

FREIE UNIVERSITÄT BERLIN

The Impact of Concept Representation in
Interactive Concept Validation (ICV)

Maximilian Mackeprang, Abderrahmane Khiat,
Maximilian Stauss, Tjark Sascha Müller, Claudia
Müller-Birn

TR-B-19-03
July, 2019



**FACHBEREICH MATHEMATIK UND INFORMATIK
SERIE B • INFORMATIK**

Abstract

Large scale ideation has developed as a promising new way of obtaining large numbers of highly diverse ideas for a given challenge. However, due to the scale of these challenges, algorithmic support based on a computational understanding of the ideas is a crucial component in these systems. One promising solution is the use of knowledge graphs to provide meaning. A significant obstacle lies in word-sense disambiguation, which cannot be solved by automatic approaches. In previous work, we introduce *Interactive Concept Validation* (ICV) as an approach that enables ideators to disambiguate terms used in their ideas. To test the impact of different ways of representing concepts (should we show images of concepts, or only explanatory texts), we conducted experiments comparing three representations. The results show that while the impact on ideation metrics was marginal, time/click effort was lowest in the images only condition, while data quality was highest in the both condition. **Keywords:** Crowd Ideation, Semantic Annotation, Human-Computer Interaction (HCI).

1 Introduction

Contemporary challenges in society, technology, science and nature can only be tackled by leveraging diverse, interdisciplinary sources. In business settings, collaborative ideation platforms (such as Quirky (www.quirky.com) and OpenIDEO (www.openideo.com) have emerged as a promising solution to collect creative ideas from the crowd. However, new challenges are introduced by this approach: The submitted ideas are often simple, mundane and repetitive [7, 8]. Furthermore, with a large number of ideas collected it becomes economically unfeasible to manually review all of them in detail and to identify the most valuable ones. These factors make computational support of large scale ideation a critical goal, leading to an active field of research in the realm of computational analysis of ideas, ideators and overall efforts in a large scale ideation setting. One way to transform the submitted ideas into a computationally understandable representation is the linking between words used in the idea texts and resources in so called *general knowledge graphs*. A *knowledge graph* organizes various topically different real-world entities (concepts) with their relations in a graph. It provides a schema that aggregates these real-world entities into classes (abstract concepts) that also have relationships to each other [6]. In other words, using a knowledge graph provides computational meaning to concepts.

The existing relationships between concepts in a knowledge graph allow us to identify similar ideas, even though they have no terms in common by revealing a relationship on a more abstract level, for example between abstract concepts. An idea describing modification of a door and another describing wall painting, for example, could be connected by the concept of *architecture*. Thus, as opposed to statistical methods that are based on explicit relationships, knowledge graphs allow for identifying implicit, more subtle relationships.

Enabling the annotation of concepts in idea texts would have multiple advantages. The concepts can be used, for example, as an input for a faceted search tool (e.g. finding all ideas that talk about kinds of plants), they enable aggregation and cluster visualization of the solution space [11] and they could potentially enable effective inspiration mechanisms such as *analogical transfer* [2] (via a ‘has-usage’ relationship between concepts).

2 Context: Interactive Concept Validation (ICV)

This previously described potential of annotating ideas motivated the development of our ‘Interactive Concept Validation’ (ICV) technique, which can be integrated into traditional ideation processes [5]. After submitting an idea, a person is asked to annotate the idea manually based on concepts obtained from a knowledge graph. This tool consists of an interactive component since we ask a person for their annotations and an intelligent component because the provision of the annotations uses AI technologies.

We used an ideation software developed by us as the context. Figure 1 shows the interface. It consists of five panels shown with a capital letter. The problem description, for which a person is asked to submit ideas, is the panel (A). In the panel (B), the ideator can type and submit her ideas using the button [Submit]. After submission, panel (C) shows the concepts found for each term in one of the list formats. In the bottom row, the panel (D) shows a button [Need_Inspiration] which allows a person to request inspirations. If a user clicks the button, the panel expands and shows three randomly selected ideas from previously submitted ideas. Panel (E) presents an overview of the user’s proposed ideas.

The submission of an idea by a person triggers the actual information extraction process. To realize term extraction and concept search, we employ an automatic search approach in the DBpedia knowledge graph ¹.

In the case of overlapping concept term combinations (example: Pet Food Distribution Center), we employ a greedy algorithm to expand the term in the idea to the longest continuous text. We furthermore save the original terms for each concept, to preserve the overlapping information. This enables multiple annotations of a single term. The search for concepts results in a link between a term (or set of terms for overlapping results) that are potentially linked to a list of concepts (this is called *Annotation Candidate*). If the search for a term doesn’t result in any results, the *Annotation Candidate* is discarded.

