LIVIVO | Ranking | Docker Container | | | | | |
| running | livivo rank pyserini | 2019-06-10 | LIVIVO | Ranking | Docker Container | | | | | |
| submitted | gesis rec precom | 2019-06-10 | GESIS | Recommendation | Pre-computed Run | | | | | |
| running | tekma n | 2021-04-12 | GESIS | Recommendation | Pre-computed Run | | | | | |
| running | lemuren elastic only | 2021-04-15 | LIVIVO | Ranking | Docker Container | | | | | |

# Read and write data

**POST** stella/api/v1/sessions/<int:id>/feedbacks

**GET** stella/api/v1/feedbacks/<int:id>

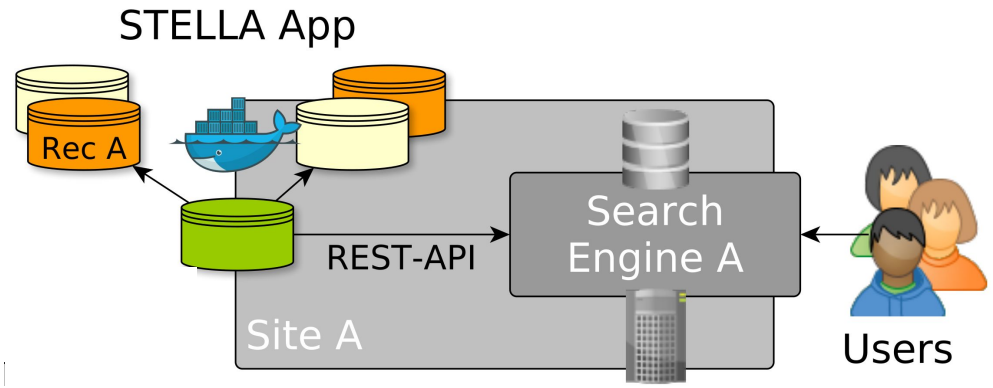
https://github.com/stella-project/stella-server


```
{
  "start": "2019-11-04 00:06:23",
  "end": "2019-11-04 00:10:38",
  "interleave": "True",
  "clicks": [
        {"1": {"docid": "doc1", "clicked": "False", "date": "None", "system": "EXP"}},
        {"2": {"docid": "doc11", "clicked": "True", "date": "2019-11-04 00:08:15", "system": "BASE"}},
        {"3": {"docid": "doc2", "clicked": "False", "date": "None", "system": "EXP"}},
        {"4": {"docid": "doc12", "clicked": "True", "date": "2019-11-04 00:06:23", "system": "BASE"}},
        {"5": {"docid": "doc3", "clicked": "False", "date": "None", "system": "EXP"}},
        {"6": {"docid": "doc13", "clicked": "False", "date": "None", "system": "BASE"}},
        {"7": {"docid": "doc4", "clicked": "False", "date": "None", "system": "EXP"}},
        {"8": {"docid": "doc14", "clicked": "False", "date": "None", "system": "BASE"}},
        {"9": {"docid": "doc5", "clicked": "False", "date": "None", "system": "EXP"}},
        {"10": {"docid": "doc15", "clicked": "False", "date": "None", "system": "BASE"}}
       ]
}
```

# STELLA App

# STELLA App

- “Broker” between Sites (GESIS, LIVIVO) and STELLA infrastructure
- Every site will deploy one instance of the STELLA app
- Multi-container application with all experimental systems



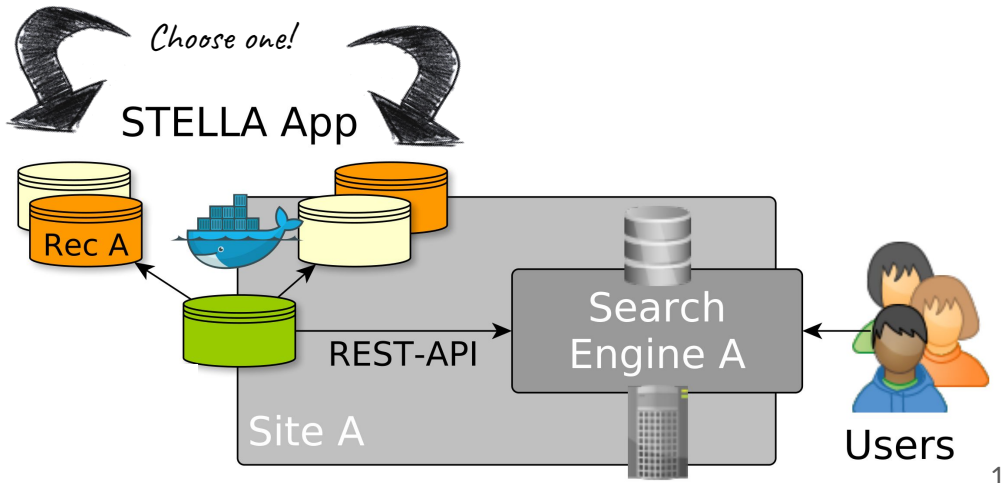
# GET ranking

**GET** /stella/api/v1/ranking?query=<string:query>&page=<int:page>&rpp=<int:rpp>&sid=<int:sid>&container=<string:container>

