

Open  
Research  
Knowledge  
Graph

# Intertwining Machine Learning and Semantic Web in the Open Research Knowledge Graph (ORKG)

Kheir Eddine Farfar

Data Week Leipzig 2024



# Introduction

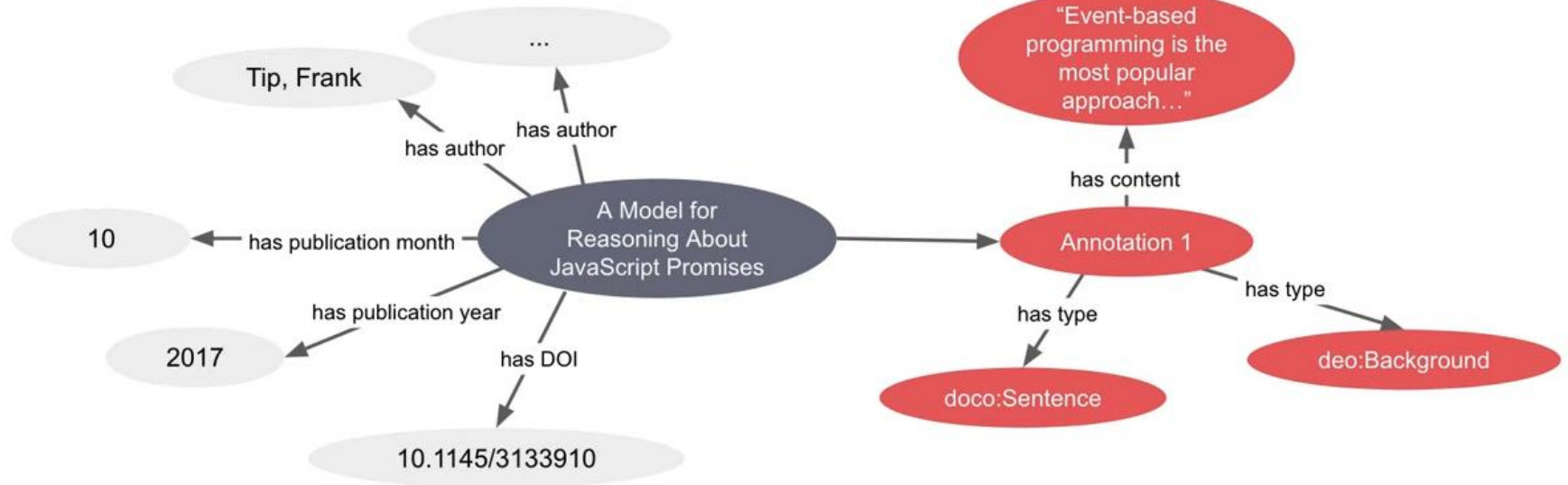
## Introduction | Scholarly communication

- Largely **document-based**
- Hinders **machine-actionability**, machines cannot easily understand the articles



# Introduction | Structured knowledge

With **knowledge graph**, the article's content can be described in a structured manner



