Summary of Heuristic Evaluation of Prototypes (Group)

Due: End of studio day (Thur/Friday, Feb. 25-26 @ 11:59 PM PST)

Overview

The goal of this assignment is to **summarize the heuristic evaluations** and **assign severity ratings to violations** so that project teams can fix the most important problems found in their prototype UI. Make a copy of this template file for your group to collaborate on. Throughout the report, use **anonymized reviewer IDs, not names**.

Report Format

1. Problem/Prototype Description

One sentence description of the UI you are evaluating.

2. List of Heuristic Violations

You will be summarizing the HE reports along with the others who evaluated the same project.

A. List each distinct problem with a unique number. You may have to merge similar violations from different evaluators. For example, two evaluators have written about the same violation:

> *Report from evaluator A:*
> ***5. H4 Consistency & Standards***
> *The interface used the string "Save" on the first screen for saving the user's information, but used the string "Update" on the second screen. Users may be confused by this different terminology for the same function.*
> *Fix: Use "Save" on all screens.*
>
> *Report from evaluator B:*
> ***18. H4 Consistency & Standards***
> *"Update" and "Save" are used interchangeably in the interface. This is confusing.*
> *Fix: Pick one!*

You should give it a unique number and list it **once** in your summary report (with the **best** description and suggested fix from the multiple evaluators that found the violation).

B. Indicate the heuristic violated.

C. Assign it a severity rating based on the following ratings defined in lecture:
0 = not a problem, 1 = cosmetic, 2 = minor, 3 = major, 4 = UI catastrophe

Your report will list each of the problems found in the following format:

Problem # [heuristic violated] / [severity rating] / [reviewer ID(s) of those who found it]:
Description of problem, reasoning why it violates the heuristic & suggestion for fix.

For the above example, you might write:

1. H4 Consistency & Standards / Severity 3 / Found by: A, B
The interface used the string "Save" on the first screen for saving the user's information, but used the string "Update" on the second screen. Users may be confused by this different terminology for the same function.
Fix: Use "Save" on all screens.

3. Summary of Violations

Your report will also summarize the **number of violations** found in each of the ten heuristic categories, the number of violations found in each of the severity categories, and the total number of violations in the entire interface. Fill in the table provided for this in the template we've provided.

For example:

Category	**# Viol. (sev 0)**	**# Viol. (sev 1)**	**# Viol. (sev 2)**	**# Viol. (sev 3)**	**# Viol. (sev 4)**	**# Viol. (total)**
H1: Visibility of Status	0	4	0	5	1	**10**
H2: Match Sys & World	1	etc...				
H3: User Control	2					
H4: Consistency & Standards	1					
H5: Error Prevention	0					
H6: Recognition not Recall	1					
H7: Efficiency of Use	0					
H8: Minimalist Design	0					
H9: Help Users with Errors	0					
H10: Help & Documentation	0					
Total Violations by Severity	**5**					

4. Evaluation Statistics

Your report will give an analysis of **how well each of the evaluators did** in finding problems in the interface (use anonymized reviewer IDs, not names). This will show you how well the technique works:

- Refer to the *total violations by severity* row of the ***Summary of Violations table*** from Part 3 to get the total number of violations detected at each severity level.
- Sum the number of violations ***each evaluator*** detected at each severity level.
- For each evaluator and each severity rating, calculate the % of the total violations of that severity level found by that evaluator. *(e.g., If Evaluator A found one of the five total "severity 1" violations, write down 20%)*
- For each evaluator, list the % of all violations found, and the % of severity 3 & 4 violations.

For example:

Severity/Evaluator	**Evaluator A**	**Evaluator B**	**Evaluator C**	**Evaluator D**	**Evaluator E**
Sev. 0	10%	etc…			
Sev. 1	20%				
Sev. 2	5%				
Sev. 3	10%				
Sev. 4	10%				
Total (sev. 3 & 4)	10%	30%	20%	50%	80%
Total (all severity levels)	25%	75%	80%	10%	40%

(**Note that the bottom two rows are *not* calculated by adding the numbers above it.)**

5. Summary of Recommendations

Merge together the general recommendations each of you made in your individual reports and make sure that they are coherent and consistent. This should be a few paragraphs.

Deliverable

Your group report is due by the end of the day of your studio (e.g., Thursday night or Friday night). Please **share your report, as a Google doc, with your TA**. **Make sure your file name contains the name of the project you evaluated (e.g., "CS147 Group HE - [Project Name You Evaluated]")**. Also make sure that everybody's name is listed at the top of the assignment (as formatted in the provided template, with letters).

Your typed write-up should follow this outline with separate sections for the top-level items:

1. Problem/Prototype Description (one-sentence description of project and the UI being evaluated)
2. List of violations found in format described above
3. Summary of violations (table provided in template)
4. Evaluation statistics (table provided in template)
5. Summary of recommendations

Examples of strong reports: Cabana, Localized

Grading Criteria

Report (100 pts)

- Prototype Description (10 pts)
    - Does the problem statement make a solid attempt to holistically capture what the application seeks to achieve?
- List of Heuristic Violations (50 pts)
    - Is the list a readable and digestible report that follows the given format for listing problems? Does it avoid reporting repeat problems? (15 points)
    - Are the descriptions of the heuristic violations detailed enough to be actionable? (15 points)
    - Does the list properly cover the individual reports that were given? (10 points)
    - Does the list give good coverage of the actual problems in the medium-fi prototype? (10 points)
- Summary of Violations (15 pts)
    - Was the provided template used to accurately sum up all of the violations?
- Evaluation Statistics (15 pts)
    - Was the provided template used to accurately reflect how well each of the evaluators did? Did the table identify evaluators by letters, and not names?
- Summary of Recommendations (10 pts)
    - Were the general recommendations made by the individual evaluators merged in a coherent manner?