**Example**

**GET** /stella/api/v1/ranking?query=vaccine&page=0&rpp=10


```
{'body': {'1': {'docid': 'M27622217', 'type': 'BASE'},
          '2': {'docid': 'M27251231', 'type': 'EXP'},
          '3': {'docid': 'M27692969', 'type': 'BASE'},
          '4': {'docid': 'M26350569', 'type': 'EXP'},
          '5': {'docid': 'M26715777', 'type': 'EXP'},
          '6': {'docid': 'M26650940', 'type': 'BASE'},
          '7': {'docid': 'M27098271', 'type': 'EXP'},
          '8': {'docid': 'M28381438', 'type': 'BASE'},
          '9': {'docid': 'M27763523', 'type': 'EXP'},
          '10': {'docid': 'M27157745', 'type': 'BASE'},
 'header': {'container': {'base': 'rank_elastic_base', 'exp': 'rank_elastic'},
            'page': 0,
            'q': 'vaccine',
            'rid': 3,
            'rpp': 20,
            'sid': 1}}
```

# POST feedback

**POST** /stella/api/v1/ranking/<int:rid>/feedback

**Example**
**POST** /stella/api/v1/ranking/3/feedback


```
{'clicks': {'1': {'clicked': False,
                  'date': None,
                  'docid': 'M26923455',
                  'type': 'EXP'},
            '2': {'clicked': False,
                  'date': None,
                  'docid': 'M25600519',
                  'type': 'EXP'},
            '3': {'clicked': True,
                  'date': '2020-07-29 16:06:51',
                  'docid': 'M27515393',
                  'type': 'EXP'}},
 'end': '2020-07-29 16:12:53',
 'interleave': True,
 'start': '2020-07-29 16:06:51'}
```


- Sites run STELLA apps in their backends
- User interactions (implicit, explicit) are logged and will be written to the STELLA app
- STELLA app temporarily stores feedback and sends it to the STELLA server

https://github.com/stella-project/stella-app

# Participation

# Two types of tasks

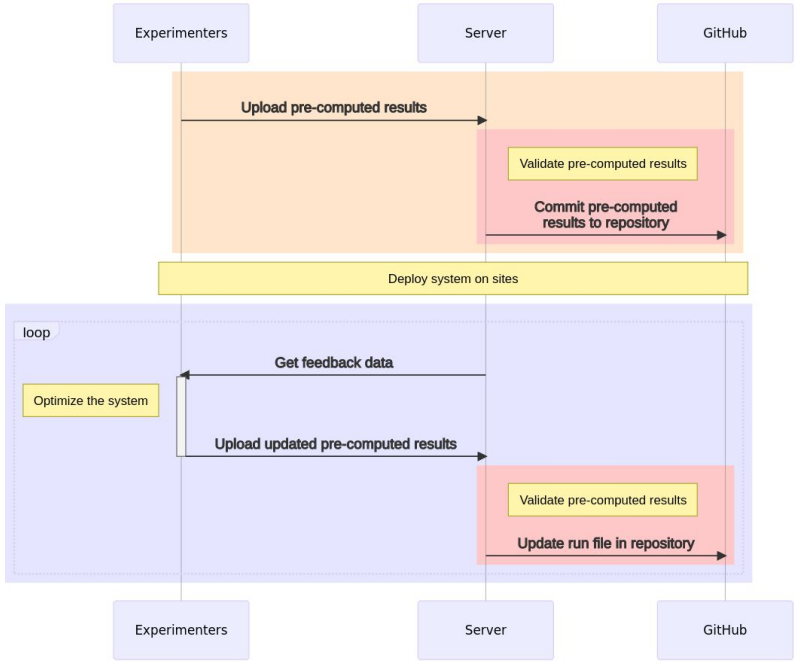
- **Task 1: Ad-hoc Search Ranking**
  - Given a query, find the most relevant documents

- **Task 2: Research Data Recommendation**
  - Given a *seed* publication, recommend research datasets


# Two types of submissions

**Type A: Pre-computed runs**

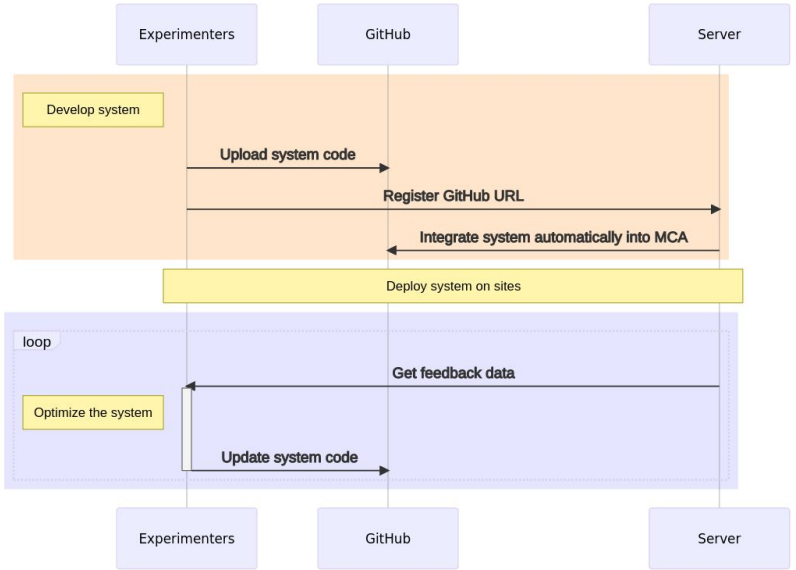
- Data provided by GESIS and LIVIVO contains
  **1. collection of documents/research datasets**
  **2. head (top-k) queries/documents**
  **3. candidate documents/research datasets**
- Submissions follow TREC run file syntax:
  `<topic> Q0 <docid> <rank> <score> <team>`
- Participants only upload the run files
- Comparable to LL4IR/TREC OpenSearch