## Introduction | Making structured knowledge

### Automated techniques

Machine learning

✘ Not sufficiently accurate

### Manuel techniques

Crowdsourcing

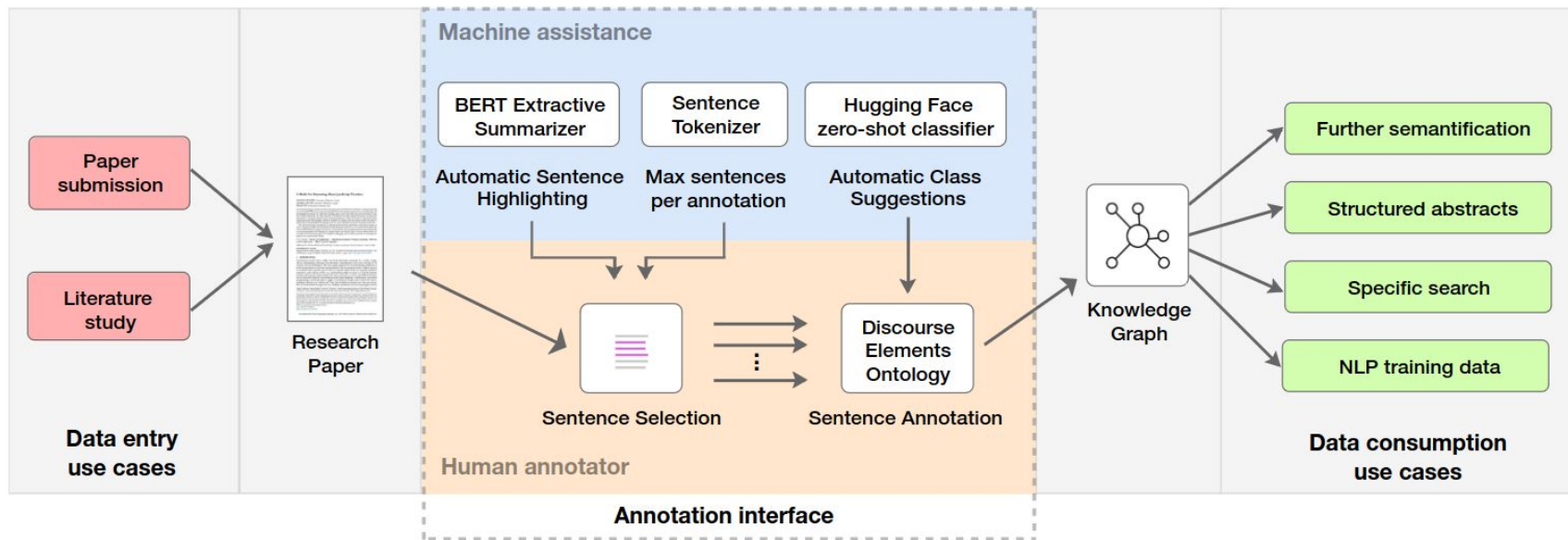
✘ Require skilled workers in Semantics

Interwinning Machine learning and Semantic web



# System design

# System design: Paper annotator



Allard Oelen, Markus Stocker, and Sören Auer. 2021. Crowdsourcing Scholarly Discourse Annotations. IUI '21. <https://doi.org/10.1145/3397481.3450685>

# User interface: Paper annotator

## Paper annotator

Save

Completion 10%

Background 1 annotation

“ The number of scholarly publications grows steadily every year and it becomes harder to find, assess and compare scholarly knowledge effectively. ”

Contribution 0 annotations

Methods 0 annotations

Problem statement 0 annotations

Results 0 annotations

---

Related work 1 annotation

“ Prominent examples of openly available knowledge graphs include DBpedia [4], YAGO [51] and Wikidata [56]. With projects such as Semantic Scholar [3], Microsoft AcademicGraph [47] and Open Research

## Crowdsourcing Scholarly Discourse Annotations

Lucas Stocker  
lstocker@tib.eu  
Information Centre for  
and Technology  
over, Germany

Sören Auer  
soeren.auer@tib.eu  
TIB Leibniz Information Centre for  
Science and Technology  
Hannover, Germany

document-based. Scholarly articles are mostly published in PDF format, which is specifically designed for human readability [38] and portability across systems. With this form of publishing, scholarly knowledge is not machine actionable [9, 41]. Knowledge graphs can be employed to represent scientific contributions semantically, render scholarly knowledge more machine actionable, and thus making it easier to find, compare and process knowledge. Knowledge graphs are defined as semantic networks describing entities and their interrelations [42]. Prominent examples of openly available knowledge graphs include DBpedia [4], YAGO [51] and Wikidata [56]. With projects such as Semantic Scholar [3], Microsoft Academic Graph [47] and Open Research Knowledge Graph (ORKG) [26], knowledge graphs are gaining popularity in the scholarly domain to structure scholarly knowledge. Except for ORKG, these graphs only capture metadata about research articles and do not describe the content of reported research work, including research contributions [44].

Populating knowledge graphs with scholarly metadata is a relatively straightforward task due to the low task complexity and high accuracy of automated parsing tools (such as GROBID [33]). In contrast, generating graphs of the contents of research articles (i.e. research contributions) is a considerably more complex task which

Select type

Select...

