InspireHEP

Search

A long history

Started as the Stanford Physics Information Retrieval System (SPIRES) in the 70s it was the main information platform for high energy physics.

Rewritten in 2010 with Invenio and became INSPIRE.

Currently a collaboration between CERN, DESY, Fermilab, IHEP, IN2P3 and SLAC.

What for?

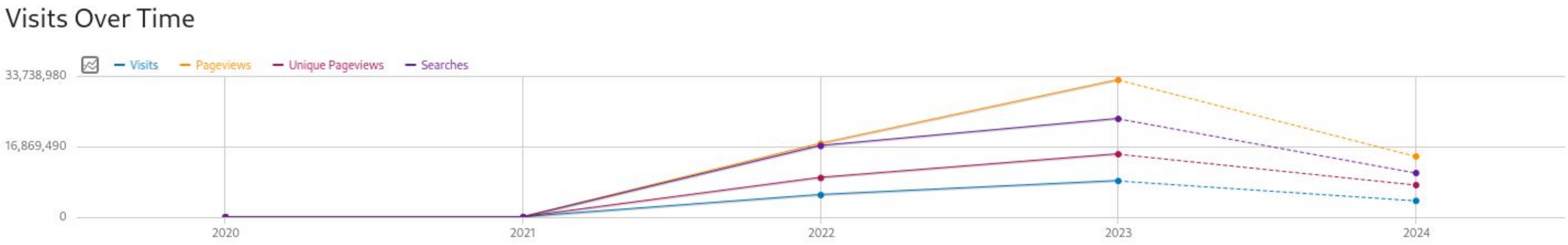
Searching for papers using complex multidimensional queries.

Review job applicants.

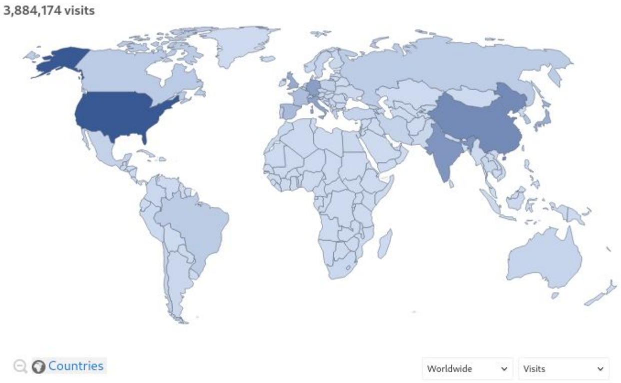
Publish or find job openings.

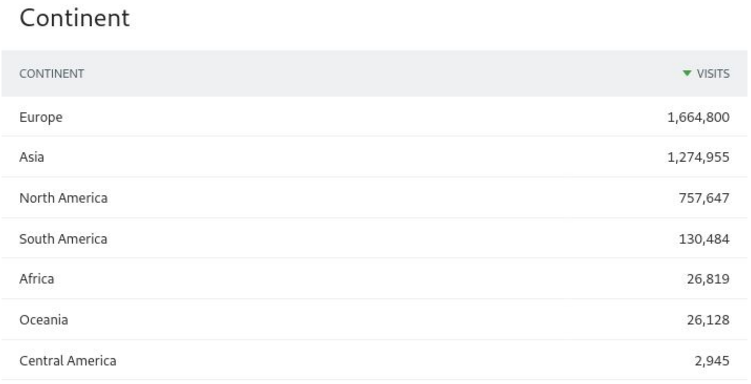
Find upcoming conferences and seminars.

