

DESIGNING A USER INTERFACE FOR EXPLORING RELATIONSHIPS BETWEEN SEMANTICALLY SIMILAR BRAIN DISEASES

SUBMITTED IN PARTIAL FULFILLMENT FOR THE DEGREE OF MASTER OF SCIENCE

FREEK BUITER
10742441

MASTER INFORMATION STUDIES
DATA SCIENCE
FACULTY OF SCIENCE
UNIVERSITY OF AMSTERDAM

2021-04-23

	Internal Supervisor	External Supervisor
Title, Name	Dr Frank Nack	Prof Dr Lynda Hardman
Affiliation	UvA	CWI & UU
Email	f.m.nack@uva.nl	lynda.hardman@cwi.nl



UNIVERSITEIT VAN AMSTERDAM

Designing a user interface for exploring relationships between semantically similar brain diseases

Freek Buiter

freek.buiter@student.uva.nl

Universiteit van Amsterdam

Amsterdam, Nederland

ABSTRACT

Relationship finding is a method that allows researchers to find relationships between concepts in literature. However, the amount of research that is currently available in neuroscience is too much for a single person to review. This poses a problem when one is attempting to find relationships between two concepts which can lead to overlooking crucial relationships. The DatAR project assists researchers in relationship finding between concepts in the neuroscience domain by visualizing linked data in an Augmented Reality (AR) environment. This visualization contains a 3D topic model of brain diseases where the distance between two diseases represents the semantic similarity. In this study, we have designed a visualization that lets neuroscientists explore the relationship between semantically similar brain diseases. We used an iterative design process consisting of three studies. In the first study, we created a user scenario on which we based our first design. After which, we conducted interviews with neuroscientists to find out what information is relevant for users. In the second study, we made two new designs based on the results of the first study. Afterwards, we conducted interviews with design experts to find out what an appropriate way is to visualize the information. The results from these two studies constructed our design requirements. In the last study, we created a final design that satisfies the design requirements and held a usability evaluation with neuroscientists. The results of this evaluation showed that the visualization had few usability problems and users gave the design an average score of 85 on the System Usability Scale. We, therefore, see opportunities to implement this design in AR. We also discuss several fundamental issues with our final design and the implications of designing in 2D for a 3D visualization.

1 INTRODUCTION

Nowadays, the information that is made available to researchers is too much for a single human being to take in. Therefore, it is important to carefully select what research problem to pursue and narrow the scope early on.

A method that can help researchers in selecting a research problem is *relationship finding*. Relationship finding is a method used by a researcher to find connecting factors to a concept of interest. For example, a concept of interest could be the disease *Depression* and the researcher is interested in all brain regions that are related to *Depression*. The researcher has to manually review the literature in search of these relationships. Although relationships can be found, knowing which brain region is most commonly related or which one is very rarely related to *Depression* is difficult. The main problems are that manual reviewing of literature is first of all extremely time-consuming and secondly, researchers are apt to stay in their

own expertise, therefore crucial relationships between concepts may be missed by accident [5].

Relationship finding and reviewing literature can be made easier for researchers by using automated tools. A technology that can be used for the purpose of relationship finding is *linked data*. Linked data is a data structure that consists of concepts and relationships between these concepts. If we take the example of *Depression* again, previously it was difficult to find which brain regions have a relationship with *Depression*. With linked data, it is possible to show all brain regions that have a relationship with *Depression* and where these relationships can be found.

In this study, we make use of a linked data repository called Linked Brain Data¹ (LBD). The LBD was created to assist neuroscientists in finding relevant literature. This repository was formed by colleagues at the Chinese Academy of Sciences, Institute of Automation, who scanned neuroscience literature. The repository consists of different brain concepts (e.g. diseases, brain regions, genes, proteins) and their relationships. A relationship is formed when two concepts are named in the same sentence in a paper's abstract: this is termed a *single co-occurrence*.

The LBD gives an overview of relationships between concepts, where these relationships can be found and also how often they are mentioned in the literature. This makes it easier for a neuroscientist to find relevant relationships.

While the LBD supports neuroscientists in getting an overview of the relationships between concepts, the discovery of new meaningful relationships is still a highly user-centered process [7]. The LBD doesn't have an application that lets users roam the relationships in a visual manner. Therefore it is important that, next to the linked data repository, there also is an application that assists neuroscientists in discovering these relationships. Even though linked data expresses many relationships, a user is needed to interpret which relationships are meaningful. Therefore interactive mechanisms are needed to assist neuroscientists in exploring these relationships.

The DatAR project is an application that visualizes the LBD repository in order to assist neuroscientists in discovering relationships [15]. The DatAR project consists of a 3D environment that supports the exploration of relationships between brain regions and brain diseases. These concepts form the basis for the two main elements of DatAR's interface:

- A brain region visualisation; here brain regions can be selected for investigation (see Figure 1)
- A topic model visualisation; here brain diseases can be selected for investigation (see Figure 2)

¹<http://www.linked-brain-data.org/>

[illegible]

2

contains three main studies (or design cycles) wherein every study one or more designs were created, evaluated and updated. With the results of these three studies, we answered the three subquestions in the introduction.

To answer the first subquestion: "Which information, present in the LBD, is relevant for a neuroscientist to help understand the relationship between two diseases in the DatAR 3D topic model?". We created a user scenario and conducted interviews with neuroscientists.

The user scenario was created with the information that is available in the LBD and based on the format of Kujala et al. [8]. Then we identified how a user would possibly interact with a system with this information. With the user scenario, we created a design which is called the information design. The focus of the information design was not on the visual design aspects but on the information and functionality we believed to be important.

We then set out to interview neuroscientists, to test if our assumptions based on the user scenario on what is relevant information to a user was correct.

To answer the second subquestion: "What is an appropriate way to visualize this information to the user?". We created two new designs and conducted interviews with design experts. The first design we explain is called the improved information design. The improved information design, as the name suggests, is an improvement of the information design. It was updated to include new functionality and a more refined visual style but to keep the main elements of the information design. The 3D-centered design is a design we created with the intention of making it more suitable for 3D, and therefore for the DatAR project as well.

Then we invited design experts to give their opinions about the two designs and see if this was an appropriate way to visualize the relevant information to the user.

During the first two design cycles, we collected design requirements that helped us shape the designs in every phase of the iterative process. In the Summary of Design requirements, we reflect on how we found these design requirements and what their purpose is.

