

## FAQ

20 March 2026

### How to use the SMAFIRA search engine

---

SMAFIRA (<https://smafira.bf3r.de/>) is a free online tool that is intended to support biomedical researchers and animal welfare officers in their efforts to screen the literature database PubMed for alternative methods to animal experiments. SMAFIRA does not evaluate the suitability of retrieved methods. This needs to be done by the involved human experts. SMAFIRA does however decrease the time needed to search the literature for potential alternatives. All you need is a PubMed identifier (PMID). Should any of your questions not be answered by the following FAQ, you may contact us directly.

## SMAFIRA Tool

---

### General issues

**Does the tool collect any information from the user? Does it use cookies or similar technical elements?**

Please refer to the [Data Protection Declaration](#).

---

### Search and processing

**Which types of searches are possible?**

Currently, you can conduct a similarity search using a PMID (PubMed identifier). Please go to PubMed first and identify an article that represents your research goal and can serve as a reference for identifying similar contents. Then, you can enter the PMID of the reference into the input field of SMAFIRA.

**Which PMIDs are allowed in the search?**

A valid PMID retrieves a citation from PubMed. This is validated by SMAFIRA. Furthermore, the citation must contain an abstract and must be linked with similar articles.

### **Which articles are considered by SMAFIRA (coverage)?**

Currently, SMAFIRA analyses similar articles retrieved from PubMed up to a maximum number of 10,000. Usually, the number of retrieved similar articles is below 500.

### **May I enter more than one PMID in a query?**

No, each search (session) must contain only one PMID.

### **Which processing is performed on the articles?**

SMAFIRA performs an automatic classification to predict the experimental models mentioned in the abstracts, followed by a re-ranking according to the similarity to the reference abstract. For more details about our methods for classification and re-ranking, see "Methods and data" below.

### **Can I start my search from a project summary in other languages than English?**

Yes. You can start from a detailed summary of your experimental project in a language other than English. Then, you should use translation tools to translate your summary into English, and the [MeSH on Demand tool](#) to identify an equivalent citation in PubMed. Use the PMID of the latter citation for your search.

---

## **List of results**

### **What information is shown for each citation in the list?**

SMAFIRA presents the following information:

- Rank (number) of the citation in the list, according to the chosen method.
- PMID, with a link to the side-by-side visualization of the reference and similar abstract (user opinion page).
- External link to the article in PubMed. (<sup>1</sup>)
- SMAFIRA Rank (default ranking): rank of the citation according to the SMAFIRA re-ranking.
- PubMed Rank: rank of the citation according to PubMed.
- My Rank: an icon that indicates the similarity between the reference abstract and the current abstract as judged by the user. This icon is only available after a user has activated the respective checkbox on the side-by-side comparison of a single article with the abstract.
- Title of the cited article.
- Labels: one or more icons that indicate the assigned classes of experimental models that were automatically predicted (indicated by ) or specified by the user (no cogwheel icon).

<sup>1</sup> The icons in this application were made by [Freepik](#), [Good Ware](#), [juicy\\_fish](#), [Roundicons](#) from [www.flaticon.com](http://www.flaticon.com).

- If available, icons that indicate whether the article is a review (📖), as specified in PubMed) or that no abstract is available (🚫).

### **Why are some filters disabled (icon 🚫) sometimes?**

This occurs when none of the articles in the list mentions the respective experimental model in the abstract text, i.e., it is not possible to filter the list by this label.

### **How are the citations ranked?**

By default, the tool ranks the citations according to the SMAFIRA re-ranking. However, the user has two other options for ranking the articles: (a) the PubMed ranking, i.e., the one that is carried out by the PubMed ranking algorithm; and (b) the user-defined ranking (My Rank) based on the similarity validation, and in this order: Thumb up, thumb up/down, thumb down.

### **Why are similar articles without abstracts included in the list of results?**

For the sake of completeness, we do not remove them, but they are usually located at the end of the list. Since no abstract is available, we can neither predict the labels nor calculate its rank with the SMAFIRA re-ranking.

### **What exactly does each label mean?**

We give a short definition below, please refer to our [publication](#) for details.

- in vivo: living vertebrates (including cephalopodes)
- tissue/biopsy: isolated vertebrate organs and tissues
- primary cell: finite vertebrate primary cells or stem cells
- immortal cell: established vertebrate cancer or immortalized cell lines
- invertebrate: invertebrates or invertebrate material (excluding cephalopods)
- human: human/patients and or any kind of human material (including organs/tissues, primary cells and immortal cells)
- virtual model: computer simulations
- others: other types of research, including observational studies

### **How were the filters (labels) decided?**

We chose labels that indicated not only the level of invasiveness but also the species or the origin of material used. Please refer to our [publication](#) for details.

### **May I come back to my results later? Or share it with someone else?**

Yes, please save (or bookmark) the URL of the session that is specified at the top of the results page.

## Can I download or export the list of results?

Yes, please click on the export icon  on the top right of the results page. All articles in the list will then be exported, along with the feedback provided by the user for the labels and similarity.

