SErAPIS Online: User Guide

Yiannos Stathopoulos and Angeliki Koutsoukou-Argyraki

June 2021

SErAPIS (**S**earch **E**ngine by the **A**lexandria **P**roject for **Is**abelle) is an experimental search engine for Isabelle 2021 and the Archive of Formal Proofs (AFP), 2021 release[1]. SErAPIS has two goals:

1. Allow Isabelle users to discover what is already available in the libraries and AFP using *mathematical concepts* – direct references to mathematical content and ideas.

2. Conduct research in information retrieval and machine learning of mathematical knowledge using Isabelle.

A mathematician using Isabelle to formalise e.g. new material on multisets might be wondering

"What is there formalised in Isabelle, either in the Libraries or the AFP, related to partitions of multisets that I could use?".

Our mathematician has a fairly precise idea of the result they are looking for but not enough knowledge of the naming conventions and organisation of the Isabelle libraries to answer their question quickly. This is the main use-case SErAPIS is designed to serve.

SErAPIS tries to understand what you are looking for by utilising *mathematical concepts*: natural language phrases that refer to mathematical objects, constructs and named results.

Of course, SErAPIS supports search by other means, such as keyword matching, topic and theory filtering and can help users of all levels to find what they are looking for (cf. Section 3). Furthermore, SErAPIS aims to present search results in a way that allows its users to quickly assess their relevance (e.g., using preview snippets and links to browsable theories) and to understand how facts and theories relate to each other for the purpose of discovery. SErAPIS is fast and searches all theories in the libraries and AFP simultaneously (unless the user has explicitly chosen to exclude certain theories as we will see).

In order for us to meet the second goal above, we kindly ask that you provide us feedback about which results are relevant to your searches. This is easy to do, as described in Section 4. However, SErAPIS records more than just the results relevant to each query. It also records your search session: how you refine your search and what you find interesting enough to click on in the set of results.

The data we record is extremely useful for understanding the search needs of Isabelle users, for evaluating the current efficiency of the search engine and for building intelligent tools for Isabelle based on Machine Learning. All data is recorded **anonymously** and cannot be traced back to you or your location. We are very grateful for your participation.

Please make sure you look at the "Constructing Effective Queries" portion of this guide (Section 3) so that you can make the best use of SErAPIS.

---

[1]Date of last indexing: 31 May 2021 (All AFP entries up to 31 May 2021 are included in the index.)

1 Issuing a Query to SErAPIS

SErAPIS is designed to allow you, the Isabelle user, to search the Libraries and Archive of Formal Proofs (AFP) easily, with as little input as possible. The main interaction with SErAPIS is done using 8 controls, but a search query can be issued just by filling one search box (see below).
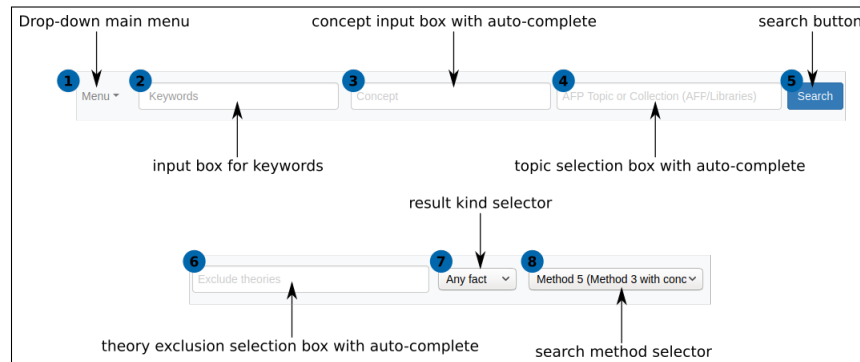

Figure 1: Search controls for the SErAPIS online Isabelle search engine.

Figure 1 shows the SErAPIS main UI controls[2], which are:

Drop-down main menu ①
This is a clickable, drop-down menu that allows you to navigate the sections on the SErAPIS website: "Home", "Instructions", "Feedback", "Privacy Policy" and "About SErAPIS".