Whereas in the first two design cycles, we focused respectively on the information a user wants to see and an appropriate way to visualize this information. The third design cycle focuses on the usability of the proposed final design and how intuitive this design is to new users. The proposed final design was created while trying to satisfy all design requirements.

To answer if the proposed final design is usable and clear to a neuroscientist, we conducted interviews with users which consisted of two parts: firstly, we showed the proposed final design to the users and asked them to perform tasks with it. We timed when each task was completed and asked them to think-aloud to see where the major usability problems were in the final design. Secondly, we asked them to fill in a System Usability Scale (SUS), which is a survey consisting out of 10 questions to measure the usability of a product [2].

With all the information from the three studies, we answered the main research question.

4 INTERVIEW NEUROSCIENTISTS

4.1 User tasks & user scenario

To answer the first subquestion, we first created a user scenario with complementary user tasks.

The first task was created based on the closed discovery perspective. By showing missing links to users, we hope to support them with finding relationships between semantically similar diseases and therefore furthering their research.

- Support neuroscientists in finding relations/missing links between semantically similar brain diseases.

The second task was created with to identify in which phase we are trying to assist researchers. This task has us focusing on a design that has a high ease-of-use and useful information shown at the right stage of research.

- Exploring semantically similar brain diseases at an early research stage.

Based on these tasks and the information available in the LBD, a user scenario was created. The purpose of this user scenario was twofold: firstly, we investigated how a user would possibly interact with a system for finding relationships between semantically similar brain diseases. Secondly, how users would get relevant information from the visualization to further their research.

The user tasks and user scenario were taken as inspiration to specify the following initial design requirements. Users should be able to:

- visually identify concepts that are related to both selected diseases.
- visually identify concepts that are related to only one of the two selected diseases.
- tell the number of co-occurrences for all relationships.
- visualize initially hidden relationships between a concept and the concepts it relates to on request
- investigate which articles a connection between concepts are based upon.

These design requirements were the basis for the creation of the information design that we will describe in the following section.

4.2 Information design

The information design was based on the initial design requirements and inspired by the literature in the related work.

We took an example scenario, where the user chose the brain diseases *Depression* and *Anxiety*. These diseases are close to each other in the 3D topic model, so they are semantically similar and the user would like to find out what causes this semantic similarity.

In figure 5, we see the selected diseases *Depression* and *Anxiety* in the black boxes. To fulfil the design requirement of visually indicating concepts that are related to both diseases, we look in the middle of the two diseases and see 12 colour coded concepts. These concepts are related to both diseases and the colours indicate to which brain-building group they belong: Red boxes are brain regions, yellow boxes are neurons, green boxes are proteins, blue boxes are genes, purple boxes are neurotransmitters and pink boxes are cognitive functions.

The second design requirement was visualizing concepts that only have a relation with one of both diseases (henceforth called

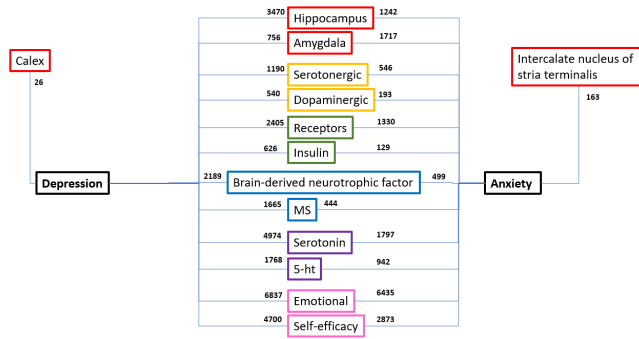


Figure 5: The information design consists of 7 different brain-building groups. Red boxes are brain regions, yellow boxes are neurons, green boxes are proteins, blue boxes are genes, purple boxes are neurotransmitters and pink boxes are cognitive functions. The numbers indicate the number of co-occurrences a link has. In the middle of the two black boxes are concepts in common. On the left and right side of the black boxes are discriminatory concepts. The number of co-occurrences of a relationship is shown on the left or right side of a concept

discriminatory concepts), we can see the result on the left side of *Depression*, where the brain region *Calex* only has a relationship with *Depression*. The same goes for the concept on the right side of *Anxiety*.

The third design requirement states that the number of co-occurrences must be shown for all relationships. This requirement was added to give the user an idea of the “strength” of the relationship by showing how often they were mentioned in the literature together. We can see the result on every line between one of the selected diseases and a concept. If we take, for example, the *Hippocampus*, we can see that it has been mentioned together with *Depression* 3470 times in the literature and the *Hippocampus* and *Anxiety* have been mentioned together 1242 times.

The fourth design requirement of showing initially hidden relationships between a concept and a concept it relates to can be seen in figure 6. This function allows the user to dive deeper into a concept of interest and see what kind of relationship this concept has with other brain-building groups. In figure 6, we see an example where the user is interested in the concept *Hippocampus* and selected the brain-building group Genes to show the additional relationships *Ca1 field* & *Brain-derived neurotrophic factor*.

The fifth design requirement of investigating which articles relationship is based upon is not included in the information design. This function already exists in DatAR and therefore we felt it was not necessary to implement it in this design.

4.3 Setup interview neuroscientists

To find out what information is relevant to a user and to test if the assumptions that we made with what relevant information is in the information design were right, we conducted interviews with participants who have a background in Neuroscience.

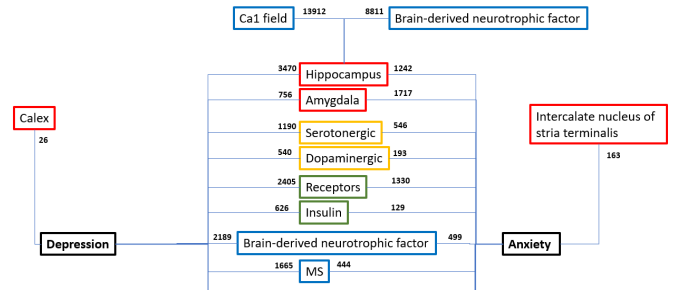


Figure 6: This is a function of the information design, where a concept can be selected and extended. In this example, we see that the *Hippocampus* has been selected and that the user would like to see additional genes which are connected to the *Hippocampus*.

Three participants, with a bachelor’s degree in Psychobiology, were invited for a one-hour semi-structured interview. The participants were encouraged to participate by offering them a 15 euro gift certificate or donating 15 euro to a charity of their choice. After giving a small introduction to the study, we started with the interview which consisted of three sections. Firstly, we asked how they explore literature related to a specific brain disease and the relationship between diseases. Secondly, we asked about their opinion about the information design. Lastly, we asked about the overall meaningfulness of the task and if they had any other comments.