Stats - traffic



Visits Over Time





Continent

CONTINENT	▼ VISITS
Europe	1,664,800
Asia	1,274,955
North America	757,647
South America	130,484
Africa	26,819
Oceania	26,128
Central America	2,945

Stats - indices

~2M records

~24M docs in the literature index

~165GB literature index primary size (x6 primaries)

~1B search per week (both end user and internal)

Two use cases for search

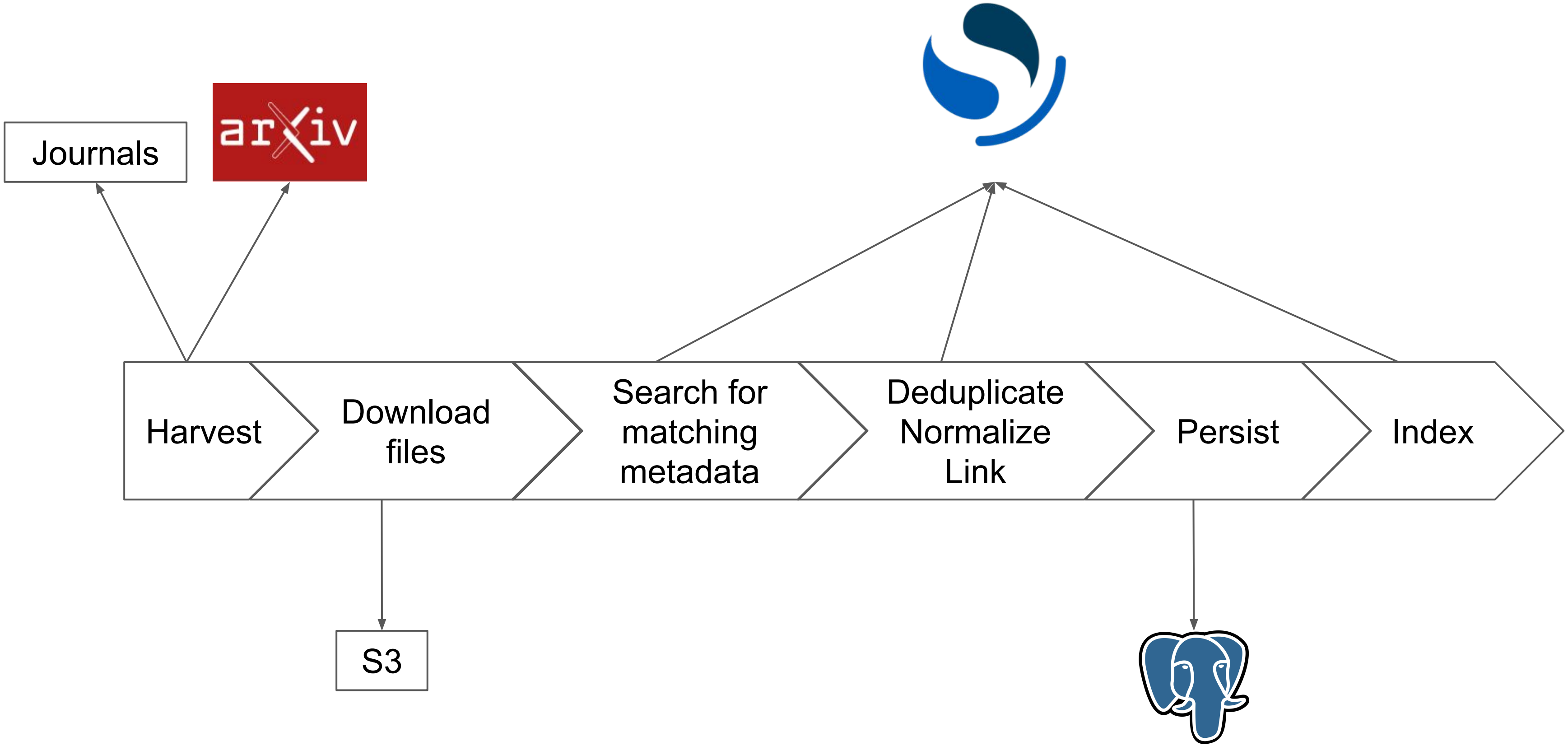
End users

- No scoring
- Several ordering options

Internal processing

- Disambiguation
- Reference matching
- Fuzzy matching and scoring

Harvesting



Journals

arxiv

Harvest

Download files

Search for matching metadata

Deduplicate Normalize Link

Persist

Index

S3

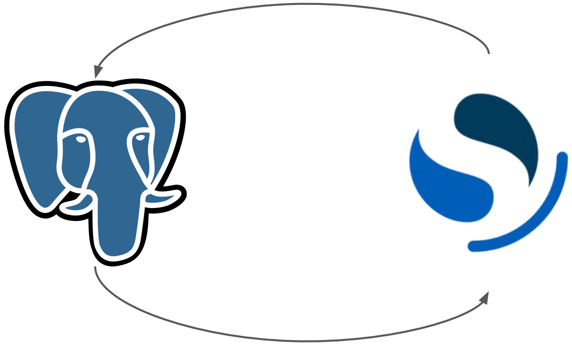
Live remapping

1) Spin-up new backend workers
2) Create new indices with new mappings
3) Start re-indexing all the records with the new workers on the new indices
4) Spin-up new frontend pod connected to the new indices
5) Remove the old frontend and backend pods
6) Re-index missing records

No downtime but some records might be temporarily missing.

Aliases can be used but are optional in our case.

The loop…



Not a recommended pattern

Possible with versioning:

- ORM automatically increment a version field in postgres during updates
- On OS docs can be updated with the same or higher version