# Two types of submissions

**Type B: Docker containers**

- Participants package their ranking and recommendation systems with the help of Docker
- Participants adapt provided Docker project templates to their requirements
- Uploaded Docker images will be deployed in the backend of the sites
- Live systems with dynamic rankings and recommendations beyond head queries/documents



# REST-API

- Participants contributing **Type B** submissions have to adapt **Dockerfiles** and implement RESTful endpoints (with the **Flask** framework)
- Ranking:

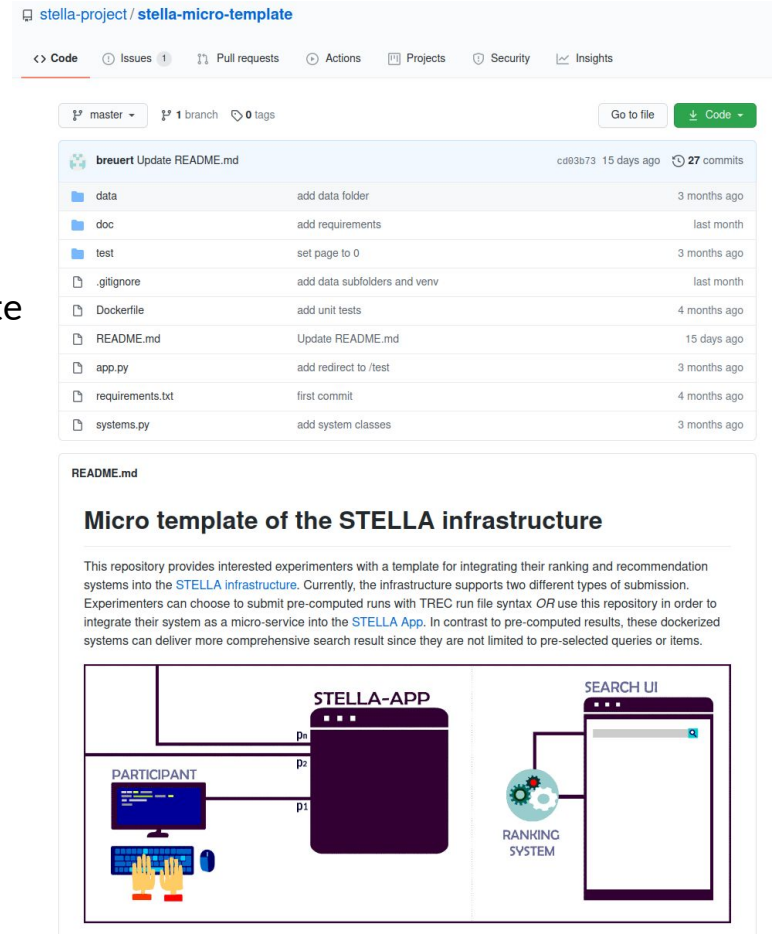
  **GET** container_name/ranking?query=<string:qstr>&page=<int:pnum>&rpp=<int:rppnum>

- Recommendation:

  **GET** container_name/recommendation/datasets?itemid=<string:itemidstr>&page=<int:pnum>&rpp=<int:rppnum>
  **GET** container_name/recommendation/publications?itemid=<string:itemidstr>&page=<int:pnum>&rpp=<int:rppnum>

# Project template

https://github.com/stella-project/stella-micro-template



# Final remarks

# Lessons learned

- **Docker pays off**: technically reproducible systems…
- … but **additional workload** for participants (and organizers).
- **Lightweight** infrastructure: requirements depend on user traffic.
- Experimental systems should **suit hardware resources** of the sites.

# Future work

- Improve **reproducibility**: technically and experimentally
- Shared **index** (cf. CIFF)
- More sophisticated **logging** functionality (cf. LogUI)
- **Scripting experiments** with PlanOut (cf. Apone)
- Gamification with public **leaderboard** (cf. MS MARCO)

# Thank you for your attention!

https://github.com/stella-project

# Experimental Evaluations & Results

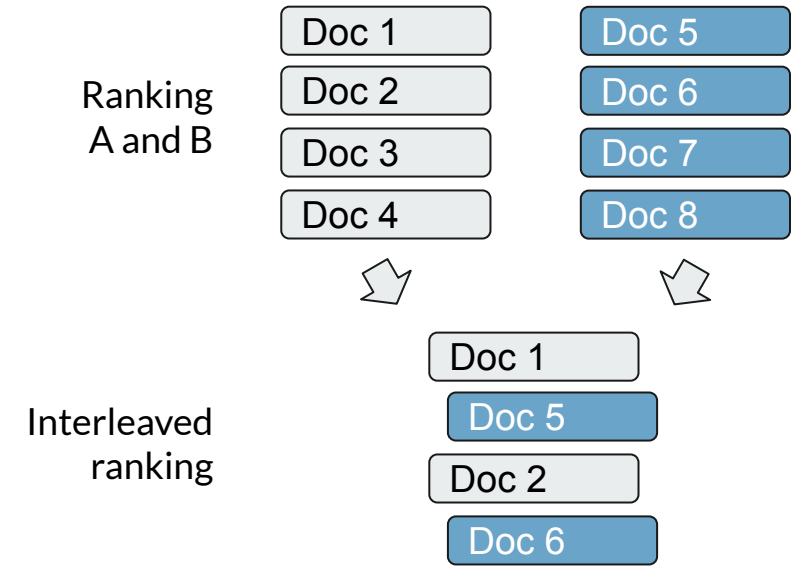
Philipp Schaer, Timo Breuer, Leyla Jael Castro, Benjamin Wolff, Johann Schaible and Narges Tavakolpoursaleh

Living Labs for Academic Search (LiLAS)


CLEF 2021; September, 21-24, 2021; online event (from Bucharest - Romania).