Smart suggestions

Related work

Methods

Contribution

Future work

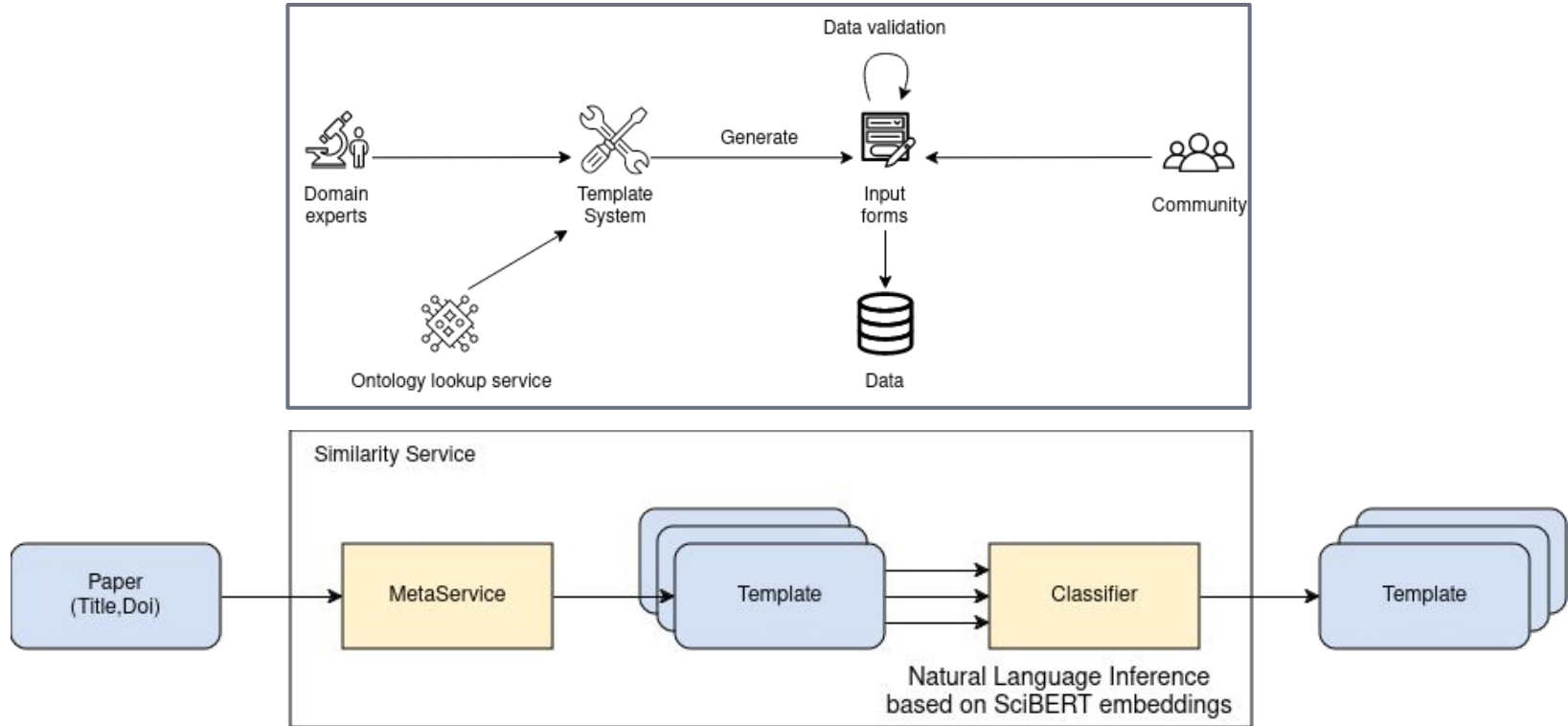
Scenario

Annotate

scholarly knowledge from paper authors with a web-based user interface supported by artificial intelligence. The interface enables authors to select key sentences for annotation. It integrates multiple machine learning algorithms to assist authors during the annotation, including class recommendation and key sentence highlighting. We envision that the interface is integrated in paper submission processes for which we define three main task requirements: The task has to be (1) straightforward (2) time efficient (3) well-defined. We evaluated the interface with a user study in which participants were assigned the task to annotate one of their own articles. With the resulting data, we determined whether the participants were successfully able to perform the task. Furthermore, we evaluated the interface's usability and the participant's attitude towards the interface with a survey. The results suggest that sentence annotation is a feasible task for researchers and that they do not object to annotate their articles during the submission process.




# System design: Template recommendations



Arab Ogli, Omar: Information Retrieval Service Aspects of the ORKG. Gottfried Wilhelm Leibniz Universität, Master Thesis. <https://doi.org/10.15488/11834>

# User interface: Template recommendations

## Transmission interval estimates suggest pre-symptomatic spread of COVID-19 ☆ 🔍



📅 March 2020
💬 142 citations
☰ Virology

👤 Lauren Tindale
👤 Michelle Coombe
👤 Jessica Stockdale
👤 Emma Garlock

👤 Wing Yin Venus Lau
👤 Manu Saraswat
👤 Yen-Hsiang Brian Lee

👤 Louxin Zhang
👤 Dongxuan Chen
👤 Jacco Wallinga

👤 Caroline Colijn

DOI: <https://doi.org/10.1101/2020.03.03.20029983>

✎ Edit metadata
🗑 Delete paper

✔ Verified

Contribution 1
🗑 ✎

Contribution 2

+

🔗 Templates
🔗 Help
⚙ Preferences

🔗 Applied template: Contribution

95% Confidence interval

1.45-2.48

+

location

Singapore

+

Add to comparison

?