4.4 Results interview neuroscientists

4.4.1 Results Brain disease literature exploration. After giving the introduction and asking some questions about general demographics, we asked the participants how they search for literature right now and if they ever researched the relationship between diseases. All participants used Google Scholar, PubMed or Web of Science and the most used strategy was typing in keywords and going further from there. Some participants had researched the relationship between diseases before but didn’t use a different search strategy.

When asked what they found more interesting to research, diseases that are close in the 3D topic model or ones that are farther away, two participants found that diseases that are close together were more interesting because of underlying causes or treatments that could work for both diseases. One participant disagreed with that statement and said the following: “Well, that depends because I think sometimes diseases are having similar symptoms, but it doesn’t necessarily (mean that they) are linked to each other causally. So I don’t actually think per se that diseases that look similar are also interesting to research together. Yeah, I think, for example, Parkinson’s disease and dementia don’t really have the same symptoms but they co-occur often so that’s more interesting to research for me”. This is another approach of research we initially didn’t think about but would be very interesting as well.

4.4.2 Results Information design visualization. In the second part of the interview, we asked the participants about the information and functionality that was present in the information design. Firstly, we

asked what they liked and disliked about the information design, as can be seen in figure 5.

It was agreed on by all participants that the overview of which concepts were overlapping was easy to understand and that the colours, which discriminate the brain-building groups, were useful in case they wouldn't recognize a concept but then still would know which brain-building group it belonged to.

Their overall opinion about the discriminatory concepts was that they were insignificant or only useful for crossing off concepts that they are not going to research. One participant said the following about the discriminatory concepts: *"I suppose there are a lot of interesting concepts that don't overlap. But you're interested in the ones that do overlap and those are the ones that are shown in the hierarchy of most overlapping, to say it like that, to the least. So that's the most interesting part of this structure, I think. Good and what's not overlapping, of course, also interesting but that's not what you're looking for"*. When asked if the discriminatory concepts should be removed, they all disagreed but still found the discriminatory concepts to be less relevant.

Their most pressing concern was the fact that if more concepts are added, the visualization could become chaotic. The initial design only shows 14 concepts but there are many more to show. The participants valued having a clear overview above showing all information there is available.

We then continued by asking what the participants opinion was of the function of extending concepts, as can be seen in figure 6. One participant called it a specific and detailed way to further research the relationship you're interested in. Another participant called it useful because in the early stages of research you don't exactly know what you are looking for and this could be a helpful tool in that regard. On the other hand, one participant found the links confusing. The participant thought at first glance that the links with the *Hippocampus* that were shown were also related to *Depression* and *Anxiety*. When learning that this was not the case, the participant argued that you were initially interested in the relationship between *Depression* and *Anxiety* and not per se links individually with the *Hippocampus*. They still think it is an advantage that you can extend links and see more information about certain concepts, but only when it's related to the two diseases as well. After these questions, the participants had no further comments, which brings us to the final questions about meaningfulness.

4.4.3 The meaningfulness of the information design. In the final part of the interview, we asked the participants if they found the task we are designing the system for is meaningful. After seeing the design and answering questions about the functionality and information shown, all participants agreed that looking at relationships between diseases could be useful in an early research stage.

Although, they had some comments about improvements that could be made. Two participants made suggestions for an extension of the data, which they would find useful. One participant found that the data repository should include symptoms as well, because of the importance they have towards comparing diseases. Processes, such as *neurodegeneration*⁶ where neurons slowly decay over time, was another suggestion that was made. Another participant suggested adding the prevalence to the diseases, that way it is clear which diseases are occurring more often in the population.

Finally, we asked if they would use a visualization like this in their daily practice, which they all agreed to. Two of them stated that by looking at the design/system they would get a plan and a strategy to further research the relation on their own. One mentioned that they would mostly find this interesting in an early research stage: *"I would think it would be most useful when you're just starting research. Because once you exactly know what topic you are looking for you are probably going to do more, like, not literature work, but like lab work. Especially once you're looking for something to research or, um yeah, especially in that phase"*. This information is particularly beneficial for us because it is one of the tasks (defined in the user scenario) we are designing for.

4.5 Implications of Interview Neuroscientists on follow-up design

The conclusion that we drew from these interviews is as follows. Firstly, that the task we are designing it for is meaningful and that most information shown on the information design is relevant to the user. On the other hand, there are also some improvements to be made and things to watch out for. The main concern for the user is that if more or all concepts are shown on the screen, that it would be too chaotic. Secondly, that many different functions or widgets can be added to help users understand the information on the screen. For example, one participant mentioned a few times to use visual cues to make it easier to interpret the number of co-occurrences at first glance. These small details can make the design clearer and easier to understand for users. Lastly, the initial design is not very suitable for a 3D environment. A few participants mentioned this as well. If we want the design to be an extension of the DatAR project, it needs to be suitable for 3D.

The information we received from the interviews were the basis for the following design requirements. Users should be able to:

- visually differentiate between six different brain-building groups (e.g., brain regions, genes etc).
- read the concept names shown without clutter at all times
- tell the number of co-occurrences by (multiple) visual cues.
- (temporarily) hide/show individual brain-building groups.
- understand that relationships between concepts are based on correlation, not causation.

These newly found design requirements and the initial design requirements guided us in creating two new designs. The improved information design and the 3D-centered design. We will discuss how we incorporated the design requirements and why we chose to create two designs in the following section.

5 SUMMARY OF DESIGN REQUIREMENTS

In this section, we will shortly summarize the design requirements we have created based on the information from the user scenario, the interviews with neuroscientists and the interviews with design experts. Previously, we hadn't ordered these design requirements and just used them as a guideline for the designs. In table 1, we ordered the design requirements into two groups: task-dependent and task-independent design requirements. We have defined seven task-dependent design requirements, which are important to support the tasks we have defined in the user scenario and assist the

user in achieving their goals. On the other hand, we have eight task-independent design requirements that focus on the visualization and additional functionality the application might need.

6 UI EVALUATION OF FINAL DESIGN WITH NEUROSCIENCE STUDENTS

To answer the last subquestion: "Is the design for the discovery of relationships between semantically similar brain diseases usable and clear to a neuroscientist?". We conducted interviews with neuroscientists where they had to perform tasks with the final design and fill out a System Usability Scale.