# Evaluation Setup & Metrics

# Team Draft Interleaving (TDI) [Radlinski et al., CIKM, 2008]




**Algorithm 2** Team-Draft Interleaving

**Input**: Rankings $A = (a_1, a_2, \dots)$ and $B = (b_1, b_2, \dots)$
**Init**: $I \leftarrow ()$; $TeamA \leftarrow \emptyset$; $TeamB \leftarrow \emptyset$;
**while** $(\exists i : A[i] \notin I) \wedge (\exists j : B[j] \notin I)$ **do**
  **if** $(|TeamA| < |TeamB|) \vee$
    $((|TeamA| = |TeamB|) \wedge (RandBit() = 1))$ **then**
    $k \leftarrow \min_i\{i : A[i] \notin I\}$ ..... *top result in A not yet in I*
    $I \leftarrow I + A[k]$; ..... *append it to I*
    $TeamA \leftarrow TeamA \cup \{A[k]\}$ ..... *clicks credited to A*
  **else**
    $k \leftarrow \min_i\{i : B[i] \notin I\}$ ..... *top result in B not yet in I*
    $I \leftarrow I + B[k]$ ..... *append it to I*
    $TeamB \leftarrow TeamB \cup \{B[k]\}$ ..... *clicks credited to B*
  **end if**
**end while**
**Output**: Interleaved ranking $I$, $TeamA$, $TeamB$

[Radlinski et al., CIKM, 2008]

# Wins, Losses, and Ties [Schuth et al., CLEF, 2015]

- **Wins**
  A system 'wins' if it has more clicks on results assigned to it by the interleaving than clicks on results by the baseline system.

- **Loss**
  Opposite of 'Win'. Number of times when the system has less clicks on results than the baseline system.

- **Tie**
  Equal number of clicks for your system and the baseline. Only results having at least two clicks are included.

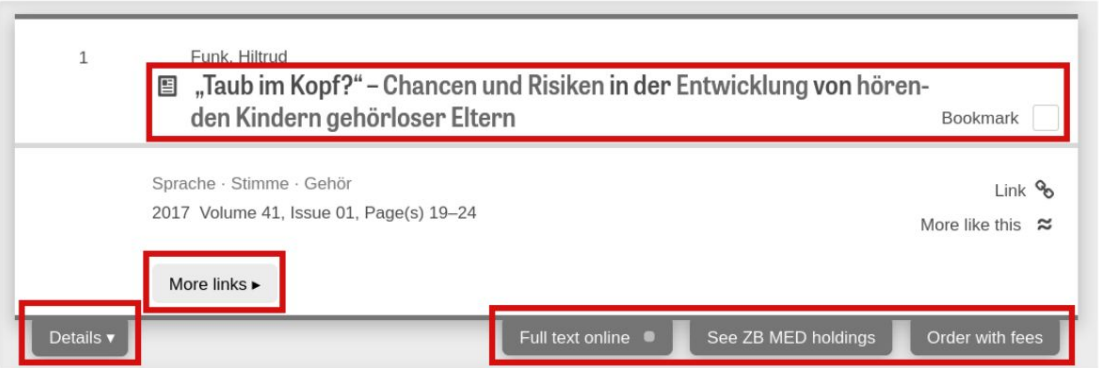
- **Outcome**
  #Wins / (#Wins + #Loss)

# (Normalized) Reward [Gingstad et al., CIKM, 2020]

$$Reward = \sum_{s \epsilon S} w_s c_s$$

$$nReward = \frac{Reward_{\text{exp}}}{Reward_{\text{exp}} + Reward_{\text{base}}}$$

# Logged SERP elements at LIVIVO




| SERP Element | $w_s$ |
|---|---|
| Bookmark | 10 |
| Order | 10 |
| Fulltext | 8 |
| In Stock | 8 |
| More Links | 2 |
| Title | 1 |
| Details | 1 |