✎ Paper abstract

✎ Annotator

Templates

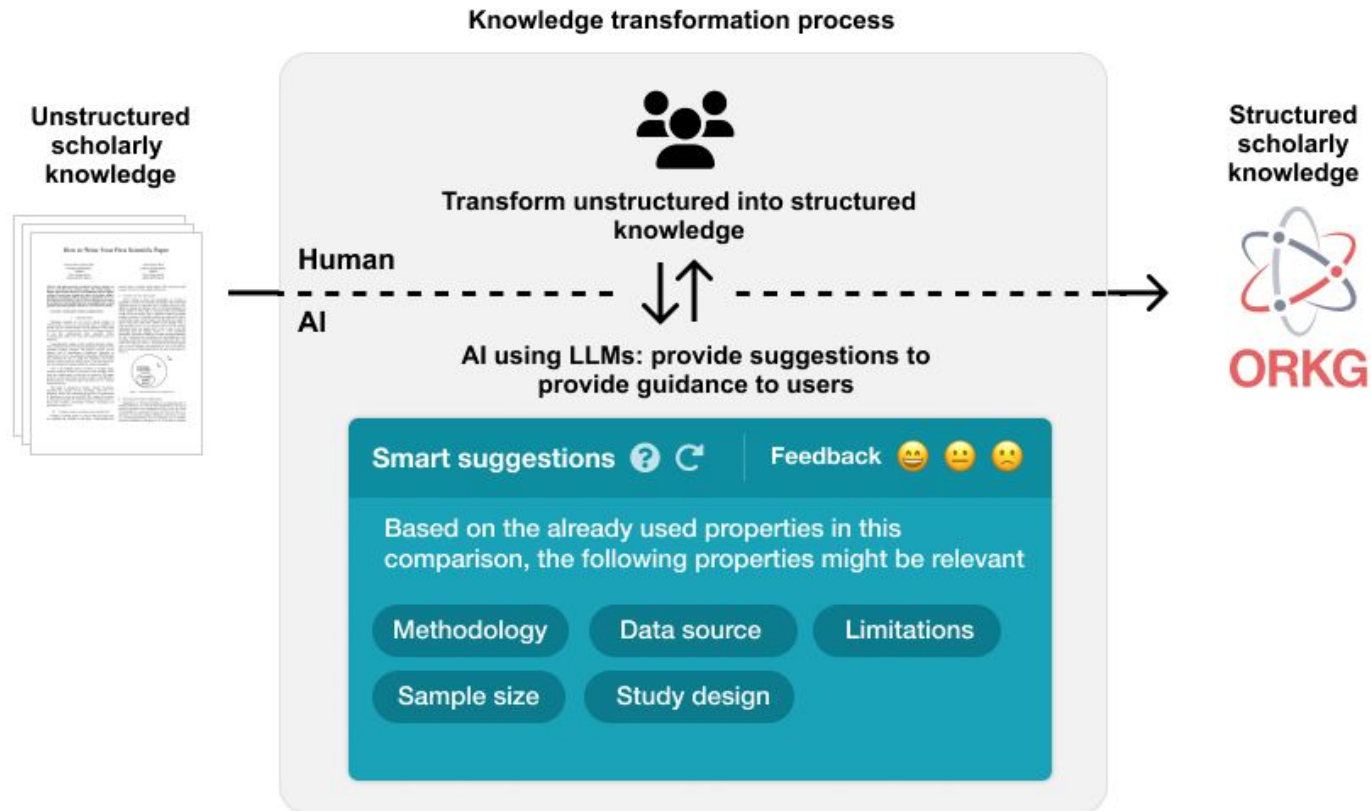
+
Basic reproduction number estimate

Provenance

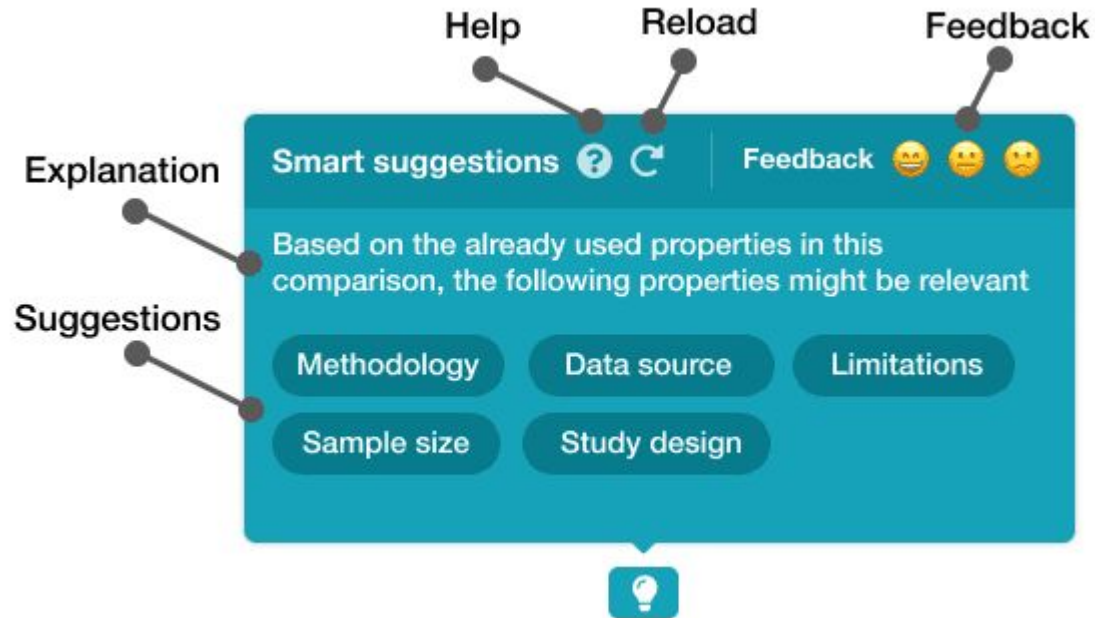
Timeline

Belongs to observatory

# System design: Smart suggestions



## User interface: Smart suggestions



## User interface: Collecting feedback

Positive feedback

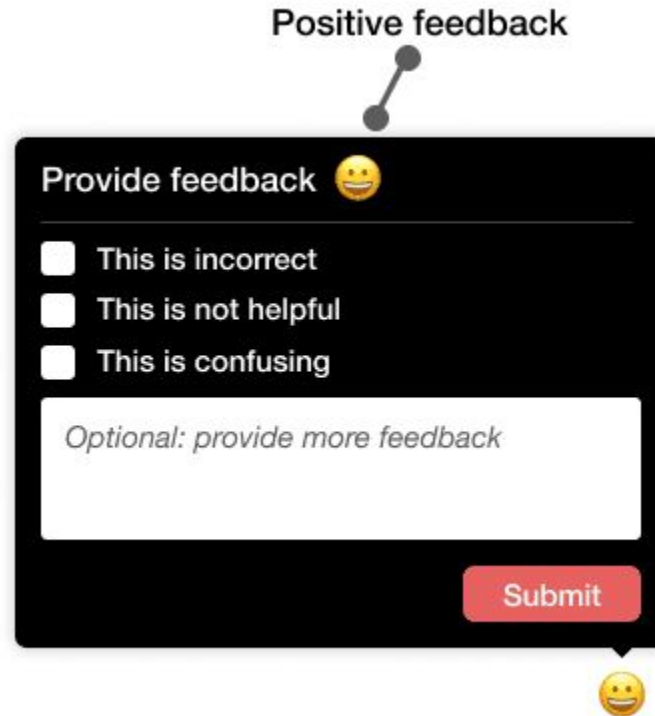
Provide feedback 😊

- This is incorrect
- This is not helpful
- This is confusing

*Optional: provide more feedback*

Submit

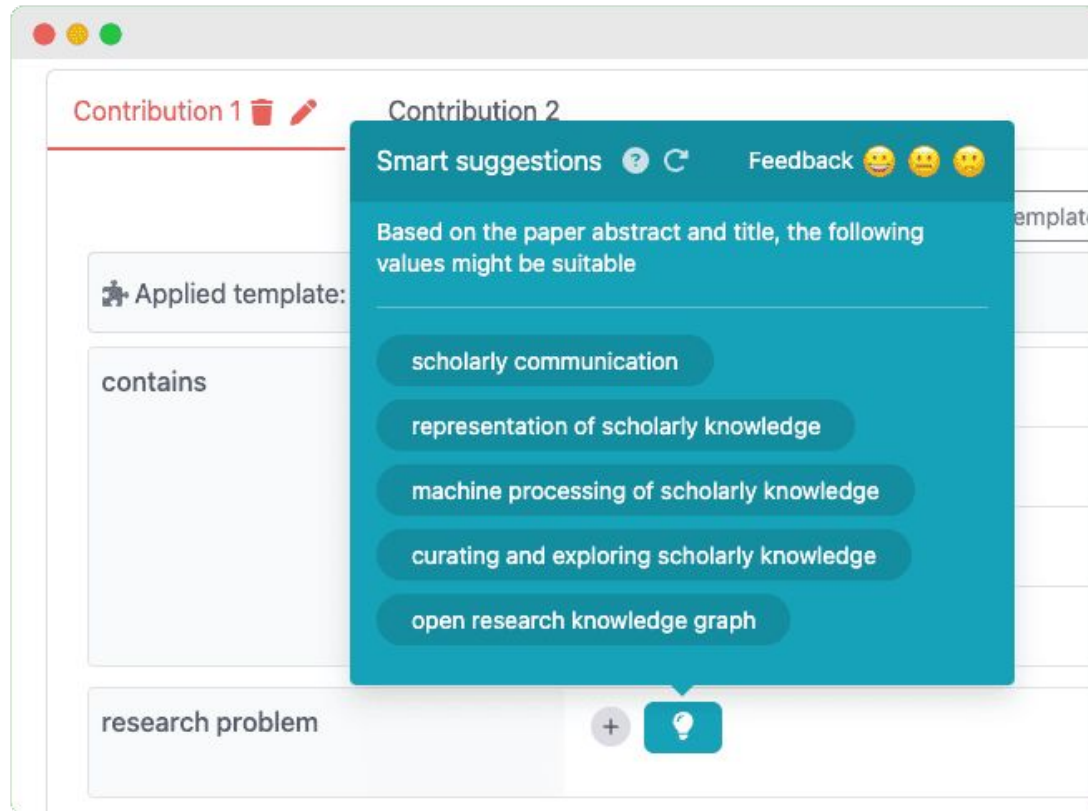
😊





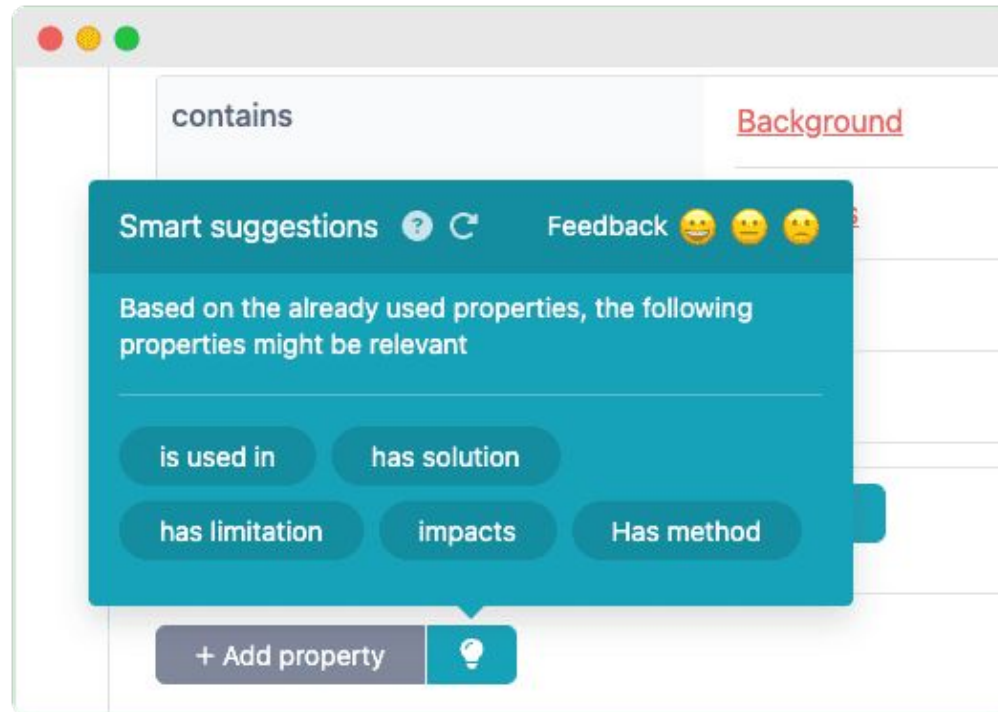