Documents can be updated with the same version if the record didn’t change in Postgres.

Might happen with calculated fields.

When merging new and existing records versions are checked on both OS and Postgres.

Indexing

Bulk indexing API for performance

Different analyzers for different needs:

- ASCII folding analyzer
- ICU analyzer
- …

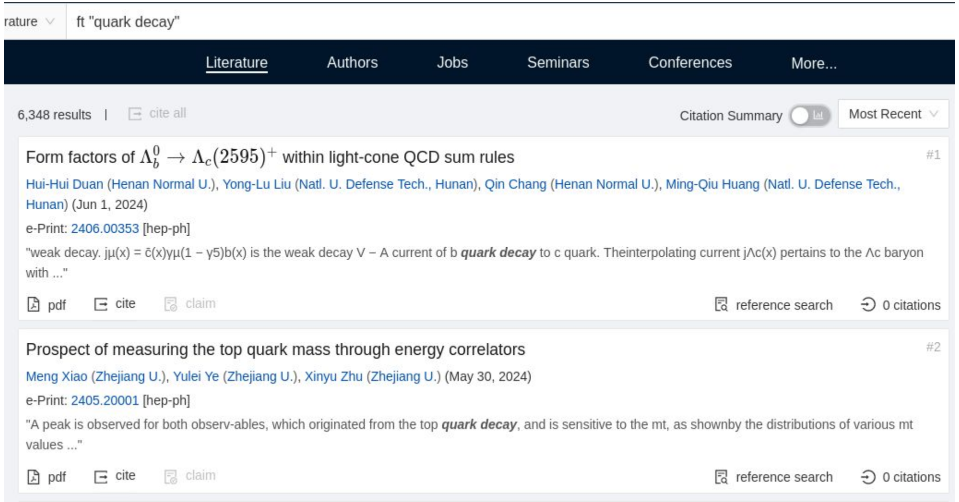
Documents are split by types only

- Literature
- Authors
- Jobs
- …

Lot of pre-processing in app before indexing !!!

(more on this later)

PDF full text search



6,348 results | cite all

Citation Summary

Most Recent

Form factors of $\Lambda_b^0 \to \Lambda_c(2595)^+$ within light-cone QCD sum rules #1

Hui-Hui Duan (Henan Normal U.), Yong-Lu Liu (Natl. U. Defense Tech., Hunan), Qin Chang (Henan Normal U.), Ming-Qiu Huang (Natl. U. Defense Tech., Hunan) (Jun 1, 2024)

e-Print: 2406.00353 [hep-ph]

"weak decay. jμ(x) = c̄(x)γμ(1 − γ5)b(x) is the weak decay V − A current of b ***quark decay*** to c quark. Theinterpolating current jΛc(x) pertains to the Λc baryon with ..."

Prospect of measuring the top quark mass through energy correlators #2

Meng Xiao (Zhejiang U.), Yulei Ye (Zhejiang U.), Xinyu Zhu (Zhejiang U.) (May 30, 2024)

e-Print: 2405.20001 [hep-ph]

"A peak is observed for both observ-ables, which originated from the top ***quark decay***, and is sensitive to the mt, as shownby the distributions of various mt values ..."