6.1 Final design explained based on design requirements

Because of some remarks of the design experts that the improved information design and the 3D-centered design would be clearer to them if there were animations, we choose to create the final design as a prototype in Invision².

The final design was created with the design requirements, which can be seen in table 1, in mind. We will explain on the hand of five figures, how we incorporated the design requirements into the final design.

We will start with the task-dependent design requirements. In figure 7, we see the two selected diseases, *Anxiety* and *Depression*, and in the middle of these two diseases, we see the concepts they share in common (D1). On the left side of *Anxiety* and the right side of *Depression*, we can see the discriminatory concepts of both diseases (D2). We chose the positioning of these concepts to make it immediately clear if the concept belongs to one of the diseases or both of the diseases. On the sides of the concepts we see grey bars with white numbers inside of them, the white numbers indicate the number of co-occurrences a concept has with a certain disease (D3).

If we click on a concept, for example, *Hippocampus*, we can extend the relationship as shown in Figure 8 (D4). In this figure, we see the selected concept *Hippocampus* on the left side and the selected diseases on the right side. The concepts in the middle all have a relation with these three concepts. On the left side of the concepts, the number of co-occurrences is shown with the *Hippocampus*. On the right side of the concepts, we have two bars, where the black bars indicate the number of co-occurrences with *Anxiety* and the grey bars indicate the number of co-occurrences with *Depression*.

D5 is not included in the design because this function already existed in the DatAR system. Just for example's sake, you could implement the functionality in the following way: if a grey bar on a concept is clicked, the system shows you the articles that are included in that particular relationship. D6 is also not added in the design, it is difficult to defend but the most obvious way to make clear that relationships are based on correlation, not causation, is by putting a disclaimer somewhere in the application or give this disclaimer beforehand. Due to space limitations, it would not seem

to be a great idea to add it in this design but in 3D, with "unlimited" space, it would be easier to show.

D7 is a design requirement we haven't incorporated in this design. In 3D there are endless possibilities to "play" with positioning. For example, you could indicate the ratio between the number of co-occurrences of both diseases by placing it somewhere in the middle of both diseases. We believed it was therefore necessary to include the design requirement that explains this distance to users. We eventually chose to make the design in a straight line and not use positioning to relay information. A disadvantage is that if this was implemented in 3D, you wouldn't use all the available space. An advantage of the straight line is that we could use this to implement a sideways histogram to indicate the number of co-occurrences, which satisfies design requirement I3.

In figure 7, we can see the bar with the names of the brain-building groups and to which colour they belong (I1). This was implemented to show which concepts belong to which brain-building group. For these colours, we chose to use the Colorbrewer application from Cynthia Brewer [6]. In this application, it is possible to select colours that are also distinguishable by colourblind people and can be of great help for colour selection. This bar has another function as well. When clicking on one of the brain-building group, for example *Proteins*, we can show or hide the group (I4). In figure 9, we see the result of hiding the brain-building group *Proteins*. We can see that the *Proteins* aren't in the design anymore and that the button *Proteins* in the bar below is made greyish to indicate that it is hidden.

I2 is a design requirement that in this design is easy to fulfil, but when there are many more concepts shown on the screen it could become chaotic and not as easy to fulfil, which is also the main point where the neuroscientists in the first round of interviews were afraid of. Therefore, we added a slider which can be seen in the lower right corner of the design to adjust how many concepts are shown on screen per group (I2). We can see what happens if the slider is put on two in figure 10. Where we can now see that two concepts per brain-building group are shown instead of three.

The black bar on the lower-left corner of the screen is used to change the ranking of the concepts (I5). In figure 7, we see that now the ranking is grouped for every brain-building group from highest to lowest summed co-occurrences. If we wanted to see, for example, which concepts have the highest relation with *Anxiety*, we can click on the button *Anxiety* in the lower black bar and that would bring us to figure 11. In figure 11, the concepts are not ranked according to brain-building groups but ranked on the highest number of co-occurrences with *Anxiety* to lowest. Because the discriminatory concepts with *Depression* don't have a relation with *Anxiety*, we chose to order them from highest to lowest with *Depression* to keep the design uniform.

The small bubbles beneath the groups indicate that for that particular group there are more concepts to show (I6). We differentiated between the legend and the concepts, as stated in I7, by putting the legend below the concepts. You could argue that this distinguishment could be made clearer by putting a line in between them for example, but we found this unnecessary. We made the distinguishment between the selected diseases and the concepts, by giving the diseases a black colour, making them larger and giving them

²The prototype of the final design is available online and can be tried out by clicking this link: <https://projects.invisionapp.com/prototype/20210301-Finaldesign-ckmn8z12u007r8601zdx3vqd6/play/271b43ee>

Design Requirement	Task dependency	Source	Section	Figure Shortcut
Visually identify concepts that are related to both selected diseases.	Dependent	Initial	4.1	D1
Visually identify concepts that are related to only one of the two selected diseases.	Dependent	Initial	4.1	D2
Tell the number of co-occurrences for all relationships.	Dependent	Initial	4.1	D3
Visualize initially hidden relationships between a concept and the concepts it relates to on request.	Dependent	Initial	4.1	D4
Investigate which articles a connection between concepts are based upon.	Dependent	Initial	4.1	D5
Understand what that relationships between concepts are based on correlation, not causation.	Dependent	Neuroscientist	4.5	D6
Understand what the distance between concepts in the UI represents.	Dependent	Design expert	5.5	D7
Visually differentiate between six different brain-building groups (e.g., brain regions, genes etc).	Independent	Neuroscientist	4.5	I1
Read the concept names shown without clutter at all times.	Independent	Neuroscientist	4.5	I2
Tell the number of co-occurrences by (multiple) visual cues.	Independent	Neuroscientist	4.5	I3
Temporarily hide/show individual brain-building groups.	Independent	Neuroscientist	4.5	I4
Order the relevance of related concepts.	Independent	Design expert	5.5	I5
Visually aware if there are concepts that are (temporarily) hidden to prevent interface clutter.	Independent	Design expert	5.5	I6
Distinguish between concepts and the legend.	Independent	Design expert	5.5	I7
Distinguish between the selected diseases and the concepts to which they are related.	Independent	Design expert	5.5	I8

Table 1: Design requirements

a white text colour instead of black (I8). In the following section, we'll discuss how we evaluated the final design.

6.2 Setup UI evaluation with neuroscientists

To answer the third subquestion, we conducted interviews with potential users to investigate the usability and clearness of the design.