Input box for keywords ②
This is where you can specify space-separated keywords that SErAPIS should match against theories or facts. The input can be references to fact or theory names, strings you would like matched against results and so on.

Concept input box with auto-complete ③
This box allows you to enter one or more mathematical concepts. There is a pre-defined set of concepts that SErAPIS has identified in the libraries and AFP and these will be presented to you as a drop-down list. You can start typing a concept you think is relevant to what you are looking for and SErAPIS will update its drop-down list of suggested concepts in real time.

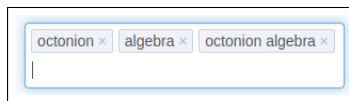Figure 2 shows how the concept input box might look like in your browser with two selected concepts.

Figure 2: Example of how concept input in SErAPIS looks like.

Concepts can be removed by clicking the "X" button at the top-right of each entry (cf Figure 2), or by pressing backspace to highlight and remove the concept. This behaviour is consistent across all SErAPIS input boxes that support auto-complete.

---

[2]on the actual website the controls are aligned in sequence in a search bar. They are divided here for better use of space.

Topic selection box with auto-complete ④

This box allows you to **optionally** specify which Isabelle collection you want to search. The search box allows you to enter one or more collections or topics and will automatically recommend topics known to be represented in our index as you type.

There are two main collections: the Libraries and AFP represented by the options `Libraries` and `AFP`, respectively. By adding `Libraries` in the box, you will not be able to further select the collection `AFP` (or its sub-topics) and vice-versa because the two collections are mutually exclusive.

However, you can select one or more sub-topics in the `AFP`, for example `Analysis` (the subset of the AFP `AFP/Mathematics/Analysis`) and `Algebra` (`AFP/Mathematics/Algebra`). This will make SErAPIS perform a search of the AFP index in a disjunctive manner: it will search for facts or theories that belong to either of the input topics.

If you wish to search specific sub-topics of the AFP, you only need to input the corresponding entries – you do not need to include the collection `AFP`. You should, however, enter `AFP` as the collection if you want to search the entirety of the AFP specifically.

If you leave the box blank, the search will be performed both in the AFP and the Libraries, through the entirety of both collections.

Search button ⑤

Clicking this button will result in your query being submitted to SErAPIS. You can also press enter to trigger your search.

Theory exclusion box with auto-complete ⑥

SErAPIS allows you to exclude specific theories from your search results. This is **optional** but very useful in cases when you know that what you are looking for comes from a particular topic and you want to filter results coming from unrelated theories.

There are two ways you can specify theories to be excluded. One way is to type the name of a theory directly in this box and SErAPIS will automatically recommend theories to be excluded from its index. The other way is to use the exclude button included with each search result to signal that you want results from the result's source theory to be excluded. This functionality is described in Section 2. Regardless of the method you choose, excluded theories are collected in this box and can be removed by editing the list in this box.

Result kind selector ⑦

SErAPIS indexes facts such as theorems, lemmata, corollaries, definitions and axioms. The first three kinds of facts fall under the "lemma" category but sometimes it is useful to search for definitions. This selector box allows you to select the kind of results you want SErAPIS to present to you.

Currently, you have three options:

1. `Any fact`: Show results of any kind.

2. `Definition`: Only show results that are definitions.

3. `Theory`: Find theories, rather than specific facts.

The third option accesses the theory search functionality of SErAPIS. You can search for theories, as well as facts, using the inputs described in this section.

Search method selector ⑧ Mathematics search is usually highly targeted. It is thus unlikely that one single search method will perform well in all mathematical searches. This is why we have 8 search algorithms in SErAPIS available to you and changing the default is optional.

As a user of SErAPIS you don't need to worry about how each method works but you have a few methods to try in case the default method does not give you good results. In that case, be aware of the following:

- Methods 1, 2, 3, 5, 6, 7 and 8 are *concept-oriented methods*: they use input concepts differently when trying to satisfy your queries. If one of these methods is not producing good results in response to the concepts you input, another method might.

- Methods 1, 2, 3, 5, 6, 7 and 8 can be used with keywords, concepts or both.

- Method 4 requires keyword input and does not use concepts to search the SErAPIS index. SErAPIS cannot execute a query using method 4 unless at least one keyword is provided.