Ingest-attachment plugin

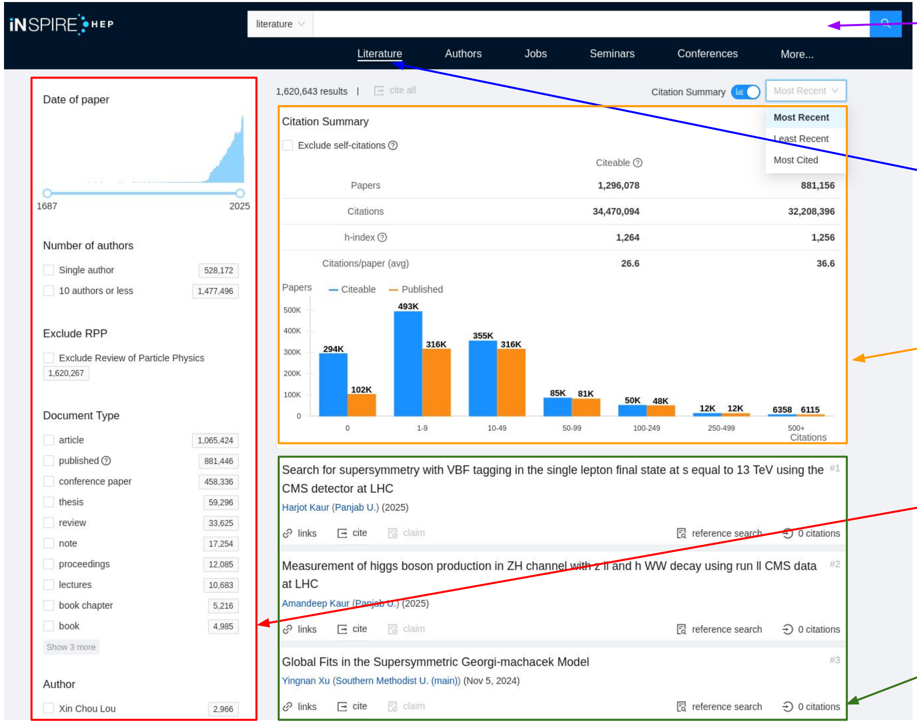
Index a document with a PDF attached.

The PDF is parsed and the content loaded into a field of the document.

For fulltext search

Not visible to the users (licensing)

How does it look?



Citation Summary

□ Exclude self-citations

	Citeable	
Papers	1,296,078	881,156
Citations	34,470,094	32,208,396
h-index	1,264	1,256
Citations/paper (avg)	26.6	36.6

Search for supersymmetry with VBF tagging in the single lepton final state at s equal to 13 TeV using the CMS detector at LHC #1

Harjot Kaur (Panjab U.) (2025)

links cite claim reference search 0 citations

Measurement of higgs boson production in ZH channel with z ll and h WW decay using run ll CMS data at LHC #2

Amandeep Kaur (Panjab U.) (2025)

links cite claim reference search 0 citations

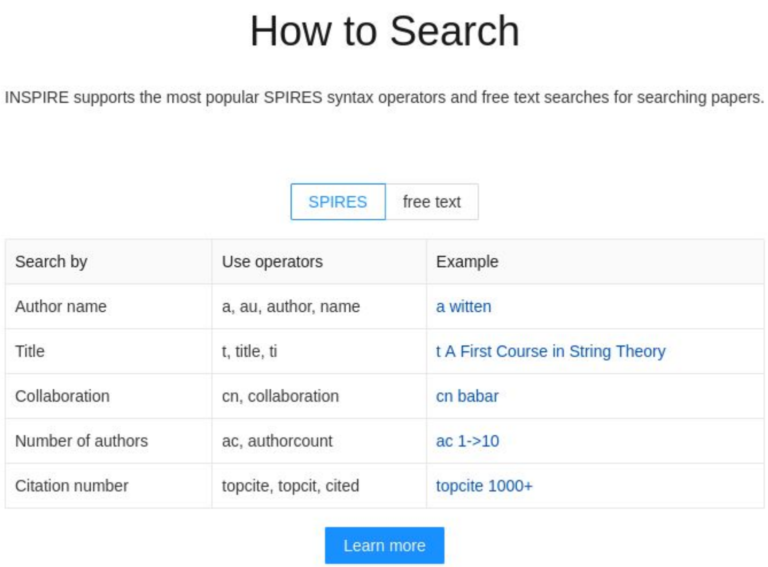
Global Fits in the Supersymmetric Georgi-machacek Model #3