¹To obtain a ranked list of terms and concepts, we use the candidate search API of DBpedia Spotlight (<http://api.dbpedia-spotlight.org/en/candidates>)

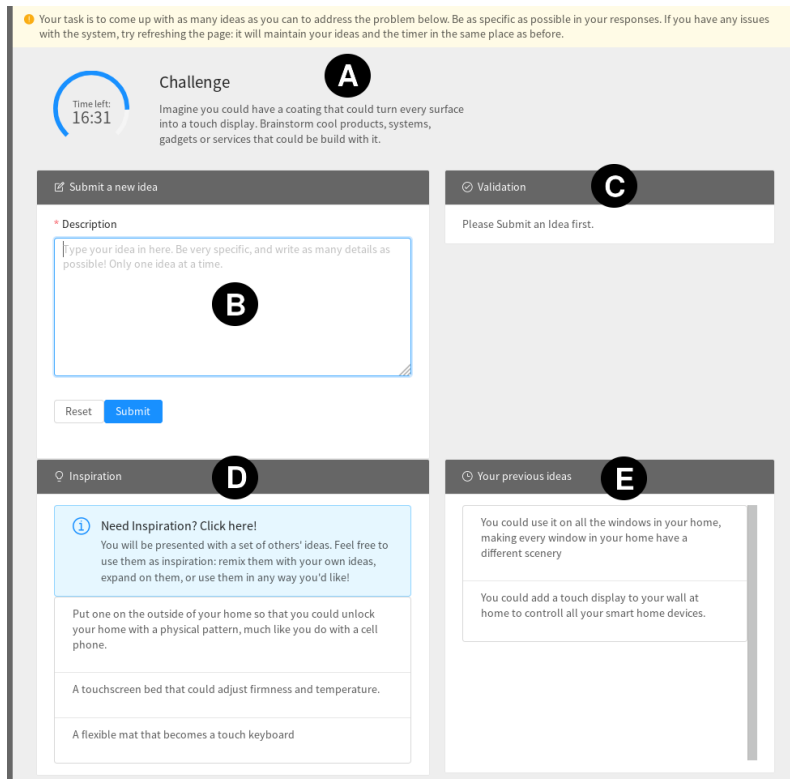


Figure 1: Baseline Ideation Interface employed in all conditions. The interface features panels for the problem description (panel (A)), Idea Text Submit (B), concept list (C), inspirations (D) and a history of submitted ideas (E).

3 Study: Understanding the impact of concept representation

A crucial design decision in ICV is the representation of the concepts during the validation task. We wanted the task to distract as little as possible but still provide as much information about the concept as needed. In this study, therefore, we evaluated the impact of different concept representations during the disambiguation task. We determined the influence of each of these representations on the ideation outcome, the quality of the concepts chosen and the user's satisfaction with the design. We hypothesized that the visual representation of concepts by images helps humans to disambiguate terms, motivated by the *picture superiority effect*. This phenomenon describes the recurrent observation that pictures are better stored in and faster retrieved from human memory than words [10, 1]. Another motivation for the use of images is that showing a diverse set of images to the ideators could increase the diversity of ideas by serving as

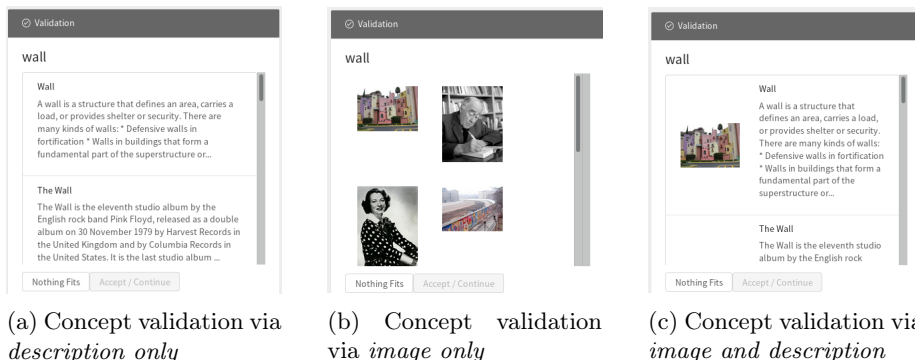


Figure 2: Ideation interfaces across different alternatives for concept validation.

unintended inspiration.

