Study Findings

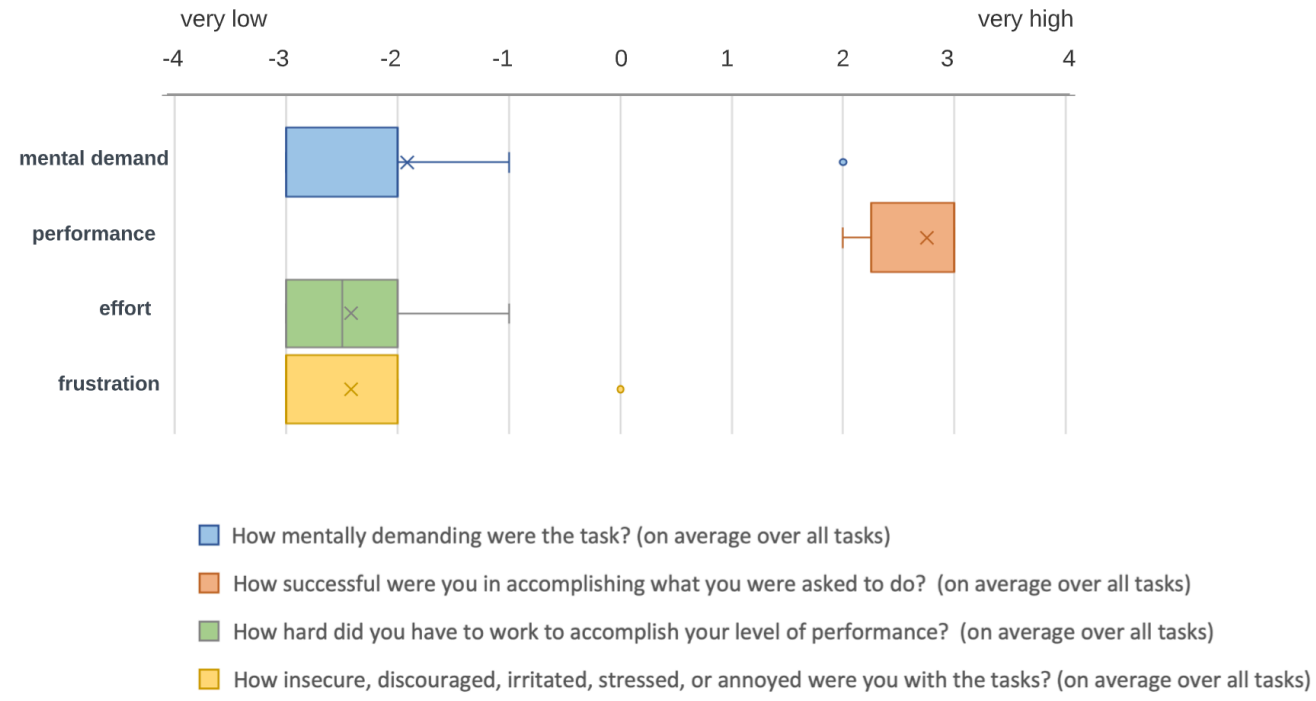
Part A: Task completion rates

	Task 1.1. Identify the top 2 articles with the highest overall similarity to the given article for which you are receiving recommendations (so the central node).	Task 1.2. Explore the recommendations. Then go ahead and save 2 documents that you, personally, find most interesting to your 'collected documents' panel for later reading.	Task 1.3. Identify the recommended article (or articles) containing the highest degree of only citation-based similarity.	Task 1.4. Identify the recommended article (or articles) containing the highest degree of only image-based similarity.	Task 1.5. Assume one of the documents saved in your 'collected documents' panel no longer interests you. Remove it from the collection.	Task 1.6. Collapse the side panels to give you more space to expand the recommendations.
P. ID	**Successful task completion? (1 = success, 0 = failure)**					
1	1	1	1	1	1	1
2	0	1	1	1	1	1
3	1	1	1	1	1	1
4	1	1	1	1	1	1
5	1	1	1	1	1	1
6	1	1	1	1	1	1
7	1	1	1	1	1	1
8	1	1	1	1	1	1
9	1	1	1	1	1	1
10	0	1	1	1	1	1
11	1	1	0	1	1	1
12	1	1	1	1	1	1
Successfull task completion rate per task	0.83	1.00	0.92	1.00	1.00	1.00
Over all Tasks	**0.958**					

