**HCI 기말**

**<Evaluation and user study design>**

**\* Evaluation Techniques**

\* **Evaluation (평가)**

: Evaluation은 어디서 하는가에 따라 다른 기술을 사용한다.

- 시스템의 **usability**와 **functionality**를 평가한다. => **design**과 **implementation**을 평가

- design의 수명 주기의 모든 단계에서 고려되어야 한다.

**1) Usability testing**

: 실제로 UI/UX를 평가하는 것이다. 사용자가 얼마나 사용하기 쉽고, 얼마나 잘 사용하게 하는지에 대해서 평가하는 것. **Product 자체를 향상시키기 위해** 하는 testing이다.

- Improve products 제품을 향상시키기 위해 usability testing을 함.

- Few participants

- Results inform design (결과가 design에 정보를 제공한다.)

- Usually not completely replicable (일반적으로 반복을 할 수 없다.)

- Conditions **controlled as much as possible** (가능한한 많이 제어).

**2) Experiments for research**

: 보통 연구목적으로 많이 사용한다. 참여자가 많으면 좋고, 통계적으로 유효하게 결과가 나왔는지를 확인해야 한다.

- Discover knowledge 지식을 발견하기 위해 research를 함.

- Many participants

- Results validated statistically (통계적으로 검증된 결과)

- Must be replicable (반복해야 함)

- **Strongly controlled conditions** (강력하게 제어)

**\* Type of evaluation**

1) Evaluation place: 어디에 일어나는 일인지에 따라 나눌 수 있다.

- Laboratory studies

: 제한된 장소에서 일어나는 평가.

- Field studies

: 어느정도 개발이 된 이후에 일어나는 평가. Longitudinal studies에 적절함.

**2) Evaluation method => 이 chapter에서 다루는 것.**

- Observational evaluation

- Heuristic evaluation

- Analytical evaluation (use previous theory, conduct your own user study design)

**\* Observational evaluation**

: 관찰하면서 평가하는 방식

1. **Cognitive walkthrough**

: 새로운 user가 새로운 system을 사용하려고 할 때 얼마나 쉽게 사용할 수 있는지를 볼 수 있도록 하는 평가 방식. Design이 사용자가 **올바른 goal**에 도달하는가를 분석.

Walkthrough는 실제 평가를 할 때, 한발씩 나아가면서 전체 step을 천천히 리스트업 해서 평가하는 방식이다.

**\* 고려사항**

- 상호작용이 사용자에게 미치는 영향은?

- 어떤 인지 과정이 필요한가?

- 어떤 학습 문제가 발생할 수 있는가?

1. **Think Aloud (구두 서술)**

: 사용자가 수행하는 작업을 관찰한다. 사용자는 자신이 하고 있는 일과 왜 하고 있는지, 그리고 왜 그 작업이 일어나고 있는지에 대해 설명하기를 요청받는다.  
\* verbalize (말로 표현하다)

**\* 장점**

1) **simplicity** => 적은 전문가가 요구된다. (사용자에게 직접 물어보기 때문에)

2) 시스템이 실제로 어떻게 사용되는지를 보여줄 수 있다.

**\* 단점**

: **Subjective and selective** 사용자에게 물어보기 때문에 주관적이고 선택적이다.

1. **Cooperative Evaluation**

: 사용자에게 말하게 하고 관찰하는 방법이다.

Thick Aloud를 기반으로 변형시킨 방법. 사용자들은 평가안에서 협력하게 된다.

사용자와 평가자는 각각의 질문을 중간에 할 수 있다. 이 질문들을 통해 그 사람이 어떻게 행동하고 있는지를 평가하게 되는 것이다.

**\* 장점**

1) 제약이 적고, 사용하기에 쉽다.

2) 사용자들은 system을 비판하도록 권장된다.

3) 설명이 깔끔하다.

**\* Heuristic Evaluation**

: 전문가가 있고, 전문가가 interface를 평가하는 것. 일반적으로 알려져 있는 usability 원칙을 얼마나 만족하는지 **경험에 의해 평가**하는 방식이다.

Evaluators = Experts

전문가의 수가 많아질 수록, 제품의 문제를 더 많이 발견할 수 있지만 비용이 많이 든다.

**\* Nielsen’s Original Heuristics**

**1) Visibility of system status**

: 시스템은 타당한 시간 내에 **적절한 피드백**을 통해 사용자에게 항상 **현재 일어나고 있는 일**에 대해 알려야 한다.

**2) Match between system and the real world**

: **system과 현실이 일치**하는가에 대한 평가.

1. **User control and freedom**

: 사용자가 기능을 쉽게 제어하고 자유롭게 행동할 수 있는가에 대한 평가.

Ex) Undo, redo가 있는가?

1. **Consistency and standards** 일관성과 표준

: 사용자들은 다양한 단어, 상황 또는 행동들이 같은 것을 의미하는지 아닌지에 대해 궁금해하지 않아야 한다. 즉, 아이콘들이나 기능들이 consistency를 지켰는지에 대한 평가.

1. **Error prevention**

: 에러를 방지하기 위한 옵션을 제공했는지에 대한 평가.

1. **Recognition rather than recall**

: 개체, 작업 및 옵션을 표시하여 사용자의 메모리 로드 최소화, 객체, 액션, 옵션 및 방향을 쉽게 검색 가능

1. **Flexibility and efficiency of use** 유용성 및 사용 효율성

: 사용자가 자주 사용하는 행동들을 조정할 수 있는가에 대한 평가 (ex. **shortcut**)

1. **Aesthetic and minimalist design** => 디자인을 최소화했는가? 불필요한 정보가 없는가?