Yingnan Xu (Southern Methodist U. (main)) (Nov 5, 2024)

links cite claim reference search 0 citations

Custom Search	Custom Query Parser
Collections	Indices
Summary	Custom Aggrs
Facets	Aggregations
Results	

Custom query parser



How to Search

INSPIRE supports the most popular SPIRES syntax operators and free text searches for searching papers.

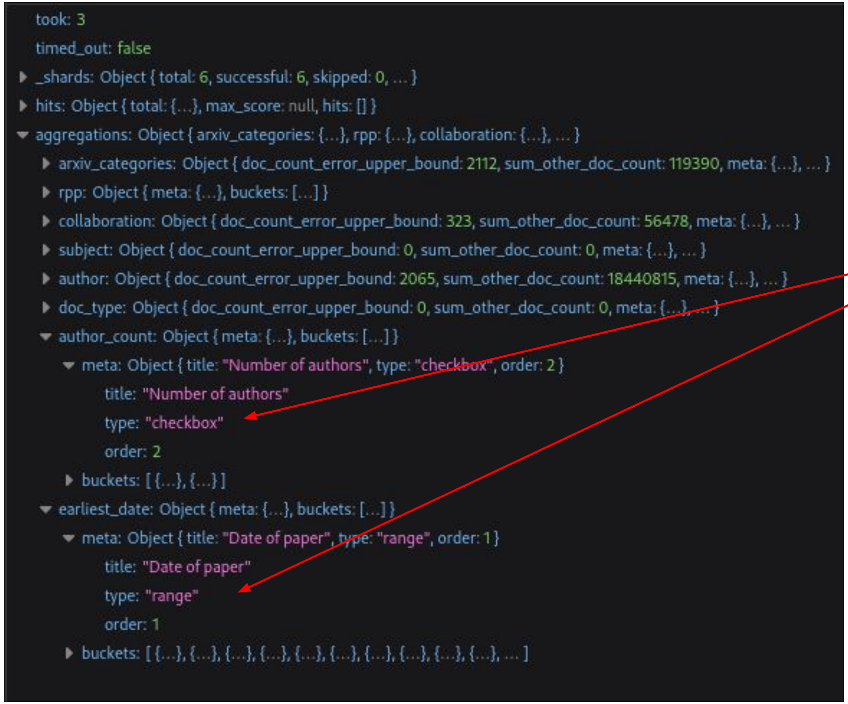
Search by	Use operators	Example
Author name	a, au, author, name	a witten
Title	t, title, ti	t A First Course in String Theory
Collaboration	cn, collaboration	cn babar
Number of authors	ac, authorcount	ac 1->10
Citation number	topcite, topcit, cited	topcite 1000+