We designed three possible concept representations to evaluate our hypothesis. In this study, we used the same interface and workflow as introduced in the previous section. However, when users clicked the [Submit] button, the representation of concepts for validation is varied between the conditions. In the condition *description only*, we present each concept by its description (cf. Figure 2a).

In the *image only* condition, we show a grid of images (one per concept; cf. Figure (one per concept; cf. Figure 2b) and, in the condition *image with description*, we provide at least one image and description for each concept (cf. Figure 2c).

We extracted the description and image for each concept from the knowledge graph. Given the concept "Touchscreen" on DBpedia², for example, we can obtain the following description: "A touchscreen is an important source [sic!] of input device and output device normally layered on the top of an electronic visual display of an information processing system [...]". For the *image only* condition, we excluded all concepts where the picture was missing. For all conditions and for each extracted term, the users selected the most suitable concept from the list of alternatives or rejected all concepts shown via the [Nothing_Fits] button. Once the users had validated all suitable concepts, they clicked on [Submit].

3.0.1 Task and Participants

We provided the following ideation task for all studies: "Imagine you could have a coating that could turn every surface into a touch display. Brainstorm cool products, systems, gadgets or services that could be built with it." We chose this task because it is both unspecific enough to enable workers from various backgrounds to contribute and, at the same time, resembles a real-world challenge

²Please check out the concept URL for further information <http://dbpedia.org/page/Touchscreen>.

from the innovation area.

We recruited participants on Amazon Mechanical Turk (Mturk) for both studies. We limited recruitment to US residents who had completed at least 1,000 HITS with greater than 95% approval rate. We set the compensation for all tasks so that workers received a payment rate of 12\$/h.

We conducted a pre-study to collect an initial set of ideas for use as inspirations and to build an LSA corpus. We recruited 100 Mturk workers for this task. The overall task consisted of a short text introduction, followed by an ideation session of ten minutes. At the end of the study, participants were asked to provide demographics, i.e. age group, gender and, optionally, qualitative feedback. We received a total of 761 ideas.

For the evaluation of the different concept representations (cf. section 3), we recruited participants from Mturk as well. We used a between-subjects study design and recruited 50 participants for each condition (150 in total). Participants were paid 5 \$ for a session in one condition. In each condition, we used the ideation interface as shown in Figure 1 and one of the list presentations as shown in Figure 2.

The task consisted of a short tutorial on concept validation followed by an ideation session of twenty minutes. After the ideation session, participants were asked to provide selected demographic data (e.g., age group, gender), fill out a questionnaire) and, optionally, provide qualitative feedback.

3.0.2 Metrics

The overarching goal of the ICV is to provide precisely annotated ideas that can be used for supporting the collaborative ideation process in later stages, for example by providing suitable creativity enhancing interventions. To further investigate this task, we defined evaluation metrics, by looking at the outcome of our conditions from either a *human* or a *system perspective*.

Human Perspective We want to ensure that the ICV is as non-invasive as possible regarding the ideation process. We can capture the effort by collecting data on human activities or asking a person to rate their perceived effort. Regarding data collection we can measure the *interaction effort* for a person by instrumenting the software with a tracking system for recording the number of *clicks* and the *time* needed to perform the ICV.

To capture mental effort, a widely used approach is the *Nasa Task Load Index (TLX)* questionnaire [4]. We employed the ‘Raw TLX’ by asking an ideator for their perceived mental and temporal demand, combined with their self-assessment of performance, effort and frustration during the ICV task³.

Furthermore, the influence of the ICV on the human ideation processes (their creativity) needs to be captured as well. In collaborative ideation, the operationalization of creativity is most often based on two metrics: fluency [9] and

³As the TLX physical demand scale refers to physical activities (pushing, pulling, etc), we excluded it from our evaluation.

dissimilarity of ideas [3]. Fluency describes the number of ideas produced by a participant. The *diversity* of ideas generated is operationalized by the *depth* and *breadth* of the set of ideas generated by each participant. We constructed an idea tree based on LSA similarity between ideas to compute the *depth* and *breadth*. Following Giroto et al. [3], we classified ideas with a high similarity into branches of a tree⁴. *Breadth* relates to the number of children in the root node of the idea tree and *depth* is the number of nodes in the branch with the most nodes in the idea tree. We used the corpus of 761 ideas from the pre-study to determine the LSA similarity of ideas.