Details about how methods 1 to 4 work are discussed in [1, 2].

2 SErAPIS Results

SErAPIS will respond to your query by returning a fixed number of results it thinks are most relevant to your query, in descending order of relevance. Each result is presented as a result card that presents information about the fact or theory and allows you to interact further with the search engine to refine your search. The information displayed in a card varies depending on whether you are searching for a fact or theory.

2.1 Fact Search

Result cards produced by SErAPIS in response to a search for facts will look like the card in Figure 3.

Figure 3: A fact result card.

Each card has four elements: (a) A header, (b) a "Used by" block, (c) a "Uses" block and (d) a "Preview snippet" block.

The header displays the eight pieces of information highlighted in Figure 4.

All header information elements are buttons. The rank, name and kind of fact, source collection and topic can be clicked to collapse the card into a single row. This enables you to reduce the quantity of information displayed and to focus on results that are of direct interest to you.

Clicking on the source theory will open the source theory in a new tab in your browser. The red "exclude source theory" button (item 7 in Figure 4) will add the source theory in the theory exclusion box and as a result it will be automatically excluded from the next search results (you
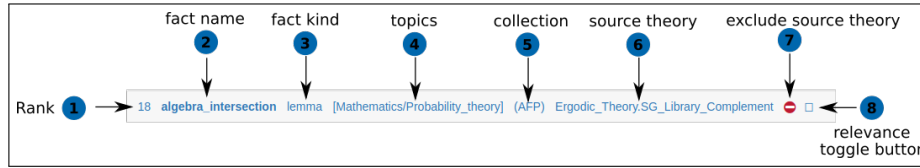
Figure 4: A fact result card.

will have to click on the Search button again to repeat the search this time considering the excluded theory/theories).

The green "relevance toggle button", which looks like a tick, can be used to give us feedback on the relevance of a particular result to your query. We give detailed instructions on how to do this in Section 4 and we kindly ask that you give us feedback when you find results relevant to your query – it's easy!

The "Used by" and "Uses" blocks list facts that use the fact in the card and other facts it uses, respectively. Both blocks are displayed if applicable; that is when the fact represented by the card is used by other facts and/or uses other facts in its body and are collapsible by clicking on their title. The items listed in this block are colour-coded:

- ■ for axioms;

- ■ for lemmata, theorems and corollaries;

- ■ for definitions.

This convention is used in both fact and theory search.

The "Preview snippet" block displays a code snippet so that you can assess on the relevance of each result to your query quickly. You can also collapse this block by clicking on the text in the title header.

2.2 Theory Search

You can search for theories, rather than facts, by selecting "Theory" from the result kind selector (Section 1) and creating a query in the same way you would for finding facts: keywords,, concepts, topics and exclusions.

However, the way theory results are presented differs slightly from the way fact results are presented. As shown in Figure 5, theory results are presented collapsed by default, so you can skim through the theory names and topics quickly and have more sub-blocks once expanded. Header information and header buttons in theory results are identical to those for facts. You can expand the cards by clicking on the theory name, topic, collection or session. A link to the source code of the theory is accessibly by clicking on the full name of the theory presented in the format "Session.Theory" (e.g., click on "Octonions.Octonions" in the expanded card in Figure 5).