# Round 1 & 2 Overview

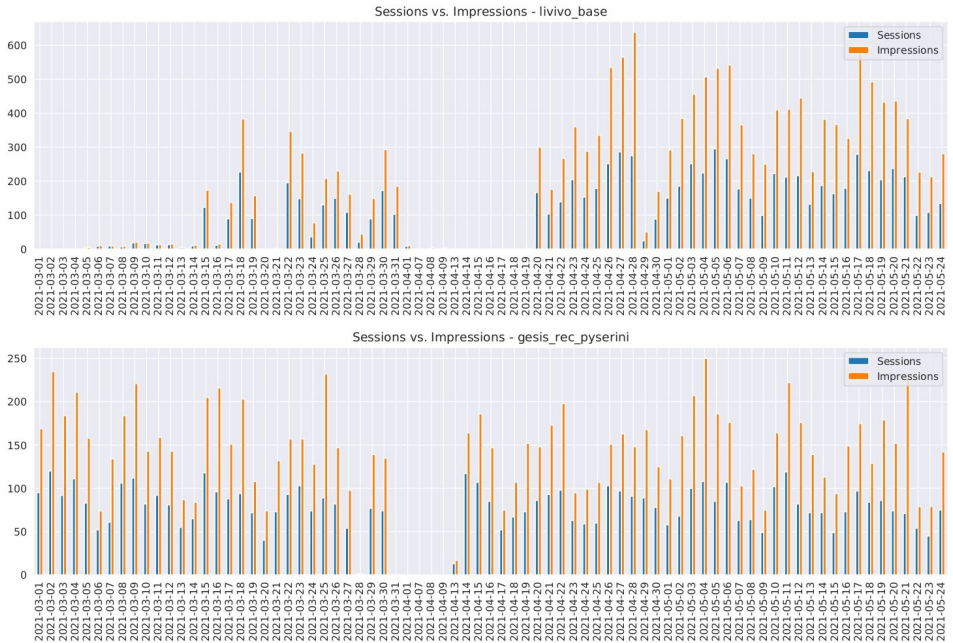
# System overview

| System name | Task | Type | Experimental | Round 1 | Round 2 |
|---|---|---|---|---|---|
| lemuren_elk | 1 | A | ● | ● | ● |
| tekmas | 1 | A | ● | ● | ● |
| save_fami | 1 | A | ● | ● | ● |
| livivo_rank_pyserini | 1 | B | ● | ◑ | ◑ |
| lemuren_elastic_only | 1 | B | ● | ○ | ● |
| lemuren_elastic_preprocessing | 1 | B | ● | ○ | ● |
| livivo_base | 1 | B | ○ | ● | ● |
| tekma_n | 2 | A | ● | ○ | ● |
| gesis_rec_precom | 2 | A | ● | ● | ○ |
| gesis_rec_pyterrier | 2 | B | ● | ● | ● |
| gesis_rec_pyserini | 2 | B | ○ | ● | ● |

## Number of sessions, impressions, clicks and click through rate (CTR)

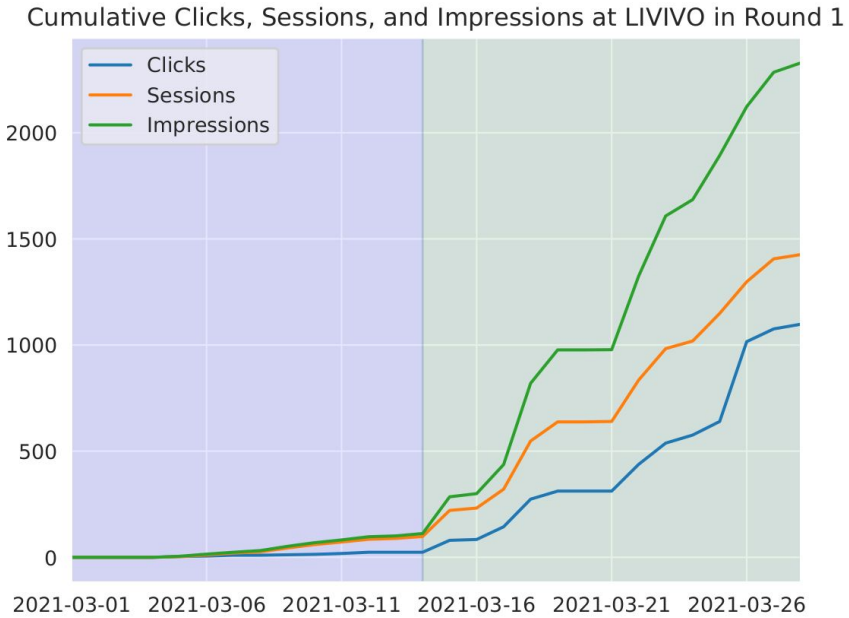
| Evaluation round | Site | Sessions | Impressions | Clicks | CTR |
|---|---|---|---|---|---|
| Round 1 | LIVIVO | 2852 | 4658 | 2452 | 0.5264 |
| Round 1 | GESIS | 4568 | 8390 | 152 | 0.0181 |
| Round 2 | LIVIVO | 12962 | 25830 | 11562 | 0.4476 |
| Round 2 | GESIS | 6576 | 12068 | 250 | 0.0207 |