# Use cases

## Use cases: Resource suggestions



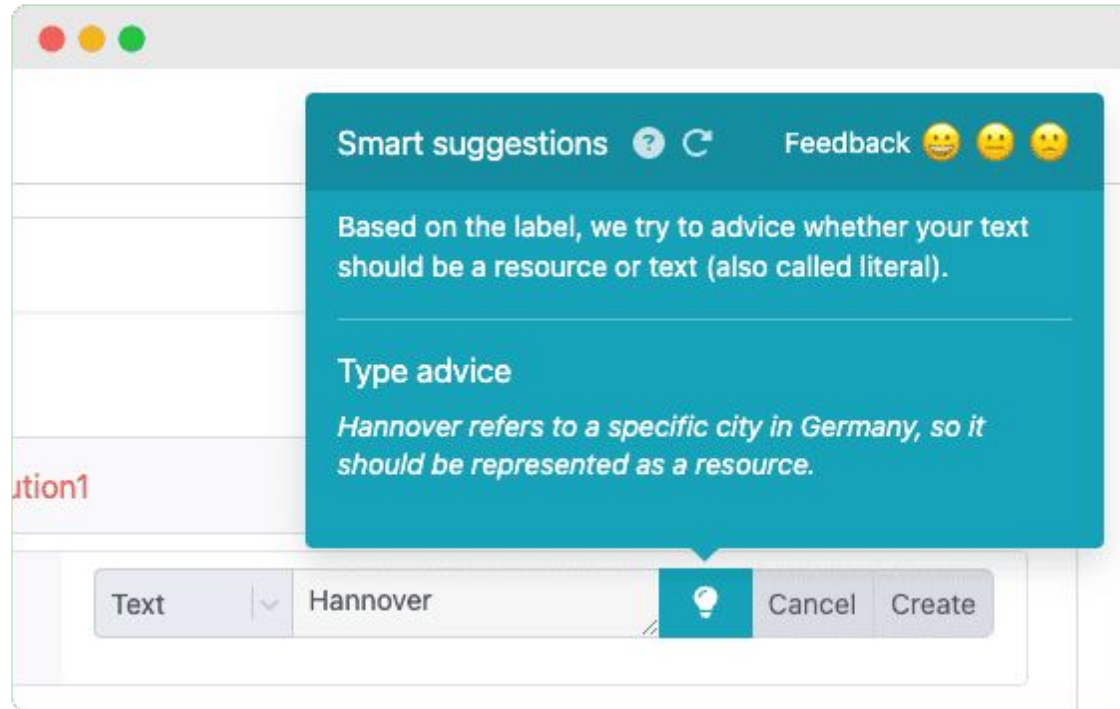
The screenshot shows a web interface with two contribution cards. The first card, 'Contribution 1', is partially visible and has a red trash icon and a red pencil icon. The second card, 'Contribution 2', is the focus. A teal popup window is overlaid on the top right of the 'Contribution 2' card. The popup has a title 'Smart suggestions' with a question mark icon and a refresh icon, and a 'Feedback' section with three emoji icons (happy, neutral, sad). The main text in the popup reads: 'Based on the paper abstract and title, the following values might be suitable'. Below this text are five rounded rectangular buttons, each containing a suggestion: 'scholarly communication', 'representation of scholarly knowledge', 'machine processing of scholarly knowledge', 'curating and exploring scholarly knowledge', and 'open research knowledge graph'. In the background, the 'Contribution 2' card shows an 'Applied template:' section with a puzzle piece icon, a 'contains' section, and a 'research problem' section with a plus icon and a lightbulb icon.