The "concept set" block enables you to view some of the concepts identified in the theory. The "dependencies" block lists all theories imported via `include` commands and you can choose to view them by clicking on their names. The "facts" block enumerates all facts defined in the theory, color-coded using the scheme described in Section 2.1. Lastly, the "comments" block gives a quick overview of author comments, so you can get an idea of what a theory result is about without having to pause your searching to open and examine the theory file.
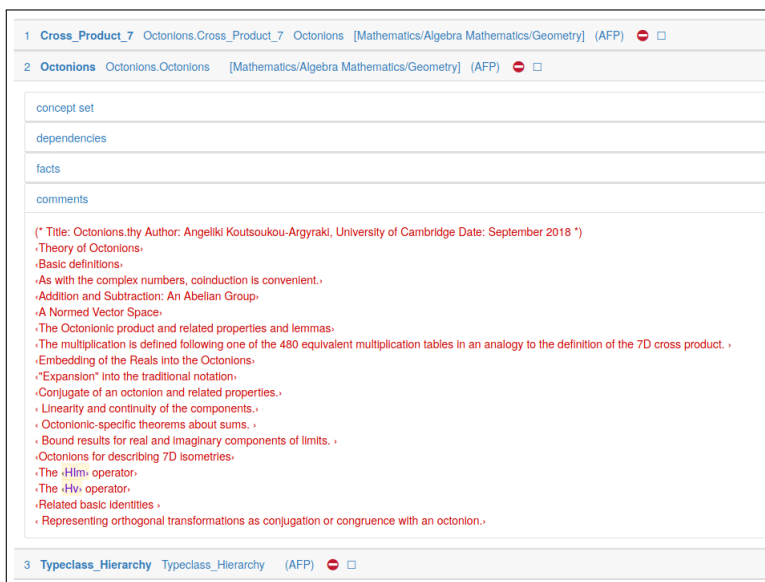
Figure 5: Results of a theory search, including theory result cards.

3 Constructing Effective Queries

Broadly speaking, the combination of keywords, concepts and topics affect the quality (in terms of relevance) of the results you will obtain. The search method also affects the results you will see: the same query (keywords, concepts and topics) will produce different results across methods. SErAPIS allows you, the user, to explore results from all methods by choosing retrieval methods from the method selector (cf. Figure 1).

Every search query and user is unique and we cannot anticipate everything. However, the following rules of thumb might help you find what are you are looking for when that becomes illusive:

- Using concepts that are too general might produce many irrelevant results while using too specific concepts might not produce useful results at all. The key is trying to balance generality/specificity by using concepts that describe the most specific generalisations of the objects or mathematical ideas you are looking for.

- Methods 1,2,3, 5, 6, 7 and 8 are concept-oriented methods: they use input concepts differently when trying to satisfy your queries. If one of these methods is not producing good results in response to the concepts you input, another method might.

- Recall that keywords are used for pattern matching, so use these when you expect the desired result to include the keywords in its name (i.e,. when you use a characteristic name of a theorem, such as "Zorn","Nullstellensatz", "Caratheodory", "Lebesgue", "Cauchy", etc.) or you are already aware of the name of the theory in which your desired result resides. Avoid using generic concepts, such as "division", as these are common in the index and will likely introduce irrelevant results.

- Try to give as much information to SErAPIS as possible. Specifically, if you know that what you are looking for is in the AFP and have a rough idea of its topic, it might be better to put this information in the topic box, rather than adding topic keywords in the keyword box.

- Use the theory exclusion box with successive queries. Consider this: after issuing a query you see results coming from theories related to Algebra and Computer Science. Suppose you

are interested only in Algebra. You can use the "exclude source theory" button in fact card headers to exclude specific theories. Using this functionality in successive queries you can iteratively improve your results quickly until a relevant fact or theory is found.

- Be aware that some concepts have synonyms. In general, there are mechanisms behind the SErAPIS search engine that address synonymy and polysemy of mathematical concepts. However, sometimes you, the user, can benefit by including an additional concept that is synonymous to what you are looking for, if one is available. For example, searching using the concept `directed graph` might benefit by including its synonym, `digraph`.

- A useful search technique for the user who wishes to find something specific (or perhaps even to just explore the Libraries and AFP for material related to a certain mathematical object) is the following: Instead of "Any fact" first look for "Theory" using ⑦ in Figure 1. You can then locate the most interesting theories/y for you in the search results. If these are only in the AFP or only in the Library, repeat your search having specified AFP/Library accordingly using ④ when now looking for facts. You can also repeat your searches by excluding theories from the results by using the red "exclude source theory" button; the excluded theories will appear at ⑥ and you can now repeat your search ensuring exclusion of the unwanted theories.