Nielsen et al. has shown that five participants can identify up to 85% of the major usability problems in early interface development [9]. Although, this number of participants hasn't been without debate. Spool et al. debated that the formula used by Nielsen for finding usability problems is a fixed number, which is incorrect and that more users are necessary to find the major usability problems [12]. On the other hand, the metric held by Spool is also considered to be fairly extreme. Woolrych et al. found that we can only determine the number of users needed by user testing [17]. In his study, he performed usability testing with five candidates but concluded that three participants were enough to find most of the major usability problems. There is an ongoing debate about this topic, but we'll hold on to five participants in this study.

We asked the participants to perform predefined tasks with a prototype of the design made in Invision³. The five participants were all MSc students at the UvA or VU with a background in

Neuroscience. We encouraged the participants by offering them 10 euros, which they could donate to one of three predefined charities. We timed how long it took participants to complete a task and wrote down where they got stuck or were confused.

After a short introduction, we showed the prototype to the user without them having any knowledge about it beforehand. We then asked what all the elements on the home screen (see Figure 7) were and after that, we started with the following tasks:

- (1) Name the *Neurotransmitters* that have a relation with both *Anxiety* and *Depression*.
- (2) Name the *Protein* with the highest co-occurrence number with *Depression*.
- (3) Name the genes that are only connected to *Anxiety*.
- (4) Change the number of concepts shown per category from three to two.
- (5) Change the number of concepts shown per category from two to three. (Go to home screen)
- (6) Hide the brain concept group: *Proteins*.
- (7) Show the brain concept group: *Proteins*. (Go to home screen)
- (8) Change the ranking of the view to show the concepts that have the most co-occurrences with *Anxiety*.
- (9) Change the ranking of the view to show the concepts grouped. (Go to home screen)
- (10) Zoom in to the concept *Hippocampus*.

³<https://www.invisionapp.com/>

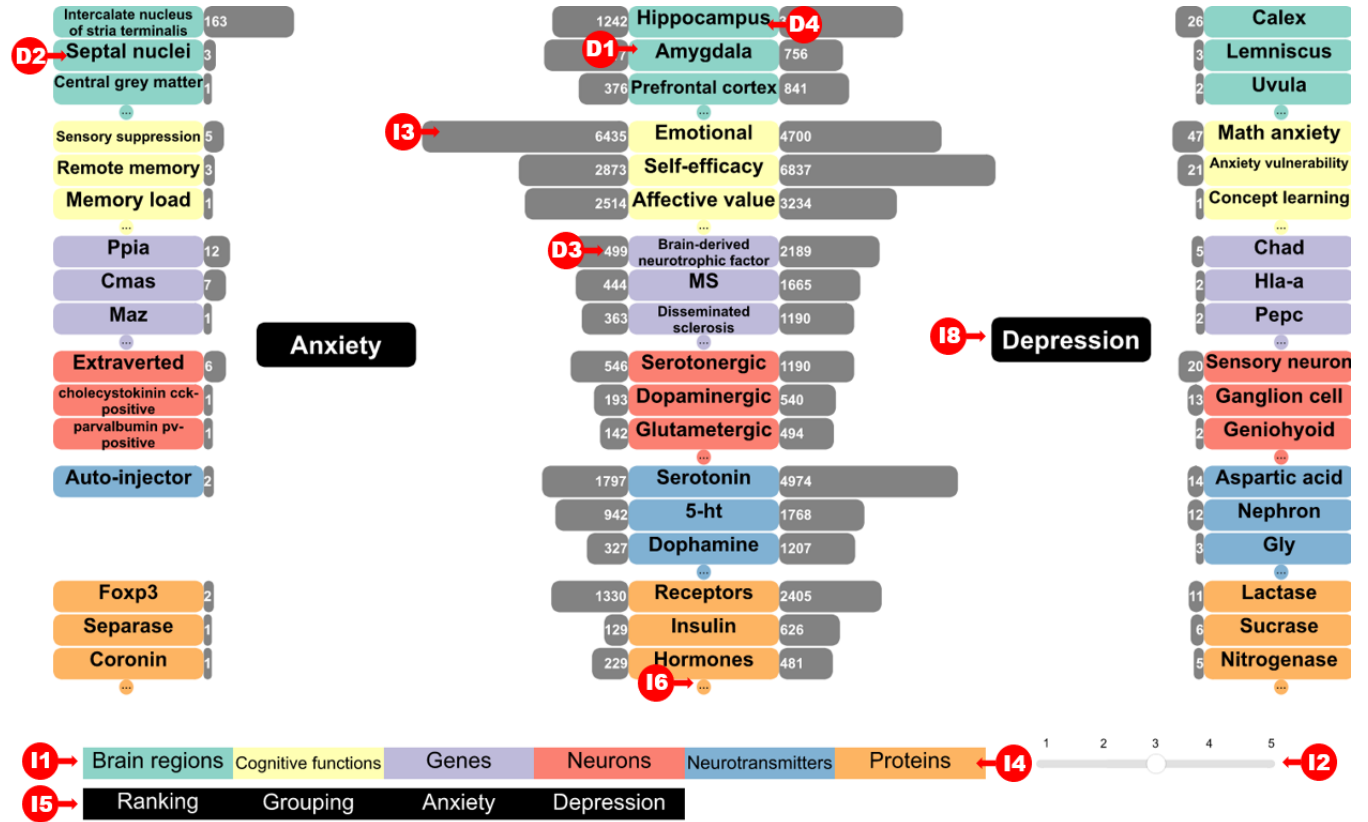


Figure 7: The final design consists of seven different brain concept groups. The black boxes are the selected diseases *Depression* and *Anxiety*. In the middle of the selected diseases, we see the concepts they share in common. On the left and right side of the diseases, we see discriminatory concepts. The grey bars next to concepts represent the number of co-occurrences for every relationship. On the bottom of the screen, we see the other six brain-building groups. These are six buttons that serve two purposes. Firstly, it is a visual reminder to know which colour belongs to which brain-building group. Secondly, they can be clicked to show/hide the brain-building group. Just below that is a black bar with 4 buttons: Ranking, Grouping, Anxiety & Depression. With this bar, the order in which the concepts are aligned can be changed. On the right bottom corner, a slider was added to change the number of concepts per brain-building group.