## Use cases: Predicate suggestions

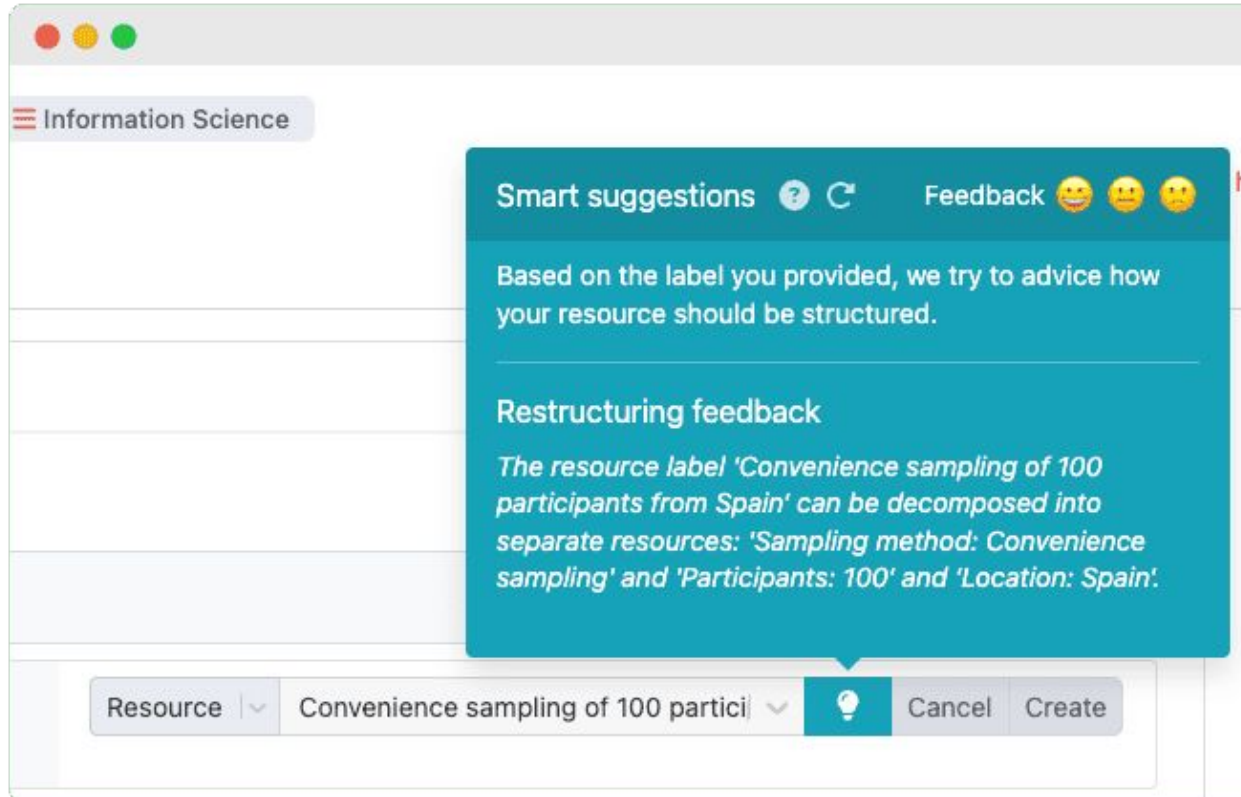




## Use cases: Feedback on whether literal should be a resource



## Use cases: Feedback on resource structure



The screenshot shows a web application window titled "Information Science". A teal dialog box is overlaid on the page, titled "Smart suggestions" with a question mark icon and a refresh icon. To the right of the title is a "Feedback" section with three emoji icons: a smiling face, a neutral face, and a sad face. The dialog contains the following text:

Based on the label you provided, we try to advise how your resource should be structured.

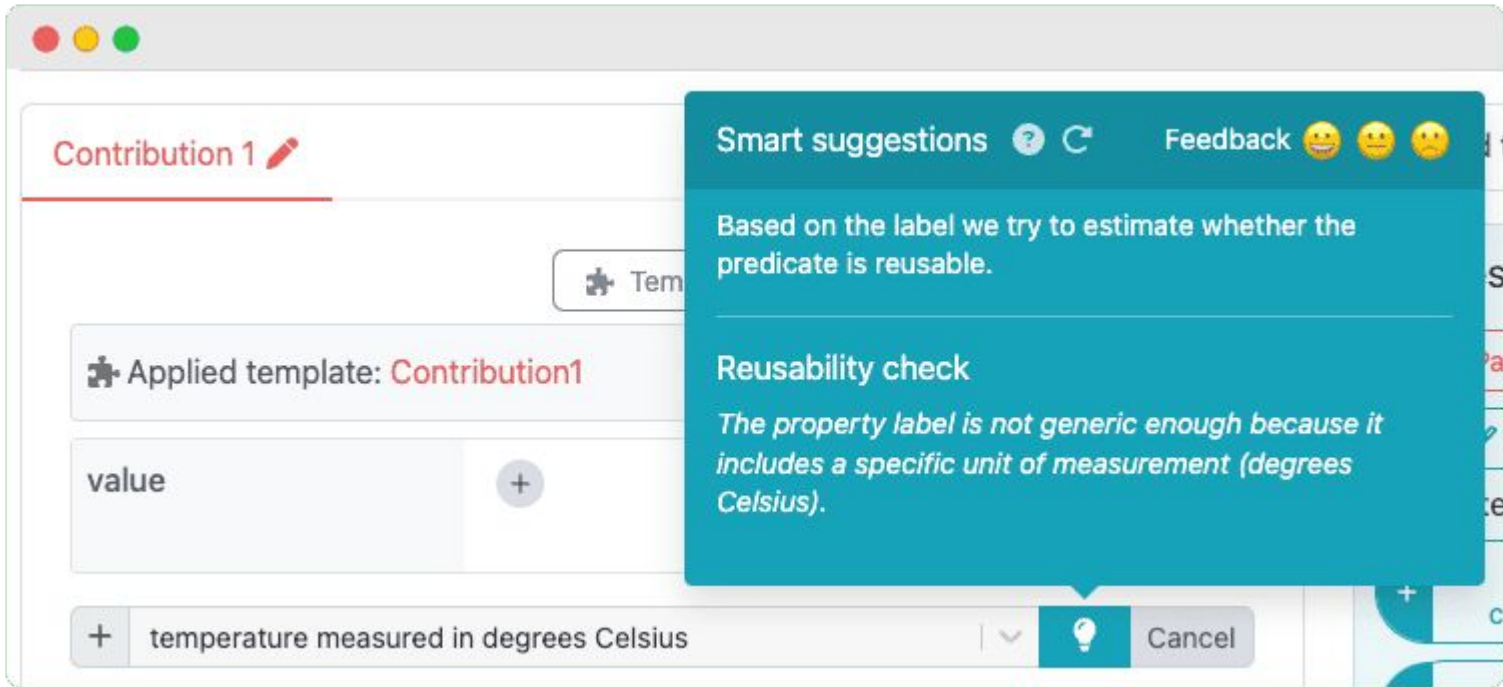
---

**Restructuring feedback**






*The resource label 'Convenience sampling of 100 participants from Spain' can be decomposed into separate resources: 'Sampling method: Convenience sampling' and 'Participants: 100' and 'Location: Spain'.*