4   Relevance Feedback

Please give us feedback on at least the first 10 results of your search top 10 results (you can give feedback for more if you so choose). The process to do so is simple. Your help is much appreciated.
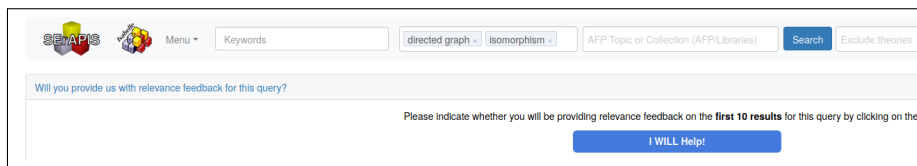

Figure 6: Results of a theory search, including theory result cards.

At the top of every set of results, SErAPIS will ask

Will you provide us with relevance feedback for this query?

as shown in Figure 6.

Step 1   The first step in providing us relevance feedback is to click the "I WILL Help!" button. By doing so, you notify us of your intention to provide relevance feedback and will bring your query to our attention in our research. Once you click it, the "I WILL Help!" button will change to an "UNDO" button, as shown in Figure 7, which allows you to undo your decision.

Figure 7: Results of a theory search, including theory result cards.

If you choose to help we kindly ask that you look through at least the first 10 results of your search closely (feel free to look through more if you so choose) and make a decision on whether the

fact or theory listed is relevant to your query. The following are good reasons to mark a result as "relevant":

- the result fulfills your enquiry: it is a fact or theory that represents the specific methods, techniques or results you are looking for;

- it is something of interest in your exploration of the libraries and AFP: a theorem, definition or other construct that you have been wondering about its inclusion in the Isabelle collections;

- it is something useful in your own work or a satisfactory answer to your question (query) or curiosity;

- it is a surprising or interesting instantiation of the mathematical concepts in your query.

Step 2 To indicate that a search result is relevant, please click the "relevance toggle button" located at the far-right of the fact result card header. This looks like an unfilled check box, as shown in Figures 4 and 8 (top).
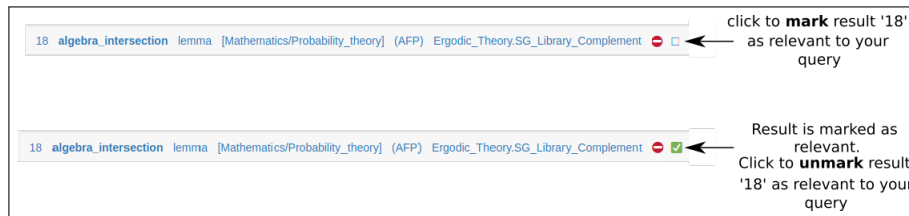

Figure 8: Relevance toggle buttons. Top: toggle button is an unfilled box that allows you to mark a particular result as relevant. Bottom: clicked once, the toggle button becomes a green tick indicating that the result is marked as relevant; if clicked again the result will be unmarked as relevant and reset to an unfilled box.

Once you mark a result as relevant, the toggle button on its header will change from an unfilled box to a green tick, as shown in Figure 8 (bottom). This is to indicate that clicking the toggle button again will allow you to undo your previous click (e.g., if you made a mistake or changed your mind). Clicking the tick will unmark the result as relevant and return the relevance toggle button to its original form (i.e., a an unfilled check box) and function (mark as relevant).

Step 3 Please review the at least the top 10 results (you can review more if you so choose) and click on the toggle button on each result that you find relevant to your query, as explained in step 2.

Acknowledgements

References

[1] Stathopoulos, Y., Koutsoukou-Argyraki, A. and Paulson, L. C., *SErAPIS: A Concept-Oriented Search Engine for the Isabelle Libraries Based on Natural Language*, online proceedings of the

Isabelle Workshop 2020 affiliated to IJCAR 2020, (in Virtual Space), June 30, 2020, available on `https://files.sketis.net/Isabelle_Workshop_2020/Isabelle_2020_paper_4.pdf`.

[2] Stathopoulos, Y., Koutsoukou-Argyraki, A. and Paulson, L. C., *Developing a Concept-Oriented Search Engine for Isabelle Based on Natural Language: Technical Challenges*, online proceedings of the 5th Conference on Artificial Intelligence and Theorem Proving (AITP 2020), Aussois, France, Mar. 22-27 (event postponed to Sept. 13-18, 2020), available on `http://aitp-conference.org/2020/abstract/paper_9.pdf`.