System Perspective From an algorithmic perspective, we need to measure the quality of the information extraction task, i.e. the suitability of the annotated concepts in describing the terms. We measured data quality by manually annotating a randomly chosen subset from all ideas, i.e., 20 ideas per condition. Two of the authors extracted and validated concepts manually by first annotating 10 ideas together, then annotating the remaining 50 ideas separately. Afterwards, the authors compared their annotation results and resolved conflicting annotations together. The manually annotated subset was used as gold standard and allowed us to calculate precision, recall, and F-measure. Whereas *precision* defines the number of concepts correctly annotated, *recall* defines the number of correctly disambiguated concepts relative to the number of all concepts found. The *F-measure* is the harmonic mean of precision and recall.

3.0.3 Expected Results

Our evaluation of different concept representation in the ICV task was guided by multiple assumptions about their impact. Based on the picture superiority effect we hypothesized that including images in the concept representation, i.e., the list, could lead to faster annotation times (H_1). We expect ideators that have access to images to be able to decide more quickly if the concept reflects their intended meaning of the term. However, the concepts obtained from DBpedia sometimes lack an image. The number of possible concept candidates is, therefore, lower for the image only condition. This consideration led to the other hypothesis (H_2) that data quality, in terms of the F-measure, is lower for the image only condition.

3.1 Results

We received a total of 1,006 ideas from 150 participants with 6,703 annotated terms. On average participants created 7,25 ideas in their session (across all conditions).

Human Perspective The effort needed by each participant was dependent on the number of ideas submitted and consequently the number of terms that had

⁴A detailed description for constructing the idea tree and the algorithm used can be found in Giroto et al.[3].

to be annotated in them. Therefore, we normalized the results, and subsequently report the human effort in Clicks per Annotation (C/A), Time per Annotation (T/A) and TLX Score per Annotation (TLX/A).

To evaluate H_1 we conducted an ANOVA between the conditions (with time per annotation as the variable). The results show that there are significant differences between conditions ($F = 9.85$ ($p = 0.001$)). Tukey post-hoc analysis revealed a significant difference between Image/Both ($p = 0.001$) and Image/Description ($p = 0.01$).

Condition	Clicks/Annotation	Time/Annotation (s)	TLX/Annotation
description	2.32 (0.59)	8.93 (3.44)	18.9 (4.85)
image	2.13 (0.42)	6.56 (3.4)	19.11 (4.28)
both	2.18 (0.34)	10.38 (5.36)	19.4 (4.11)

Table 1: Evaluation results for the concept-representation conditions in terms of **human effort metrics**. All results are reported in mean (standard-deviation)

Condition	Fluency	Depth	Breadth
description	7.38 (2.91)	3.62 (2.38)	4.17 (2.42)
image	7.96 (3.51)	4.32 (2.83)	4.17 (2.29)
both	6.4 (2.81)	3.08 (1.82)	3.96 (2.37)

Table 2: Evaluation results for the concept-representation conditions in terms of **ideation metrics**. All results are reported in mean (standard-deviation)

System Perspective We compared the resulting ideator-based validated concepts from each condition with our expert-based validations. Table 3 (row 1-3) shows the *Precision*, *Recall* and *F-measure* for each condition.

To evaluate H_2 we conducted an ANOVA on F-measure for all conditions. The results showed significant differences ($F = 6.02$, $p = 0.004$), a Tukey HSD test showed significantly higher F-Measures for the Both condition, when compared to Image ($p = 0.007$) and Description ($p = 0.01$) conditions.

Condition	Precision	Recall	F-Measure
description	0.54 (0.3)	0.39 (0.32)	0.44 (0.3)
image	0.44 (0.29)	0.42 (0.28)	0.41 (0.25)
both	0.71 (0.23)	0.65 (0.21)	0.67 (0.2)

Table 3: Evaluation results for the concept-representation conditions in terms of **data quality metrics**. All results are reported in mean (standard-deviation)

4 Discussion

The results from the three conditions helped us to make an informed decision of concept representation in our domain context. There are two central insights: first, showing images reduces the time a person needs to annotate images. Second, representing concepts with images and descriptions impacts data quality positively. We discuss these insights in detail next.

Showing images only reduces annotation time. The representation of concepts was an important design decision for the concept validation task. Motivated by the *picture superiority effect*, we hypothesized that including images would be beneficial for the concept recognition time. By conducting an ANOVA between conditions we could see that the time per concept was significantly lower for the *image only* condition. However, including both image and description didn't have the expected effect of lowering annotation time when compared to the description only condition. We could explain this result by the fact that participants read texts as soon as they were available to them, eliminating the effect. Another possible explanation would be that individual differences both in annotation difficulty and personal speed outweigh the impact of concept representation. Future work could further investigate the effects of images on concept extraction tasks, by fixing the input data set and then compare user performance for same input annotations.