At the bottom of the dialog, there is a lightbulb icon. Below the dialog, the main interface shows a search bar with the text "Resource" and a dropdown menu containing "Convenience sampling of 100 partici". To the right of the search bar are buttons for "Cancel" and "Create".

## Use cases: Feedback on predicate reusability



The screenshot shows a web interface with a teal feedback dialog box overlaid on top. The dialog box contains the following text:

Smart suggestions   Feedback   

Based on the label we try to estimate whether the predicate is reusable.

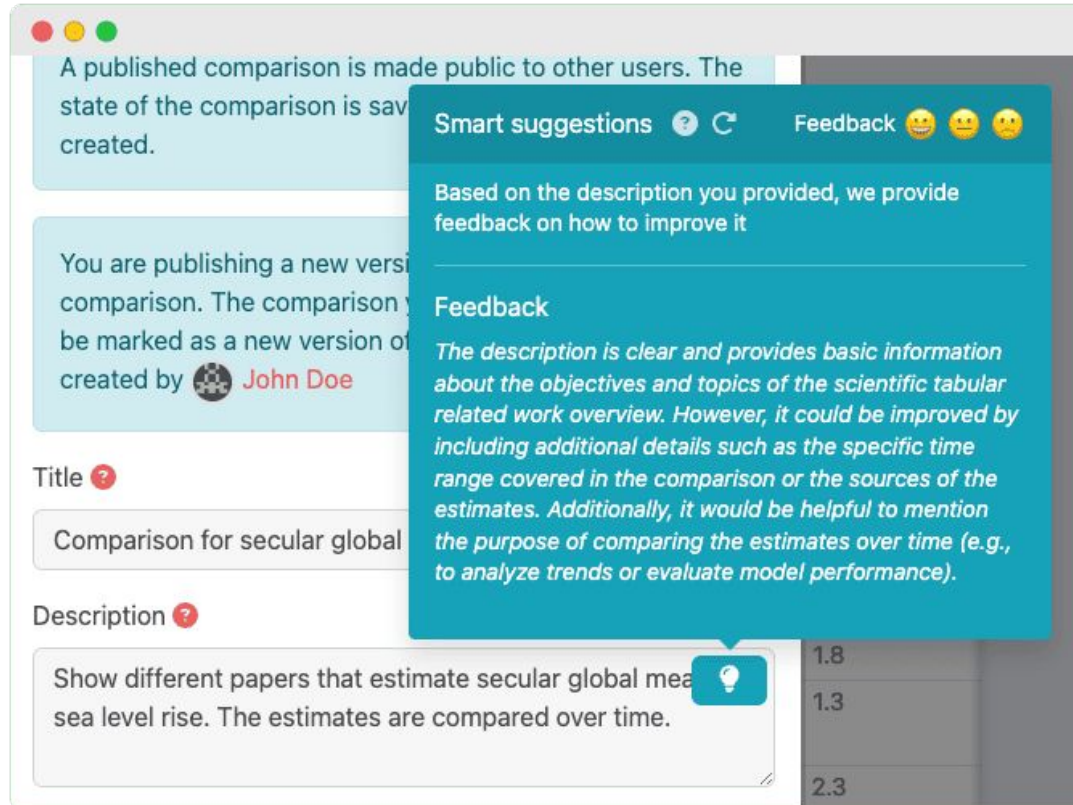
---

**Reusability check**


*The property label is not generic enough because it includes a specific unit of measurement (degrees Celsius).*


The background interface shows a form with a red header "Contribution 1" and a "value" input field. Below the input field, the text "temperature measured in degrees Celsius" is visible, along with a "Cancel" button and a lightbulb icon.

## Use cases: Feedback on comparison descriptions





A published comparison is made public to other users. The state of the comparison is saved and a new version is created.






You are publishing a new version of the comparison. The comparison will be marked as a new version of the comparison created by  John Doe

Title 

Comparison for secular global

Description 

Show different papers that estimate secular global mean sea level rise. The estimates are compared over time. 

Smart suggestions   Feedback   

Based on the description you provided, we provide feedback on how to improve it

**Feedback**

*The description is clear and provides basic information about the objectives and topics of the scientific tabular related work overview. However, it could be improved by including additional details such as the specific time range covered in the comparison or the sources of the estimates. Additionally, it would be helpful to mention the purpose of comparing the estimates over time (e.g., to analyze trends or evaluate model performance).*

1.8  
1.3  
2.3

# Thank you!

Visit the orkg at [orkg.org](https://orkg.org)

Kheir Eddine Farfar

[kheir.farfar@tib.eu](mailto:kheir.farfar@tib.eu)