Historical search format

Inherited from SPIRES

Still used by our community

Facets



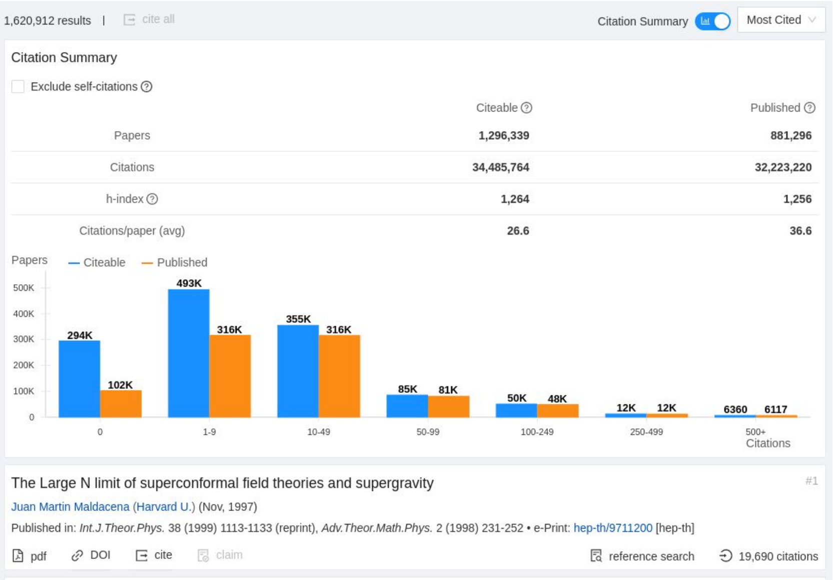
```
   took: 3
   timed_out: false
 ▶ _shards: Object { total: 6, successful: 6, skipped: 0, … }
 ▶ hits: Object { total: {…}, max_score: null, hits: [] }
 ▼ aggregations: Object { arxiv_categories: {…}, rpp: {…}, collaboration: {…}, … }
   ▶ arxiv_categories: Object { doc_count_error_upper_bound: 2112, sum_other_doc_count: 119390, meta: {…}, … }
   ▶ rpp: Object { meta: {…}, buckets: […] }
   ▶ collaboration: Object { doc_count_error_upper_bound: 323, sum_other_doc_count: 56478, meta: {…}, … }
   ▶ subject: Object { doc_count_error_upper_bound: 0, sum_other_doc_count: 0, meta: {…}, … }
   ▶ author: Object { doc_count_error_upper_bound: 2065, sum_other_doc_count: 18440815, meta: {…}, … }
   ▶ doc_type: Object { doc_count_error_upper_bound: 0, sum_other_doc_count: 0, meta: {…}, … }
   ▼ author_count: Object { meta: {…}, buckets: […] }
     ▼ meta: Object { title: "Number of authors", type: "checkbox", order: 2 }
         title: "Number of authors"
         type: "checkbox"
         order: 2
     ▶ buckets: [ {…}, {…} ]
   ▼ earliest_date: Object { meta: {…}, buckets: […] }
     ▼ meta: Object { title: "Date of paper", type: "range", order: 1 }
         title: "Date of paper"
         type: "range"
         order: 1
     ▶ buckets: [ {…}, {…}, {…}, {…}, {…}, {…}, {…}, {…}, {…}, {…}, … ]
```

Out of the box, standard aggregations

Search and aggregation query sent in parallel

Aggregation metadata attached to the response to simplify the presentation on the UI

Citation Summary



1,620,912 results | cite all

Citation Summary | Most Cited

Citation Summary

□ Exclude self-citations

	Citeable	Published
Papers	1,296,339	881,296
Citations	34,485,764	32,223,220
h-index	1,264	1,256
Citations/paper (avg)	26.6	36.6

The Large N limit of superconformal field theories and supergravity #1

Juan Martin Maldacena (Harvard U.) (Nov, 1997)

Published in: *Int.J.Theor.Phys.* 38 (1999) 1113-1133 (reprint), *Adv.Theor.Math.Phys.* 2 (1998) 231-252 • e-Print: hep-th/9711200 [hep-th]