- (11) Name the *Proteins* that have a relation with *Hippocampus*.
- (12) Zoom out of the concept *Hippocampus*. (Go to home screen)

Because of Covid-19, we had to perform these tasks online which caused that we weren't able to see where the participants were on their screen and if they completed the tasks given. Therefore, we encouraged the users to think-aloud when performing the tasks to gather as much qualitative data as possible and to check if they completed the tasks [18]. After the tasks were completed, we asked if the participants had any additional comments or questions about the system.

Lastly, we let the user fill in a System Usability Scale (SUS), which is a survey consisting out of 10 questions to measure the usability of a product [2]. We changed the original 10 questions slightly by replacing the word cumbersome for awkward and replacing the word system for product. Bangor et al. suggested that these changes make the SUS more recognizable by using words that are more commonly used in the English language [1].

6.3 Results UI evaluation with neuroscientists

6.3.1 *Tasks and completion time.* While performing the tasks with the participants, we stumbled on the following usability problems:

At task 1, a participant misread the *Neurotransmitters* for *Neurons* and gave the wrong answer. It was caused by *neurons* and *Neurotransmitters* having similar terminology according to the participant.

At task 4, all participants used the slider to change the number of concepts correctly, but a few of them were confused about what happened when they used the interaction. After sliding a few times back and forth, they understood the functionality. This slider could be made clearer by using some kind of motion or indicator of what changed in the visualization.

At task 6, a participant clicked the indicator of more concepts (I6 in figure 7), after that the participant understood that the *Proteins* button should be clicked. The indicator, for now, is only a visual indicator that concepts are temporarily hidden. In the future, this

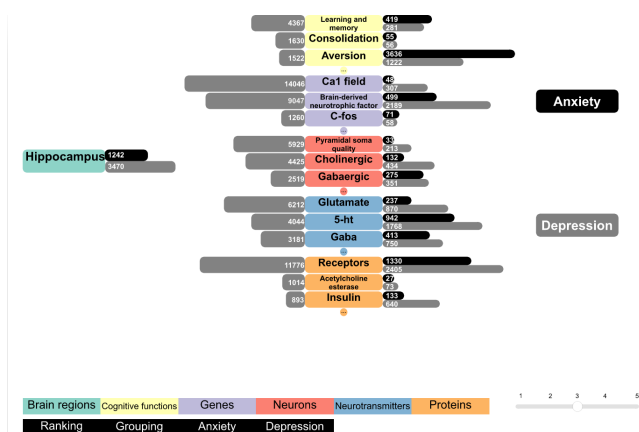


Figure 8: This visualization directly relates to design requirement D4. We see on the left side the concept *Hippocampus* that is selected. On the right side, we see the two selected diseases: *Anxiety* and *Depression*. In the middle, we see the concepts that have a relation with all three concepts. On the left side of the concepts, the number of co-occurrences with the *Hippocampus* is shown. On the right side of the concepts, the number of co-occurrences with *Anxiety*(black) and *Depression*(grey) is given.

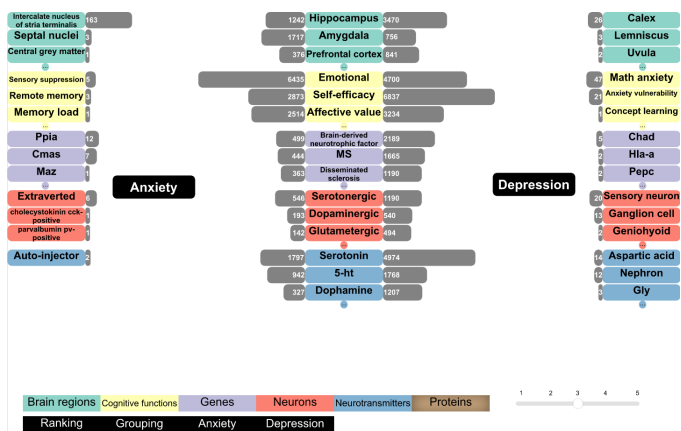


Figure 9: In this image of the final design, the button *Proteins* is clicked and we see the result of hiding the brain-building group *Proteins*. We can see that the *Proteins* aren't in the design anymore and that the button *Proteins* in the bar below is made greyish to indicate that it is hidden.

visual indicator could also be incorporated to show more concepts when selected or some other functionality.

Task 8 was by far where most people struggled and took the longest to complete as well, as can be seen in figure 12. It wasn't clear which button should be selected to show the highest to lowest co-occurrences for *Anxiety*. Two participants clicked on the button *Ranking*, one participant clicked on the disease *Anxiety*. Although participants understood what the function does, the lower black

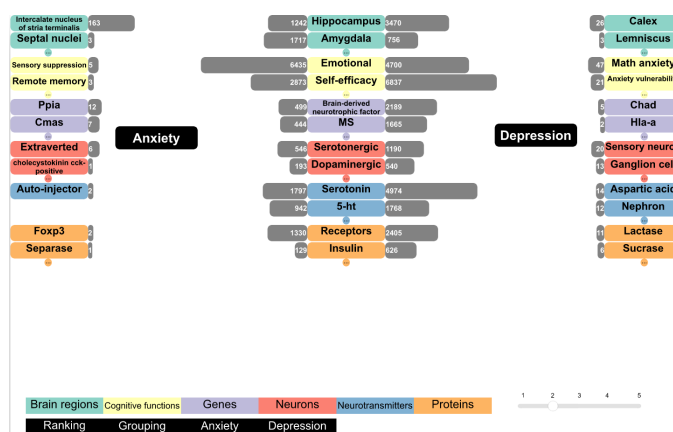


Figure 10: In this screen of the final design, we see that the slider in the lower-left corner has been moved from three to two. The result we see here is that the number of concepts per brain-building group has been changed from three to two.

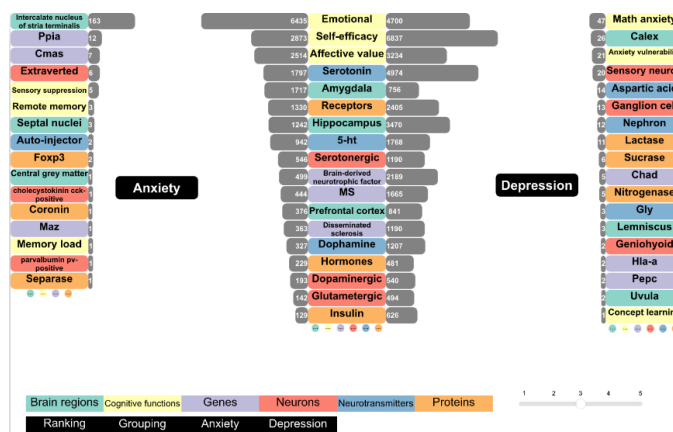


Figure 11: In this image of the final design, we see that the button *Anxiety* on the bottom left corner of the screen is pressed. The result is that we see the concepts ordered from highest to lowest number of co-occurrences with *Anxiety*. The discriminatory concepts of *Depression* are ordered based on the number of co-occurrences of *Depression* from highest to lowest.

bar wasn't immediately intuitive to participants. In the future, this bar should be reworked to make it easier for participants. This could be achieved by using different terminology and/or making it a different colour than the selected diseases.