# Sessions & Impressions distributions

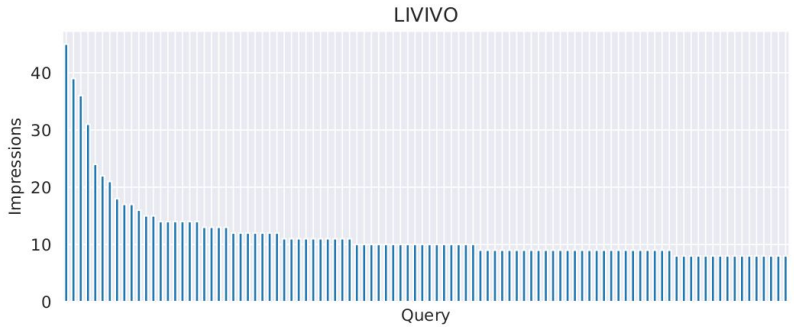


# Experimental Results

# Type A vs. B Submissions



# Top queries at LIVIVO

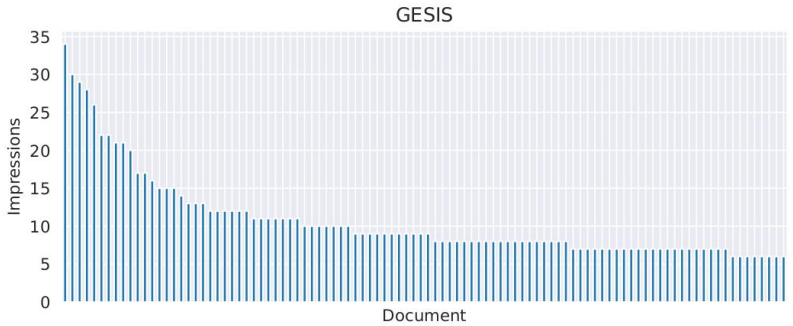



| Rank | Query string | Impressions |
|---|---|---|
| 1 | covid19 | 45 |
| 2 | demenz | 39 |
| 3 | guillian barre syndrome | 36 |
| 4 | polyvinyl and nasal and packing | 31 |
| 5 | covid | 24 |
| 6 | pflege | 22 |
| 7 | cancer | 21 |
| 8 | parkinson | 18 |
| 9 | depression | 17 |
| 10 | schlaganfall | 17 |

# Queries at LIVIVO

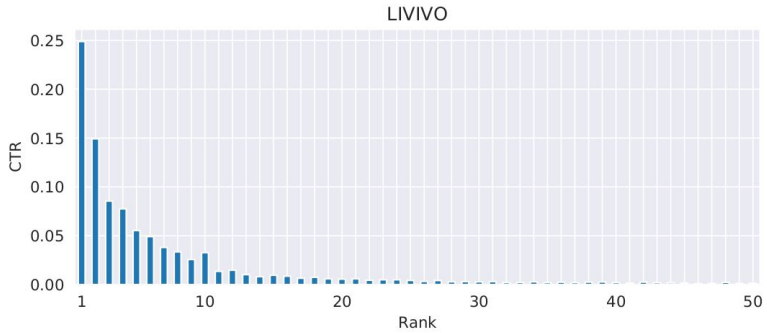
| | |
|---|---|
| Number of Unique Queries | 11822 |
| Average Query Length [Terms] | 2.9840 |
| Average Number of Queries per Session | 1.9340 |
| Average Number of Clicks per Query | 0.4547 |

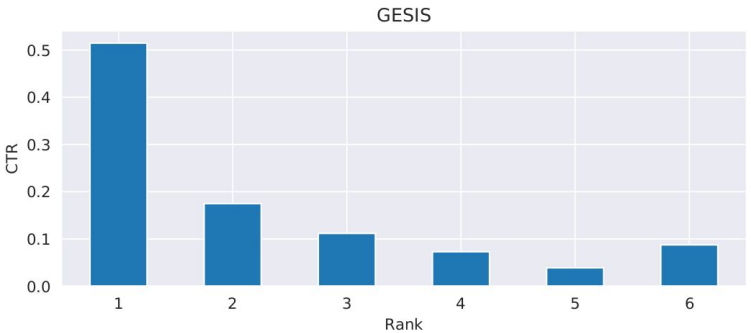
# Top documents at GESIS



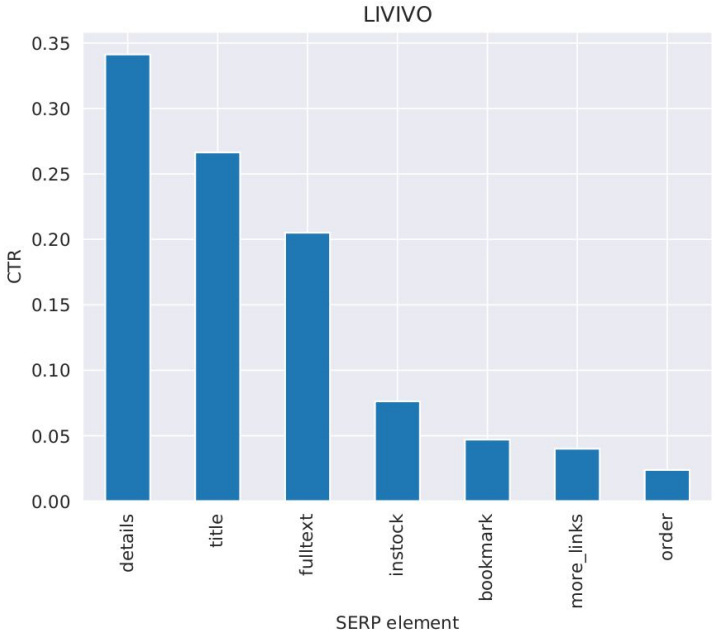

| Rank | Document title | Impressions |
|---|---|---|
| 1 | Die Nichtwähler : Politische Normalität oder wachsende Distanz zu den Parteien? | 34 |
| 2 | Doing Gender: Soziale Praktiken der Geschlechterunterscheidung | 30 |
| 3 | ZUMA-Informationssystem. Elektronisches Handbuch sozialwissenschaftlicher Erhebungsinstrumente | 29 |
| 4 | Situiertes Wissen : die Wissenschaftsfrage im Feminismus und das Privileg einer partialen Perspektive | 28 |
| 5 | Party identification, ideological preference, and the left-right dimension among western mass publics | 26 |
| 6 | Die soziale Konstruktion von Geschlecht : Erkenntnisperspektiven und gesellschaftstheoretische Fragen | 22 |
| 7 | Konsensfiktionen in Kleingruppen: dargestellt am Beispiel von jungen Ehen | 22 |
| 8 | SWLS Satisfaction with Life Scale | 21 |
| 9 | Entwicklung einer Skala zur Messung von Arbeitszufriedenheit (SAZ) | 21 |
| 10 | Gesundheitliche Ungleichheit / Health Inequalities | 20 |