Task 2.1. Identify the recommended article that has the highest overall similarity with the source document, and determine which feature contributes the most to overall similarity?	..and determine which feature contributes the most to overall similarity?	Task 2.2. Of the recommended article, identify the publication containing the highest degree of citation-based similarity.	Task 2.3. Of the recommended article identify the publication containing the highest degree of both text-based similarity and citation-based similarity.	Task 2.4. Configure the weighting of the similarity features to your own custom preference and save this weighting as a custom setting.	Successfull task completion rate per participant
1	Citation, image	1	1	0	0.90
1	Citation, image, formula	1	1	1	0.90
1	Formula and text	1	1	1	1.00
1	Formula	1	1	1	1.00
1	Image	1	1	1	1.00
1	Image	1	1	1	1.00
1	Formula	1	1	1	1.00
1	Text	1	1	1	1.00
1	Text	1	1	1	1.00
0	unsure	1	1	1	0.80
0	no answer	0	0	1	0.60
1	Text	1	1	1	1.00
				Over all participants	**0.93**
0.83		0.92	0.92	0.92	
0.896					

Part B: Task Workload Index (NASA-TLX)

Chart of user-perceived workload:



Data distribution table

Legend	The NASA-TLX survey uses a 21-point scale, which can be dived by three to scale down to 7-point scale, which is also commonly used in the literature. When entering our collected data, we scaled down the responses to fit a 7-point scale, i.e. represented as: -3, -2, -1, 0, 1, 2, 3	red = outliers	

	Mental Demand	Performance	Effort	Frustration
P. ID	How mentally demanding were the task? (on average over all tasks)	How successful were you in accomplishing what you were asked to do? (on average over all tasks)	How hard did you have to work to accomplish your level of performance? (on average over all tasks)	How insecure, discouraged, irritated, stressed, or annoyed were you with the tasks? (on average over all tasks)
1	-3	3	-3	-3
2	-2	3	-3	-3
3	-1	3	-2	-3
4	-2	3	-2	-3
5	-2	2	-3	-2
6	-3	3	-3	-3
7	-3	3	-3	-3
8	-2	3	-1	-2
9	-2	2	-2	-2
10	-2	2	-2	0
11	2	3	-3	-3
12	-3	3	-2	-2
Median	-2	3	-2.5	-3
Averages	**-1.9**	**2.8**	**-2.4**	**-2.4**

Part C: Task Workload Index (NASA-TLX)

1. The graph-based visualization gives me a useful overview of the recommended literature.

2. The graph-based view lets me quickly narrow down the recommendation results to show the articles that might be most relevant or interesting to me.

3. The visualization of document similarity using the blue color scale helps me to quickly distinguish between recommended documents with a high- vs. low-ranking similarity score.

4. The color encoding for the 4 similarity features (text, citation, image, formula) is easy to tell apart.

5. The relative size of the expanded 'feature nodes' (which indicate the presence of text, citation, image, or formula-based feature similarity) help me understand the influence of individual semantic features.

6. The expanded view for the individual recommendations can help me detect the presence of content that would otherwise be hard to identify between the source document and the recommended documents.

7. Adjusting the global similarity threshold allows an efficient removal of the documents that do not interest me.

8. Adjusting the feature-based similarity thresholds allows me to focus on properties that might interest me in the recommended literature.

9. The 'collected documents' panel provides a useful overview of documents to be saved for further examination and reading.