Showing images with descriptions impact data quality positively. The *Both* condition significantly outperformed the other two conditions in terms of data quality. When compared to the image category, the effect could be due to two causes: When compared to the Image condition, the lower data quality could be due to a reduced concept set for the image only condition. This could prevent selection of the correct concepts, as they don't have an image. When compared to the Description condition, we hypothesize that by additionally showing an image to users, we increase both task motivation and task understanding (by invoking an intuitive mental model of a 'concept' within participants). As participants develop a better understanding of what a concept looks like, data quality could be positively impacted.

5 Conclusion and Future Work

In this research, we investigated the impact of concept representation in the context of interactive concept validation. Based on our use case in large scale ideation, we started out on the assumption, that including images in the concept representation could potentially decrease annotation time due to the *picture superiority effect* and inspire ideators by providing visual stimuli. However, due to the limited availability of images in the used knowledge-graph, we expected, that providing images alone could lead to negative impacts on data quality results. To test our assumptions, we conducted an in between study, with *image*,

description and *both* as the concept representation in our software prototype. The results of the study suggest that using both images and description as a representation of a concept leads to higher data quality results. However, human effort in terms of time spend per concept was higher than in the other two tested conditions. This shows the inherent trade-off between time spent to evaluate and disambiguate each concept and quality of disambiguation when compared to a gold standard. Unfortunately, no significant results were found concerning the impact of concept representation on ideation metrics. Future work could further explore the possibilities to combine tasks needed for the algorithmic system with potential creativity enhancing influences.

Acknowledgements

The authors acknowledge the financial support of the Federal Ministry of Education and Research of Germany in the framework of Ideas to Market (project number 03IO1617).

References

- [1] Zaira Cattaneo, Albert Postma, and Tomaso Vecchi. The picture superiority effect in working memory for spatial and temporal order. *Psychologia*, 50:102–109, 2007. *Psychologia*.
- [2] Karni Gilon, Joel Chan, Felicia Y. Ng, Hila Liifshitz-Assaf, Aniket Kittur, and Dafna Shahaf. Analogy mining for specific design needs. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 121:1–121:11, New York, NY, USA, 2018. ACM.
- [3] Victor Giroto, Erin Walker, and Winslow Burleson. The effect of peripheral micro-tasks on crowd ideation. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 1843–1854, New York, NY, USA, 2017. ACM.
- [4] Sandra G Hart and Lowell E Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier, 1988.
- [5] Maximilian Mackeprang, Abderrahmane Khiat, and Claudia Müller-Birn. Concept validation during collaborative ideation and its effect on ideation outcome. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA '18, pages LBW033:1–LBW033:6, New York, NY, USA, 2018. ACM.
- [6] Heiko Paulheim. Knowledge graph refinement : A survey of approaches and evaluation methods. *Semantic Web*, 8(3):489–508, 2017.
- [7] Pao Siangliulue, Kenneth C. Arnold, Krzysztof Z. Gajos, and Steven P. Dow. Toward collaborative ideation at scale: Leveraging ideas from others


- to generate more creative and diverse ideas. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '15, pages 937–945, New York, NY, USA, 2015. ACM.
- [8] Pao Siangliulue, Joel Chan, Steven P. Dow, and Krzysztof Z. Gajos. Ideahound: Improving large-scale collaborative ideation with crowd-powered real-time semantic modeling. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST '16, pages 609–624, New York, NY, USA, 2016. ACM.
- [9] Pao Siangliulue, Joel Chan, Krzysztof Z. Gajos, and Steven P. Dow. Providing timely examples improves the quantity and quality of generated ideas. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*, C&C '15, pages 83–92, New York, NY, USA, 2015. ACM.
- [10] Lionel Standing. Learning 10000 pictures. *Quarterly Journal of Experimental Psychology*, 25:207–222, 1973.
- [11] Donghee Yoo, Keunho Choi, Hanjun Lee, and Yongmoo Suh. An ontology-based co-creation enhancing system for idea recommendation in an online community. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, 46(3):9–22, 2015.