At task 10, a participant clicked the right button but wondered what the grey bars on the left side of the concepts mean (see Figure 8). These bars indicate the relationship a concept has with an extended concept, in this case, the *Hippocampus*. The bars have the same colour as one of the selected diseases, in this case, *Depression*, which can cause confusion. Maybe the positioning of the bars isn't clear enough and in the future should be given a different colour.

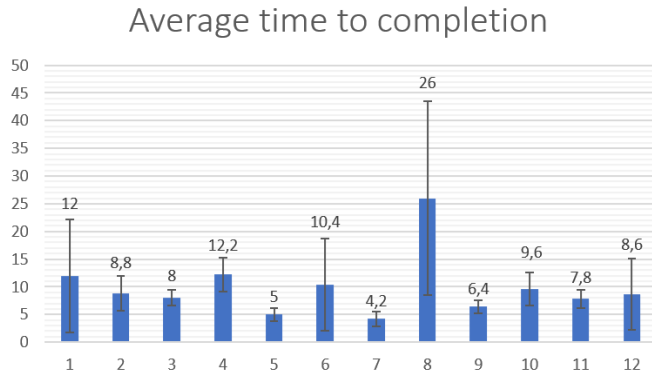


Figure 12: Average time to completion per task and the standard deviation.

At task 12, a participant clicked on the button Grouping in the lower black bar (I1) to go back to the home screen. After thinking about it a little while, the participant clicked on the right button.

6.3.2 Secondary findings. Two participants had the feeling that it was a bit much and chaotic. One of them mentioned that the main issue was the discriminatory concepts and wondered why they were even in the visualization. The other participant found that the way we visualize the application should be transformed. They suggested that instead of showing everything at once, a user should be able to select which brain-building groups they are interested in and from there be able to select more groups if interest arises. The other comments were mostly positive in the way that they found it intuitive to use and with small instructions being able to manage quite well.

6.3.3 System Usability Scale. After the interviews, the participants were asked to fill in the SUS. This scale focuses on the interface as a whole and does not assess individual items or tasks. The SUS has a range from 0 to 100. The results of the SUS per user is shown in figure 13. With an average score of 85, the design is in the higher regions of the acceptability range. The design is on the line of considering it good or excellent [1].

6.4 Implications UI evaluation

Based on the overall SUS score of 85 and that users didn't have major issues when performing the tasks, we can conclude that this design is usable and clear to a user. However, there are still improvements to be made in the final design. For example, at task 8 when changing the order of concepts, users took a long time and the majority of users clicked on the wrong button first.

7 DISCUSSION

The results of this research indicate the need for a visualization for the exploration of the relationship between semantically similar brain diseases in an early research stage. The majority of users were positive about how this could assist them with their literature search. However, during the interview with neuroscientists and the evaluation with neuroscientists, we came across some difficulties

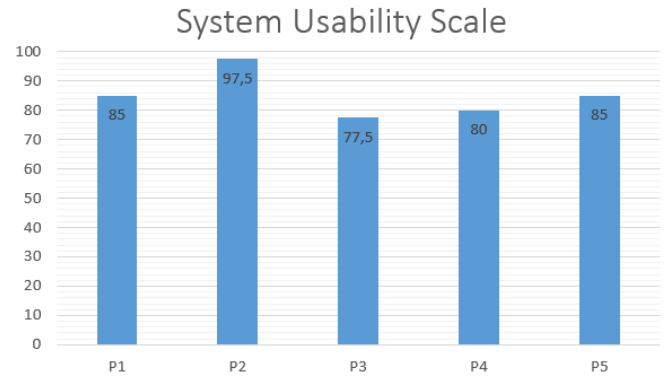


Figure 13: System Usability Scale score per user.

with the visualization.

Discriminatory concepts

The discriminatory concepts (For example, see figure 7 (D2)) have been a source of doubt as users saw them as less relevant and making the final design look chaotic. A reason we can consider for this is that during all the studies, we used the example of *Depression* and *Anxiety*. These two diseases are very close in the 3D topic model, so it is more interesting to see where this semantic similarity is based on and seeing concepts that are related to both diseases. If we used different examples of diseases, such as diseases that are further apart in the 3D topic model, we could have got different results about the discriminatory concepts and the visualization as a whole. An example that was given by a user also suggested that diseases that co-occur often in patients, such as *Parkinson's disease* and *Dementia*, could be more interesting than semantically similar diseases.

2D vs 3D

All the visualizations we created were implemented in 2D, while we are designing for an extension of the DatAR project which is in 3D. We tried to address this problem by visualizing an application that is also suitable for 3D but a major question that remains is if users would have the same opinion if this was implemented in 3D. Both design experts suggested that the functionality we have in this visualization is easily implementable in 2D, so what is the benefit of using 3D. They mentioned that only if you reach limitations in visualizing in 2D, the case could be made for implementing in 3D. Because we made a visualization that is an extension of the DatAR system and we need the 3D topic model for selecting diseases, it is implied that we design with a focus on 3D. The question nevertheless should be asked if it is necessary to design this application in 3D.

8 CONCLUSION

From the first study, we learned what information is relevant to a user. Additionally, users found the task we are designing for meaningful. However, we also learned what their main worries would be and how we could improve the design by adding functionality.

The second study gave us insight into how we could visualize the relevant information appropriately. The interviews we conducted with design experts gave us a better understanding of how to make the design suitable for 3D and that many aspects could be improved in our visualization. Based on the results of these two studies, we specified the design requirements.

In the third study, we evaluated the proposed final design on usability and clearness by letting neuroscientists perform tasks with a prototype of the final design and filling in a System Usability Scale. We learned that the design still had some usability problems but that most of the final design was clear and usable to neuroscientists.