---

## Visualization of a citation in the list

### Which information is shown in the visualization?

We show a side-by-side visualization with the reference abstract (on the left) and the abstract from the currently chosen citation (on the right). On the top left is a link  of results to go back to the list of citations, while on the top right are navigational links to the previous  (if available) and the next  (if available) citation. The page includes the possibility to give feedback on the similarity between the two abstracts (see details below). For each abstract, we present the following information:

- Information about the automatically predicted classes of experimental models (indicated with the icon .
- PMID of the citation and link to it in PubMed.
- Title of the corresponding article, as available in PubMed.

### Can I change the automatically predicted class(es) of experimental model(s)?

Yes, the user can override the automatically predicted labels. Any changes to the labels are automatically saved within the context of the respective user. To suggest a global change of a prediction – affecting all users of SMAFIRA – the change has to be forwarded to the SMAFIRA team, by sending the session URL to the contact e-mail. Then, the suggestion will be considered during a follow up training of the prediction model.

### When are two articles similar?

We consider two abstracts as similar, if the described scientific objectives agree. Please refer to our [publication](#) for a stringent specification of similarity (“equivalence”).

### How can I give feedback about the similarity?

Simply choose one of the three options (thumb up – similar, thumb up/down – uncertain, thumb down – not similar). The selected one is automatically saved. It is possible to clear your selection by clicking on the "clear" link next to the options.

### How can the user feedback be used?

First of all, users can use their own judgement of similarity to rank citations (“My rank”). Furthermore, user feedback will be shown in the other rankings (“SMAFIRA” rank, “PubMed rank”) as well. Thus, users can evaluate the performance of the ranking algorithms: the more citations on the top positions are judged as similar to the reference, the better. If

users send us the URL of their session, including such feedback on similarity by e-mail, we can internally use the data to train or evaluate the prediction models of SMAFIRA.

## Methods and Data

---

### Prediction of the methods:

#### **How does the prediction of the class of experimental model (label) work?**

We fine-tuned a machine (deep) learning model that is based on the [Transformers framework](#). This is the current state-of-the-art approach for text classification tasks. The [PubMedBERT](#) language model was pretrained on PubMed® contents (MEDLINE®). For fine-tuning we produced a data set of annotated biomedical abstracts (see next question). The obtained model is used for a real-time prediction of the labels assigned to the reference and similar abstracts. Please refer to our [publication](#) for details.

#### **What data was used for training the models?**

As described above, the pre-trained language model was further fine-tuned with a collection of around 1,600 annotated biomedical abstracts (called the GoldHamster corpus). A team of more than 10 experts manually annotated the abstracts with labels according to the mentioned experimental model(s). Please refer to our [publication](#) for details.

#### **What if the prediction is wrong?**

Users can override the automatic prediction on the user opinion page. If the URL of the respective session is sent to our contact address by e-mail, we can use this information for future fine-tuning of our model.

#### **May I get an explanation why a particular label was (or not) assigned to an article?**

Not yet, but we plan to implement such a functionality in the future.

---

### Re-ranking:

#### **How does the re-ranking of the similar articles work?**

We start by representing each article as a vector of words whose representation uses [biomedical word embeddings provided by the NCBI](#). Then we calculate the similarity score between the two vectors using a [dot similarity from Python numpy](#). The tool ranks the citations according to these scores.

**What data was used for the evaluation of the SMAFIRA re-ranking?**

We relied on our [SMAFIRA-c dataset](#). It contains four use cases for which the top 100 PubMed similar articles were annotated regarding the similarity (scientific objective) of test and reference abstract. Please see our [publication](#) for details on our experiments with various methods.

### About the BfR

The German Federal Institute for Risk Assessment (BfR) is a scientifically independent institution within the portfolio of the German Federal Ministry of Agriculture, Food and Regional Identity (BMLEH). It protects people's health preventively in the fields of public health and veterinary public health. The BfR provides advice to the Federal Government as well as the Federal States ('Laender') on questions related to food, feed, chemical and product safety. The BfR conducts its own research on topics closely related to its assessment tasks.

### About the Bf3R

The German Centre for the Protection of Laboratory Animals (Bf3R) was founded in 2015 and is an integral part of the German Federal Institute for Risk Assessment (BfR). It co-ordinates nationwide activities with the goals of restricting animal experiments to only those which are considered essential, and safeguarding the best possible protection for laboratory animals. Moreover, it intends to stimulate research activities and encourage scientific dialogue.

#### Legal notice

Publisher:

**German Federal Institute for Risk Assessment**

Max-Dohrn-Straße 8-10

10589 Berlin, Germany

T +49 30 18412-0

F +49 30 18412-99099

[bfr@bfr.bund.de](mailto:bfr@bfr.bund.de)

[bfr.bund.de/en](https://bfr.bund.de/en)

Institution under public law

Represented by the president Professor Dr Dr Dr h. c. Andreas Hensel

Supervisory Authority: Federal Ministry of Agriculture, Food and Regional Identity

VAT ID No. DE 165 893 448

Responsible according to the German Press Law: Dr Suzan Fiack



valid for texts produced by the BfR  
images/photos/graphics are excluded unless otherwise indicated)

**BfR** | Identifying Risks –  
Protecting Health