

사용성 평가 방법들의 비교 및 선택

- Comparison and Selection of Usability Evaluation Methods -

임 치 환*
Lim, Chee-Hwan

Abstract

There is a currently a focus on the usability of interactive computer products. Many usability measurement and evaluation methods have been developed in the fields of human-computer interaction and software ergonomics. In this paper, the issues of usability evaluation methods are introduced, and the usability evaluation methods are classified and compared. Also, the framework for systematic usability effort based on product' lifecycle is presented.

1. 서론

산업 기술과 정보의 발달로 제품들 (products)의 기능은 더욱 복잡해지고 이를 사용하는 주체인 인간의 능력은 제한적이므로, 사용자는 제품을 사용하는데 곤란하거나 배우기 어렵다면 좋은 제품이라도 사용하기를 꺼리게 된다. 이러한 현상은 인간이 같은 품질의 제품에 대하여 인간 자신이 느끼는 제품의 만족도(satisfaction), 유용성(utility), 사용편의성(usability; 사용성) 등에 기준을 두어 제품을 선택하기 때문이다. 특히, 컴퓨터 제품의 사용은 고도의 지적 처리 과정을 포함하기 때문에, 사용자들에 대한 이해, 학습, 숙련 등이 필요하며, 인간공학적 평가에서 사용성 문제가 강조되고 있다 [11]. 또한 컴퓨터를 비롯한 정보 기술 제품의 사용자 인터페이스에 관한 국제적 표준들이 의미하듯 컴퓨터 제품의 인간 중심적인 설계의 중요성이 증가하는 추세이다.

지금까지의 사용성 연구(usability study)에서 주요한 관심은 제품 사용시 발생하는 사용성 문제들(usability problems)을 어떻게 측정하고 평가하느냐 하는 것이었다. 제품의 사용성 문제들이 다시 일어나지 않도록 보장하고 평가하기 위하여 많은 방법들이 연구되고 있다. 실제로 제품의 사용성을 평가할 때는 비용 면에서 효과적이고(cost-effective), 고도의 기술을 필요로 하지 않으며, 적은 투자(low-investment)로 평가 할 수 있어야 할 것이다. 이것은 가용 시간과 주어진 예산 하에서 가능한 빠르고 저렴한 평가 방법이 제품의 주요한 문제점들을 발견하는데 사용되어야 한다는 것을 의미한다.

* 한국과학기술원 산업공학과

본 연구에서는 제품의 사용성 평가 방법들에 관한 주요 문제들 (issues)을 소개하고, 평가 대상자의 관점에서 사용성 평가 방법들을 분류한다. 그리고 각 평가 방법들의 장점 및 단점을 비교함으로써 주어진 예산과 시간 제약 하에서 제품 수명 주기에 가장 적합한 평가 방법들의 전략 수립에 도움을 주고자 한다. 또한 제품 수명 주기와 사용성 평가 방법들을 결합함으로써 체계적으로 제품의 사용성 보증(usability assurance)을 할 수 있는 틀/framework)을 제시하고자 한다.

2. 사용성 평가 방법들의 분류

일반적으로 제품 개발 과정에서 사용성을 측정하고 평가하기 위해서 이용되는 방법들은 다양하며, 이중에서도 널리 사용되는 방법들은 다음과 같다: 개념 실험(Concept test), 전문가의 관찰(Expert review), 모의 실험 시도(Simulation trials), 반복적인 실험실 실험(Iterative laboratory experiments), 감사(Audits), 독백 기록(Thinking aloud), Constructive interaction (codiscovery learning), 자동기록법(Autologging), 추후보완연구(Follow-up studies), 현장 연구(Field studies) 등 [12]. 지금까지 연구되어 온 제품의 사용성 평가 방법들을 분류하면 <표 1> 처럼 네 가지로 요약할 수 있다.

<표 1> 사용성 평가 방법들의 분류

사용성 평가 방법들	실 명
자동화된 (automatical) 방법	자동화 프로그램을 통해 사용자 인터페이스 명세서 (specification)를 실행함으로써 사용성 척도들 (usability measures)을 계산.
실증적 (empirical) 방법	실 사용자들을 대상으로 사용자 인터페이스를 테스트 (test) 함으로써 평가.
정형화된 (formal) 방법	사용성 척도들을 계산하기 위한 식이나 정확한 모델들을 사용.
비정형화된 (informal) 방법	평가자들의 주먹구구식 방법과 일반적인 기술, 그리고 경험에 근거한 평가.

사용성은 사용자(user), 직무들(tasks), 시스템(system), 그리고 환경(environment) 등 시스템이 사용되는 모든 상황(context)을 고려해야 하므로, 현재 기술 수준 하에서는 자동화된 방법은 잘 사용되지 않는다. 시스템 운용상의 복잡도(operational complexity)를 평가하기 위한 정형화된 방법은 사용자 행태(behavior)를 직접 측정하기 보다는 오히려 사용자가 시스템과 상호작용할때 일어나는 것을 분석하고 예측하는데 사용하며, 숙련자의 수행도를 평가하는데 적합하다. 그러나 이 방법은 사용자에 대한 학습의 전이(transfer), 사용시 범한 오류, 사용시 즐거움(pleasant-to-use) 등을 잘 설명하지는 못한다 [9]. 또한 정형화된 분석을 수행하기 위해서는 직무에 대한 상세한 묘사(description)를 필요로 하며, 직무의 작은 변동에도 많은 재분석이 요구되기도 한다. 따라서 정형화된 방법들은 규모가 큰 사용자 인터페이스에 적용하기 곤란하고 해석하기 어렵다. Karat et al. (1992)는 사용성 평가를 위해 정형화된 방법들을 실제 적용하기 전에 고려해야 할 문제들이 많이 있으며, 제품 설계 주기에서 각 단계의 요구 사항이 다르므로 각 단계에 적합한 평가 방법을 고려해야 한다고 하였다 [4].