A Screenshots of the Annotation-Interface in the different conditions

The screenshot displays a web-based interface for a challenge. At the top, a yellow banner contains instructions: "Your task is to come up with as many ideas as you can to address the problem below. Be as specific as possible in your responses. If you have any issues with the system, try refreshing the page: it will maintain your ideas and the timer in the same place as before." Below this, the "Challenge" section features a circular timer showing "Time left: 19:21" and the text: "Imagine you could have a coating that could turn every surface into a touch display. Brainstorm cool products, systems, gadgets or services that could be built with it." The interface is divided into four main panels. The "Submit a new idea" panel shows a text input field with the sentence "You can turn your bedroom wall into a touchscreen TV" where "turn", "bedroom", and "wall" are highlighted in colored boxes. Below the input are "Reset" and "Submit" buttons. The "Validation" panel shows the word "wall" in a search bar, with two search results: "Wall" (a general definition) and "The Wall" (a Pink Floyd album). Below the results are "Nothing Fits" and "Accept / Continue" buttons. The "Inspiration" panel contains a blue box with an information icon and the text: "Need Inspiration? Click here! You will be presented with a set of others' ideas. Feel free to use them as inspiration: remix them with your own ideas, expand on them, or use them in any way you'd like!". The "Your previous ideas" panel shows "No data".

Figure 3: Concept validation via *description only*

● Your task is to come up with as many ideas as you can to address the problem below. Be as specific as possible in your responses. If you have any issues with the system, try refreshing the page: it will maintain your ideas and the timer in the same place as before.



Time left:
18:52

Challenge

Imagine you could have a coating that could turn every surface into a touch display. Brainstorm cool products, systems, gadgets or services that could be build with it.





Submit a new idea

You can turn your bedroom wall into a touchscreen TV

Reset
Submit

Validation

wall

Nothing Fits
Accept / Continue

Inspiration

i **Need Inspiration? Click here!**


You will be presented with a set of others' ideas. Feel free to use them as inspiration: remix them with your own ideas, expand on them, or use them in any way you'd like!

Your previous ideas

No data

Figure 4: Concept validation via *image only*

🔔 Your task is to come up with as many ideas as you can to address the problem below. Be as specific as possible in your responses. If you have any issues with the system, try refreshing the page: it will maintain your ideas and the timer in the same place as before.



Time left:
15:26

Challenge

Imagine you could have a coating that could turn every surface into a touch display. Brainstorm cool products, systems, gadgets or services that could be build with it.


✍️ Submit a new idea

You can turn your bedroom wall into a touchscreen TV

Reset
Submit

🔍 Validation

wall



Wall
A wall is a structure that defines an area, carries a load, or provides shelter or security. There are many kinds of walls:
* Defensive walls in fortification
* Walls in buildings that form a fundamental part of the superstructure or...

The Wall
The Wall is the eleventh studio album by the English rock

Nothing Fits
Accept / Continue

🔍 Inspiration

📘 Need Inspiration? Click here!
You will be presented with a set of others' ideas. Feel free to use them as inspiration: remix them with your own ideas, expand on them, or use them in any way you'd like!

Put one on the outside of your home so that you could unlock your home with a physical pattern, much like you do with a cell phone.

A touchscreen bed that could adjust firmness and temperature.

A flexible mat that becomes a touch keyboard

🕒 Your previous ideas

You could use it on all the windows in your home, making every window in your home have a different scenery

You could add a touch display to your wall at home to control all your smart home devices.

Figure 5: Concept validation via *image with description*

B Summarized Metrics for the Evaluation

	Ideation Metrics			Effort			Data Quality		
	Fluency	Depth	Breadth	C/A	T/A (s)	TLX/A	precision	recall	F-measure
description	7.38 (2.91)	3.62 (2.38)	4.17 (2.42)	2.32 (0.59)	8.93 (3.44)	18.9 (4.85)	0.54 (0.3)	0.39 (0.32)	0.44 (0.3)
image	7.96 (3.51)	4.32 (2.83)	4.17 (2.29)	2.13 (0.42)	6.56 (3.4)	19.11 (4.28)	0.44 (0.29)	0.42 (0.28)	0.41 (0.25)
both	6.4 (2.81)	3.08 (1.82)	3.96 (2.37)	2.18 (0.34)	10.38 (5.36)	19.4 (4.11)	0.71 (0.23)	0.65 (0.21)	0.67 (0.2)

Table 4: Evaluation results for the concept-representation conditions (C/A=Clicks per Annotation, T/A = Time per Annotation(seconds), TLX/A = Task Load SUM per Annotation). All results are reported in mean (standard-deviation). The conditions are concept representation as a list of images (image condition), as a list of description (description condition) or as a list images and descriptions (both condition)