For the future, we suggest implementing this in VR or AR and testing with users to see if the final design still is usable and clear to a neuroscientist. Since we have created this visualization specifically for the exploration of semantically similar brain diseases. More research is required to find out if this design can also be used for other purposes.

More research is also needed regarding the discriminatory concepts and their role in assisting neuroscientists in exploring the relationship between semantically similar diseases.

Aside from the visualization itself, many users would like to see an extension of the data available in the LBD. For example, symptoms of diseases and their relationships would be a very useful addition. Another extension that users would like to see is the sentiment a relationship has. Currently, only the relationship can be researched between two concepts, but we can't tell if this is a positive or negative relationship. For example, if we take the brain region *Depression* and the protein *Insulin*, we wouldn't know if the *Depression* is more often mentioned with an insulin deficit or an abundance of insulin. This wouldn't work for every relationship but it could be measured with Natural Language Processing or sentiment analysis.

We believe that the implementation of our application and its availability for users would improve users ability to find relevant relationships and research projects greatly. Therefore, we think our model can contribute to the academic world by increasing the user's accuracy and efficiency while simultaneously making the researcher process more comfortable.

9 ACKNOWLEDGMENTS

First and most of all, I want to thank my supervisor Lynda from the bottom of my heart. Thank you, Lynda, for never doubting me and always supporting me even when I doubted myself sometimes. Thank you as well for all the feedback on writing, interviewing and designing. Without you, this paper would have never been written. Secondly, I would like to thank Frank for the support he gave me when I had to stop with my first project and for guiding me to this project. Then I would like to thank everyone in the DatAR team for the great support and feedback. Last but not least, I would like to thank all the participants who contributed their voices and helped in creating this design.

REFERENCES

- [1] Aaron Bangor, Philip T. Kortum, and James T. Miller. An empirical evaluation of the system usability scale. *International Journal of Human-Computer Interaction*, 24(6):574–594, 2008. doi:10.1080/10447310802205776.
- [2] John Brooke. SUS: A 'Quick and Dirty' Usability Scale. *Usability Evaluation In Industry*, (November 1995):207–212, 1995. doi:10.1201/9781498710411-35.
- [3] Federico Desimoni and Laura Po. Empirical evaluation of Linked Data visualization tools. *Future Generation Computer Systems*, 112:258–282, 2020. doi:10.1016/j.future.2020.05.038.
- [4] Vishrawas Gopalakrishnan, Kishlay Jha, Wei Jin, and Aidong Zhang. A survey on literature based discovery approaches in biomedical domain. *Journal of Biomedical Informatics*, 93(March):103141, 2019. doi:10.1016/j.jbi.2019.103141.
- [5] Carsten Görg, Hannah Tipney, Karin Verspoor, William A. Baumgartner, K. Bretonnel Cohen, John Stasko, and Lawrence E. Hunter. Visualization and language processing for supporting analysis across the biomedical literature. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6279 LNAI(PART 4):420–429, 2010. doi:10.1007/978-3-642-15384-6_45.
- [6] Mark Harrower and Cynthia A. Brewer. ColorBrewer.org: An online tool for selecting colour schemes for maps. *Cartographic Journal*, 40(1):27–37, 2003. doi:10.1179/000870403235002042.
- [7] Philipp Heim, Steffen Lohmann, and Timo Stegemann. Interactive relationship discovery via the semantic web. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6088 LNCS(PART 1):303–317, 2010. doi:10.1007/978-3-642-13486-9_21.
- [8] Sari Kujala. *User studies: A practical approach to user involvement for gathering user needs and requirements*. Number 116. 2002. URL: https://www.researchgate.net/publication/34032560_User_studies_a_practical_approach_to_user_involvement_for_gathering_user_needs_and_requirements.
- [9] J Nielsen and J Landauer. A mathematical model of finding the usability problems. *Proceedings of ACM INTERCHI'93 Conference*, pages 206–213, 1993. URL: <http://delivery.acm.org/10.1145/170000/169166/p206-nielsen.pdf>.
- [10] Dmitry Paranyushkin. InfraNodus: Generating insight using text network analysis. *The Web Conference 2019 - Proceedings of the World Wide Web Conference, WWW 2019*, pages 3584–3589, 2019. doi:10.1145/3308558.3314123.
- [11] Alison Smith, Jason Chuang, Yuening Hu, Jordan Boyd-Graber, and Leah Findlater. Concurrent Visualization of Relationships between Words and Topics in Topic Models. pages 79–82, 2015. doi:10.3115/v1/w14-3112.
- [12] Jared Spool and Will Schroeder. Testing web sites: Five users is nowhere near enough. *Conference on Human Factors in Computing Systems - Proceedings*, pages 285–286, 2001. doi:10.1145/634067.634236.
- [13] D. R. Swanson. Fish oil, Raynaud's syndrome, and undiscovered public knowledge. *Perspectives in biology and medicine*, 30(1):7–18, 1986. doi:10.1353/pbm.1986.0087.
- [14] Ghazaleh Tanhaei, Lynda Hardman, and Wolfgang Huerst. Datar: Your brain, your data, on your desk—a research proposal. *Proceedings - 2019 IEEE International Conference on Artificial Intelligence and Virtual Reality, AIVR 2019*, pages 138–143, 2019. doi:10.1109/AIVR46125.2019.00029.
- [15] Ivar Troost, Lynda Hardman, and Wolfgang Hürst. Supporting Relation-Finding in Neuroscientific Text Collections using Augmented Reality : A Design Exploration Supervised by. pages 1–21, 2020. URL: https://dspace.library.uu.nl/bitstream/handle/1874/397219/troost_2020_supporting_relation_finding_in_neuroscientific_text_collections_using_augmented_reality.pdf?sequence=1&isAllowed=y.
- [16] Fabio Viola, Luca Roffia, Francesco Antoniazzi, Alfredo D'Elia, Cristiano Aguzzi, and Tullio Salmon Cinotti. Interactive 3D exploration of RDF graphs through semantic planes. *Future Internet*, 10(8):1–30, 2018. doi:10.3390/fi10080081.
- [17] Alan Woolrych and Gilbert Cockton. Why and when five test users aren't enough. *Proceedings of IHM-HCI 2001 conference*, 2(JANUARY 2001):105–108, 2001. URL: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.85.7896&rep=rep1&type=pdf>.
- [18] Peter C. Wright and Andrew F. Monk. The use of think-aloud evaluation methods in design. *ACM SIGCHI Bulletin*, 23(1):55–57, 1991. doi:10.1145/122672.122685.