pdf DOI cite claim reference search 19,690 citations

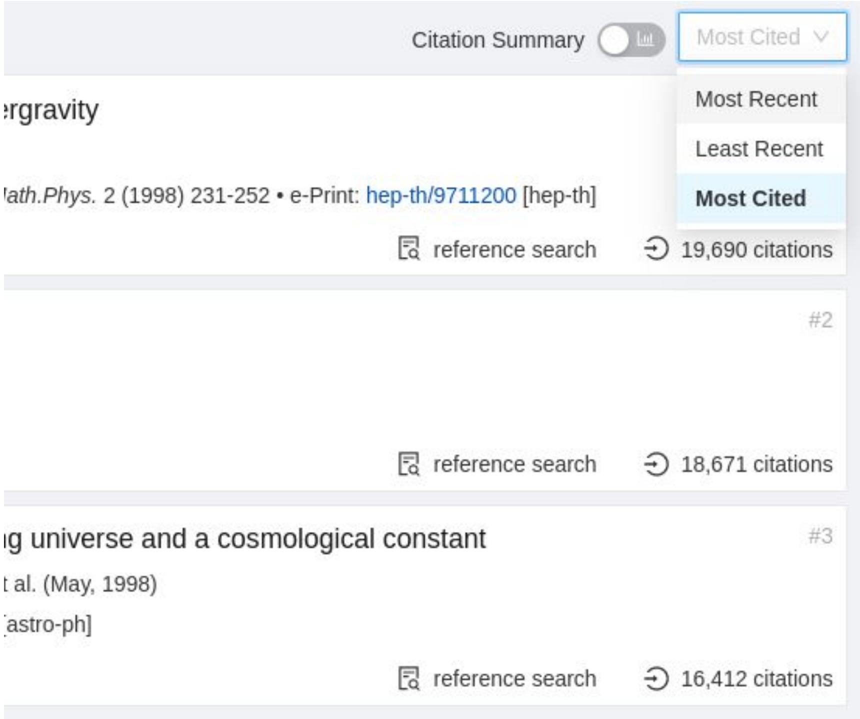
Loaded on-demand

Some out of the box aggregations

Some custom aggregations like the h-index

Written in painless

Results



ergravity

ath.Phys. 2 (1998) 231-252 • e-Print: hep-th/9711200 [hep-th]

reference search 19,690 citations

#2

reference search 18,671 citations

g universe and a cosmological constant #3

t al. (May, 1998)

astro-ph]

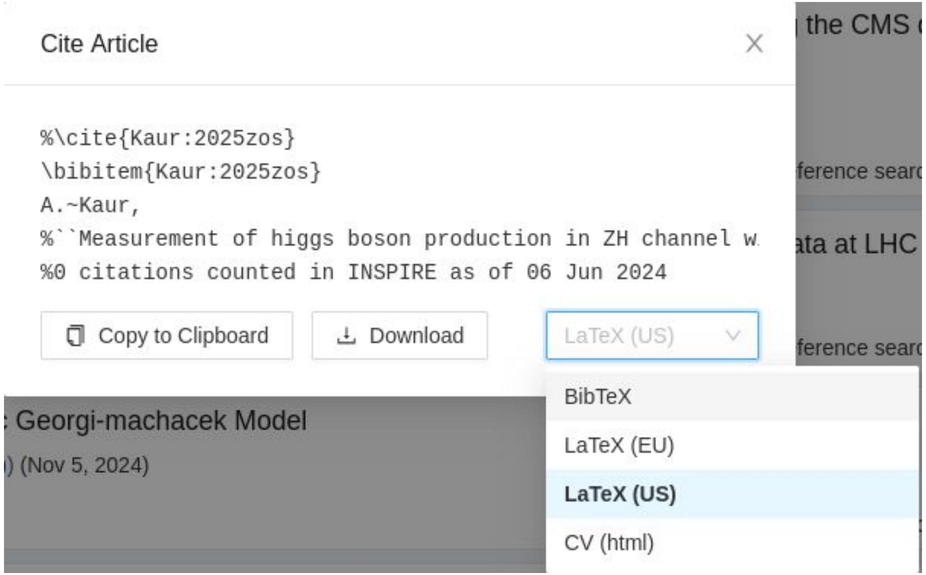
reference search 16,412 citations

No scoring

User select ordering

Default to AND instead of OR between search terms

Results - Serialization



Cite Article

```
%\cite{Kaur:2025zos}
\bibitem{Kaur:2025zos}
A.~Kaur,
%``Measurement of higgs boson production in ZH channel w
%0 citations counted in INSPIRE as of 06 Jun 2024
```

LaTeX (US)

Various serialization available for each record

- JSON (UI)
- HTML (embeddable)
- Bibtex
- Latex

Computed in the backend before indexing.

The JSON for the UI is stored as a blob to avoid any post-processing at query time.

Q&A

Any Questions?