# Click-through Rate (CTR) vs. Rank



# Click distribution on SERP elements at LIVIVO



# Round 1

| System | Win | Loss | Tie | Outcome | Sessions | Impressions | Clicks | CTR |
|---|---|---|---|---|---|---|---|---|
| gesis_rec_pyserini† | 36 | 36 | 1 | 0.50 | 2284 | 4195 | 37 | 0.0088 |
| gesis_rec_pyterrier | 26 | 28 | 1 | 0.48 | 1968 | 3675 | 28 | 0.0076 |
| gesis_rec_precom | 10 | 8 | 0 | 0.56 | 316 | 520 | 11 | 0.0212 |
| livivo_base† | 332 | 234 | 67 | 0.59 | 1426 | 2329 | 677 | 0.2907 |
| livivo_rank_pyserini | 215 | 302 | 64 | 0.42* | 1260 | 2135 | 517 | 0.2422 |
| lemuren_elk | 4 | 8 | 1 | 0.33 | 45 | 55 | 10 | 0.1818 |
| tekmas | 6 | 10 | 1 | 0.38 | 64 | 77 | 8 | 0.1039 |
| save_fami | 9 | 12 | 1 | 0.43 | 57 | 62 | 14 | 0.2258 |

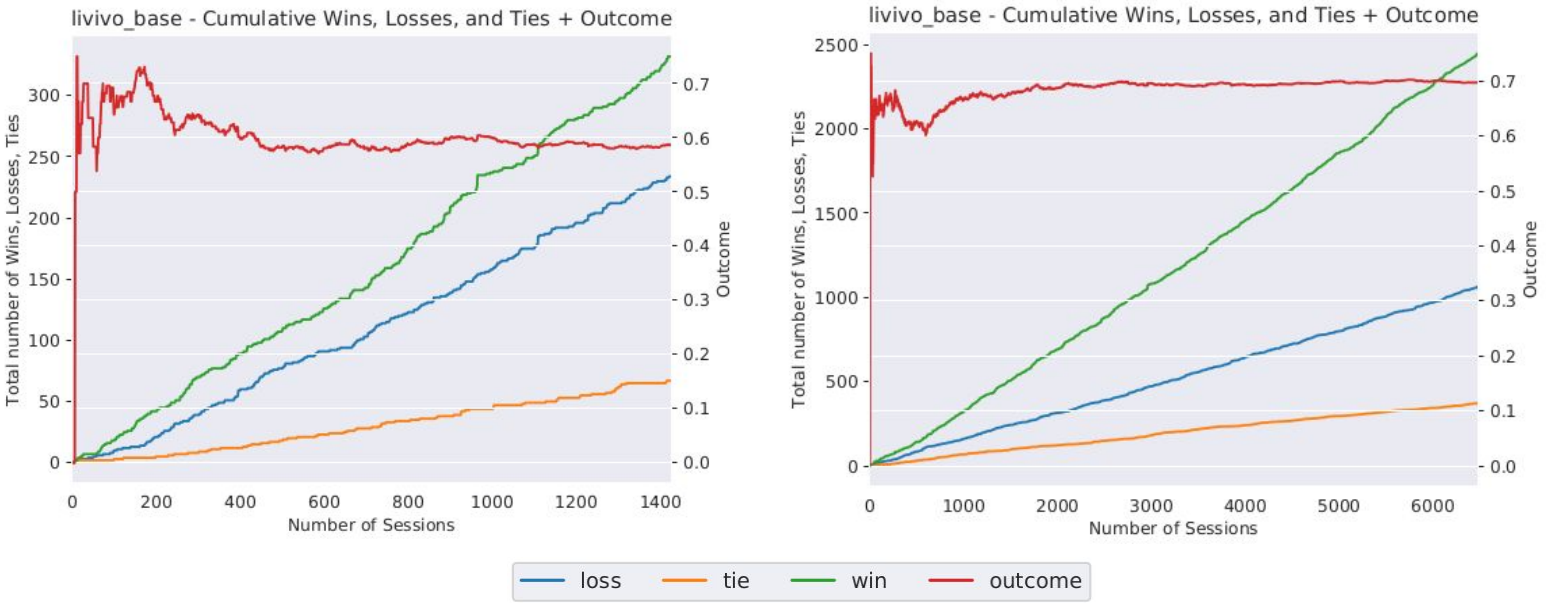
# Round 2 - 1

| System | Win | Loss | Tie | Outcome | Sessions | Impressions | Clicks | CTR |
|---|---|---|---|---|---|---|---|---|
| gesis_rec_pyserini† | 51 | 68 | 2 | 0.43 | 3288 | 6034 | 53 | 0.0088 |
| gesis_rec_pyterrier | 26 | 25 | 1 | 0.51 | 1529 | 2937 | 27 | 0.0092 |
| tekma_n | 42 | 26 | 1 | 0.62 | 1759 | 3097 | 45 | 0.0145 |
| livivo_base† | 2447 | 1063 | 372 | 0.70 | 6481 | 12915 | 3791 | 0.2935 |
| livivo_rank_pyserini | 48 | 71 | 15 | 0.40 | 243 | 434 | 112 | 0.2581 |
| lemuren_elastic_only | 707 | 1042 | 218 | 0.40* | 3131 | 6274 | 1273 | 0.2029 |
| lemuren_elastic_preprocessing | 291 | 1308 | 135 | 0.18* | 2948 | 6026 | 570 | 0.0946 |
| lemuren_elk | 6 | 13 | 0 | 0.32 | 61 | 69 | 10 | 0.1449 |
| tekma_s | 4 | 7 | 1 | 0.36 | 36 | 42 | 5 | 0.1190 |
| save_fami | 7 | 6 | 3 | 0.54 | 62 | 70 | 20 | 0.2857 |

