Assignment 8

In contrast to the previous assignments, this one is individual, i.e., you have to do it alone, not in a group. It is intended to be an "exam-style" recap of the lectures on empirical research, user studies, writing & publishing, and statistics.

Email your results in a PDF to Wolfgang (huerst@uu.nl) considering these formalities:

- For the email: Use the subject [INFOMSCIP] Assignment 8

- For the PDF file: Use the name INFOMSCIP-8-ID.pdf where ID is **your student ID**.

Include your name and student ID also at the begin of the PDF.

Deadline for this assignment: Send the email to Wolfgang before **Thu, Nov 7, 2019, 11am**.
The deadline is strict. Each additional day after the deadline will result in a grade deduction of 1.0.

Deliverables: In the following, you find some questions related to the content covered in the lectures about empirical research, user studies, writing & publishing, and statistics. Answer them, write your answers in a nicely formatted text file, and email it to huerst@uu.nl before the deadline. Do not forget to add your name and student ID at the begin of the file. There is no need to specify your hours.

Question 1: *Scientific publications*

Scientific results are often published at conferences. There are very good, highly respected conferences, medium quality conferences (which are still useful, e.g., to publish preliminary results), and very bad events, where publishing might even hurt your scientific reputation. Likewise, there are published scientific papers of high and low quality. While you are more likely to find high quality work at good conferences, it sometimes also happens that a paper with a lower quality gets accepted accidentally. And just because an event is not ranked very high, does not mean that all publications there are of low quality.

1. Give two indicators for a good conference. For each of them, state also a potential problem (i.e., why this indicator alone might not be sufficient to judge the quality of an event).

2. Give two indicators for a good scientific paper. For each of them, state also a potential problem (i.e., why this indicator alone might not be sufficient to judge the quality of a paper).

One important aspects of research is that it should make a contribution to the respective scientific field. In their paper "Research contributions in human-computer interaction", Wobbrock and Keintz define different contribution types for HCI.

3. Find one paper published at the ACM CHI 2019 conference that makes an "empirical research contribution."

4. Find another paper from this conference that makes an "artifact contribution."

In both cases, write down a proper citation of the paper. Then, shortly summarize the contribution of each of them in a way that makes it clear what the authors did, but also why their contribution is part of this category from the paper by Wobbrock and Keintz (keep it short: 2-4 sentences per paper).

Hints & comments:

On the ACM CHI 2019 website, you find a page with the abstracts of all papers, which can be helpful for identifying potentially relevant ones: https://chi2019.acm.org/for-attendees/proceedings/

You can also find the proceedings in the ACM Digital Library here: https://dl.acm.org/citation.cfm?id=3290605

In the following question, you will need to have a closer look at the paper from the "empirical research" category. Thus, it makes sense to also pick one that you personally find interesting.

Question 2: *Data, measures, and ethics*

Two important aspects of scientific research are data and measures. The data is what we usually base our conclusions on. The measures specify how we get these conclusions by allowing us to quantify our observations (although in HCI, we can also have qualitative measures). Lets have a closer look at the ACM CHI 2019 paper that you selected as example for empirical research.

1. What kind of data is used in the empirical research described in the paper? Shortly describe it and then, again shortly, critically discuss it.

   Note that for empirical studies in HCI, we usually have data that is captured by observing humans or their actions when interacting with technology. Yet, it could also be data that was generated automatically (e.g., via a simulation). What is meant by "discuss it" above is that you should shortly state why the choice of the authors to use this type of data was good. What is meant by "critically discuss it" is that you should also state possible problems and pitfalls of their choice and indicate what might have been a better choice or how they could have overcome these potential problems. Aspects to consider include, for example, who did they pick (if they used human subjects to create the data) and how did they pick them.

2. What kind of measurements are used in the empirical research described in the paper? Shortly describe them and then, again shortly, critically discuss them.

   In relation to user studies involving humans, measures are often used to quantify our observations. This can include qualitative and quantitative measures. Shortly state what the authors used, why this was a good choice, and what potential problems or limitations (if any) this choice might have had. An important question in relation to the latter is if the authors' choice of measures really allows them to draw general conclusions and convincingly prove their claimed contributions.

3. Shortly summarize all ethical aspects that are relevant in relation to this paper and the research presented in it. Indicate what the authors describe themselves. Yet, also list aspects that are not addressed in the paper but relevant as well.

Question 3: *Experiment design and analysis*

1. Give two advantages that a within-subjects design of a user study could have compared to a between-subjects design.

2. Give one convincing example for a user study where a between-subjects design would clearly be better than a within-subjects design and shortly explain why.

3. To avoid order effects, we often counterbalance the order of conditions across participants and test runs. Give a good reason why one might not do this but use a Latin square ordering instead.

4. Give one convincing example where we would not counterbalance the order of conditions, but use a fixed order for each test subject.

5. Give one convincing example where we would use a combined approach, i.e., have a fixed order for some conditions, but an order that is counterbalanced across participants and test runs for others.

6. What are descriptive statistics and what are they used for?

7. What is the power of a significance test?

8. Assume the results of your significance test do not allow you to reject the null hypothesis. Give three possible reasons why this might have been the case.