: dialogue가 관련이 없거나 거의 사용되지 않는 정보를 포함해선 안된다.

1. **Help users recognize, diagnose, and recover from errors**

: 사용자가 오류를 인식, 진단 및 복구할 수 있도록 지원하는가?

1. **Help and documentation**

**\* Mobile Heuristics**

**\* Stages of heuristic evaluation**

**1) pre-evaluation training**

: 평가자에게 시나리오에 필요한 지식과 정보를 제공한다.

**2) Evaluation**

: 개인들은 UI를 평가하고, 문제의 목록을 작성한다.

1. **Severity rating (위험성 평가)**

: 현재 여기에 있는 문제가 **얼마나 심각한지 언급**해주는 단계

1. **Aggregation 집계**

: 전문가들끼리 모여서 의견을 모으는 등의 그룹 미팅을 해서 실제 별점을 매기는 것.

1. **Debriefing**

: Design team과 함께 결과를 토론한다.

**\* HE vs. User Testing**

- HE는 전문가 5명정도만 모아 놓고 평가하기 때문에 더 빠르다.

- HE는 사용자의 행동을 해석할 필요가 없다.

- User Testing이 더 정확(accurate)하다.

**\* Analytical evaluation**

: 인터페이스 및 시스템 설계 시 모터 동작을 비교하거나 평가하는 방법.

Ex) Paul Fitt’s Law, Keystroke Level Model

\* Physiological methods

1) Eye tracking

2) Emotional response linked to physical changes 물리적 변화와 관련된 감성 반응

: Interface에 대한 사용자의 리액션을 결정하는 것에 도움을 줄 수 있다.

**\* Study design: 연구 설계**

**1) Hypothesis** 가설

: 평가를 하길 원하는 statement.

- Independent Variable (원인): 실험자에 의해 측정되어야 하는 변수

- Dependent Variable (결과): 독립변수에 의해 일어나는 변수 (즉, 종속변수)

\* Null hypothesis **H0**: 효과가 없다고 가정

: 조건 간에 차이가 없는 statement를 정의(no difference between conditions). 가설의 목적은 이를 반증하는 것이다.

\* Alternative hypothesis **H1**: H0의 반증(rejection of H0). 효과가 있다고 가정

: H1은 대답해야 할 질문(question to be answered). 반면에 H0은 H1의 반대이다.

**\* Hypothesis testing**

가설검정은 모집단의 값/분포에 두 개의 가설을 세우고, 이것의 참/거짓을 표본정보로부터 판단하는 것이다.

항상 알아야 할 것: samples(n), p value, variance/standard deviation, means

**- Alpha a-level (기준)**

: 분석전에 변경해야 함. 보통 p = 0.05 또는 0.01로 고정한다.

만약 p가 set level보다 낮다면, 중요한 결과인 것이다.

**- P-value (확률)**

: p는 관찰되는 구체적인 데이터의 확률이다. H0이 참이라고 주어졌을 때. (no difference)

- F statistic or T statistic

: 설정해 놓은 a-level 보다 p가 더 작다면, 중요한 결과인 것이고 이때에는

H0은 부정되고 H1은 승인된다. 이 작업에서 significant difference가 있다는 것이다.

**\* T-Test**

: 관측된 두 평균이 통계적으로 다른지 여부를 결정하는 데 사용되는 통계 검정.

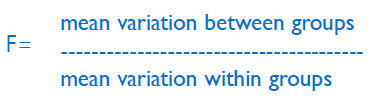
Types of T-Test

1. Independent groups
2. Dependent / paired

“Tails”

1. One sided / one tailed => A is bigger than B
2. Two sided / two tailed = > significant difference between A and B

**\* ANOVA: Analysis Of Variance**

: 두 개 이상의 집단을 비교하고자 할 때 집단 내의 분산, 총평균과 각 집단의 평균의 차이에 의해 생긴 집단 간 분산의 비교를 통해 만들어진 **F분포**를 이용하여 가설검정을 하는 방법이다.

**결론**: factor 사이에는 차이가 있다.

1. Between subjects ANOVA

: 다른 그룹들의 비교

1. Within subjects ANOVA

: 반복된 측정, 같은 참여자들의 더 많은 측정. 측정 여러번

1. One way ANOVA 일원분산분석

: 하나의 factor.

1. Multifactorial ANOVA

: factor들 사이에 각각의 factor의 주된 영향과 상호작용.

**\* Correlation 연관성**

: 두 변수가 서로 연관되어 있다. 0.7보다 높으면 연관성이 높은 것.

**\* Between subjects (그룹 간 차이)**

: 때때로 factor들은 subjects 사이에 있어야한다. (ex. Gender, age)

\* **장점**: interference effects(간섭 효과)를 피할 수 있다.

\* **중요성**: 임의의 조건 할당.

**\* Within subjects (사람 내 차이)**

: 때때로 factor들은 subjects안에 있어야 한다. (ex. measuring learning effects)

\* **장점**: 적은 참여자만 필요하다. test 조건 간에 동일한 주제 간 차이

**\* Biases 편향**

: 모집단 모수의 값을 과대(over) 또는 과소(under) 추정하는 측정 프로세스의 경향

- Hypothesis Guessing

: 참여자들은 가설을 시도하고 있는 것을 추측한다.

- Experimenter Bias

: 찾고자 하는 것을 찾기 위한 데이터 및 평가에 대한 무의식적인 편견

- Systematic Bias

: 시스템에 통합된 결함으로 인한 편향