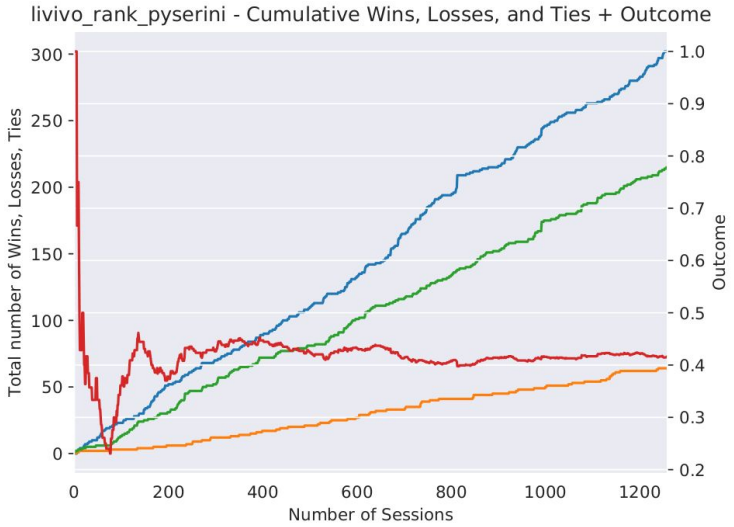
# Round 2 - 2

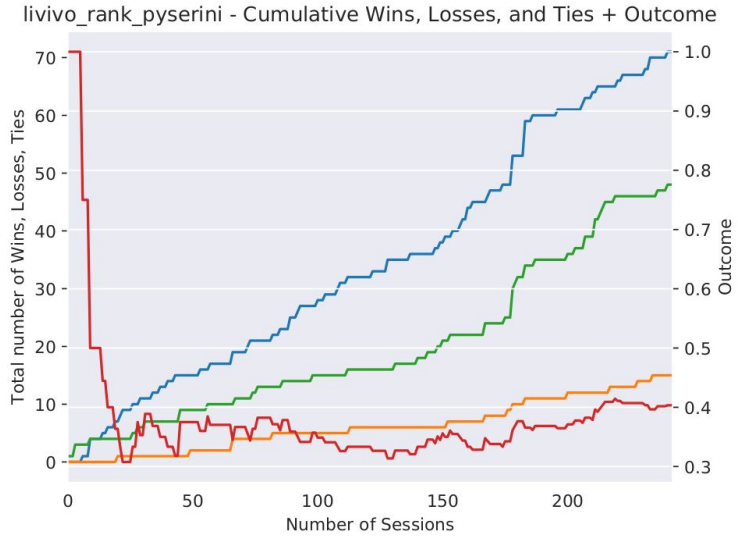
| | Bookmark | Details | Fulltext | In Stock | More Links | Order | Title | Total Clicks | nReward |
|---|---|---|---|---|---|---|---|---|---|
| livivo_rank_pyserini | 182 | 341 | 176 | 55 | 62 | 28 | 263 | 1107 | 0.4367 |
| livivo_base | 180 | 443 | 228 | 154 | 57 | 29 | 329 | 1420 | 0.5633 |
| lemuren_elastic_only | 63 | 832 | 481 | 107 | 105 | 54 | 638 | 2280 | 0.4045 |
| livivo_base | 56 | 1066 | 646 | 295 | 129 | 85 | 858 | 3135 | 0.5955 |
| lemuren_elastic_preprocessing | 23 | 355 | 257 | 23 | 28 | 21 | 285 | 992 | 0.2143 |
| livivo_base | 69 | 1190 | 762 | 301 | 119 | 82 | 934 | 3457 | 0.7857 |
| lemuren_elk | 1 | 13 | 16 | 0 | 2 | 0 | 10 | 42 | 0.4242 |
| livivo_base | 1 | 24 | 7 | 14 | 1 | 0 | 20 | 67 | 0.5758 |
| tekmas | 2 | 11 | 2 | 2 | 1 | 0 | 6 | 24 | 0.3430 |
| livivo_base | 0 | 13 | 6 | 7 | 0 | 1 | 9 | 36 | 0.6570 |
| save_fami | 11 | 21 | 9 | 3 | 1 | 1 | 16 | 62 | 0.5496 |
| livivo_base | 8 | 13 | 7 | 5 | 2 | 1 | 6 | 42 | 0.4504 |
| All experimental systems | 282 | 1573 | 941 | 190 | 199 | 104 | 1218 | 4507 | 0.3485 |
| livivo_base | 314 | 2749 | 1656 | 776 | 308 | 198 | 2156 | 8157 | 0.6515 |

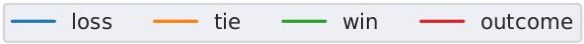
# Wins, Losses, Ties vs. Number of sessions



# Wins, Losses, Ties vs. Number of sessions



# Lessons learned

- **Docker pays off:** more feedback data as compared to precomputed results
- **Statistically significant results** with less online time
- **Power-law like distributions** of clicks and queries (complies with other studies)
- More in-depth comparison of systems with **weighted clicks and rewards**
- In the future: provide participants with **transparent baseline systems**

# Thank you for your attention!