실증적 방법들은 아마도 가장 일반적으로 사용되는 사용자 테스트(user testing)과 함께 사용자 인터페이스를 평가하는 주요한 방법들이다. 그러나 종종 전개되는 디자인의 모든 버전들의 모든 측면들을 테스트하기에 충분한 실 사용자들 (피실험자들)을 모집하는 것이 어렵고 많은 비용을 수반할 수 있다. 이러한 제약을 보완하기 위해 사용자들 (피실험자들)을 줄일 수 있는 방법으로써 사용성 검사(usability inspection) 방법이 있다. 정해진 프로젝트 스케줄이나 한정된 예산 때문에 비정형화된 평가 방법이 이용된다.

평가 방법들에 대한 다른 분류 방법은 제품을 누가 평가하느냐에 따라서 나누는 것으로 전문가에 의한 평가와 사용자에 의한 평가로 나눌 수 있다. 전문가에 의한 평가 (expert-based evaluation)는 사용자 인터페이스 전문가나 인간 공학 전문가들의 경험과 지식, 여러 가지 디자인 원칙 등에 근거한 평가이다. 사용자에 의한 평가 (user-based evaluation)는 제품의 실 사용자들에 초점을 맞추는 사용자 테스트 또는 실증적 평가에 해당되는 것으로, 평가될 시스템을 사용하는 동안 얻어진 사용자 행태 (user behaviour)의 분석에 바탕을 둔다.

2.1 전문가에 의한 평가 방법들

“소프트웨어 검사 (software inspection)”는 코드(code)를 디버깅(debugging)하고 개선하기 위한 방법으로써 소프트웨어 공학의 중요한 단계 중의 하나이다. 이와 유사한 개념으로써 비용 면에서 효과적으로 사용성 문제들 (usability problems)을 발견하기 위한 사용자 인터페이스 평가 방법들을 “사용성 검사 (usability inspection)”라고 부른다 [6]. <표 2>는 사용성 검사 방법들을 요약한 것이다. 이 방법들은 비정형화된 (informal) 방법으로써 주로 개발자, 시스템 분석자, 그리고 인간공학 전문가에 의해 수행된다. 비록 다른 방법들이 사용성 문제들의 심각성 (severity)이나 전체 디자인의 사용성 측정과 같은 내용들을 다룬다고 할지라도, 전형적으로 사용성 검사의 목적은 인터페이스 디자인에서 사용성 문제들을 발견하는데 있으며, 사용성 공학 수명 주기의 초기에 수행할 수 있는 방법이다.

2.2 사용자에 의한 평가 방법

사용자에 의한 평가에서 일반적으로 이용하는 방법들은 사용자의 행위, 내용, 생각, 그리고 시스템의 반응들을 관찰하거나 기록하는 것이다. 사용자들을 활용하여 사용성을 평가하기 위한 방법들은 <표 3>과 같이 요약될 수 있다. 사용자가 시스템과 상호작용할때 사용자의 수행도 (performance)를 측정하는 방법은 가장 널리 사용되는 것으로 보통 수행시간, 효율성, 오류시간, 오류율 등의 척도가 사용된다[1]. 이러한 수행도 척도는 매우 다양하며, 각 제품의 특성에 따라 적절한 척도를 선별(screening)하는 노력이 필요하다. 시간은 측정이 쉽고 통계적 분석이 용이하며, 오류 분석을 통해 시스템 내의 문제점을 규명하고 사용상 어려움의 원인을 파악할 수 있다. 그러나, 수행도를 측정하는 방법은 측정할 시스템이나 원형이 필요하며, 실사용자들을 잘 대표하는 피실험자의 선정과 적절한 피실험자의 수를 결정하는 문제가 있다 [10].

설문지(questionnaires)나 면담(interview)을 통한 분석 방법은 사용자의 반응을 평가 할 수 있는 직접적인 방법이다. 그러나 평가 결과에 대한 편의(bias) 문제와 많은 의문이 생길 수 있으므로 설문지의 항목을 구성하는데 주의 깊은 고려와 함께 타당성 검증이 필요하다. 사용자가 시스템을 사용할때 필요한 정신적 노력과 스트레스의 측정은 여러 생리학적 장비 및 도구를 이용하거나 사용자가 느낀 것을 주관적으로 측정한다 [1]. 그러나 실험 환경이 실제 사용자가 사용하는 환경으로 구성되도록 하는 노력이 필요하며, 정신적인 노력을 측정하고 분석하는 데 있어서 상당한 주의가 요구된다.

<표 2> 사용성 검사 방법들의 요약

Inspection Methods	평가자	설 명
heuristic evaluation	S	사용성 전문가들의 판단을 이용하여 각 인터페이스 요소들이 사용성 원칙들을 잘 따르는가를 평가하는 비정형화된 방법
cognitive walkthroughs	S	사용자의 목표와 기억 내용이 다음 행동을 올바르게 이끌 수 있는지 체크하고, 각 단계에서 사용자의 문제 해결을 모의 실험하기 위해 정확하고 상세한 절차를 이용하는 방법
formal usability inspections	C	heuristic evaluation과 cognitive walkthroughs의 단순화된 형태를 결합한 다 단계 절차를 이용하는 방법
pluralistic walkthroughs	G	사용자들, 개발자들, 인간공학 전문가들이 모여서 주어진 시나리오를 가지고 각 인터페이스 요소를 토론하는 방법
feature inspection	S	긴 작업 순서들, 다루기 힘든 단계들, 사용자들이 시도하기에 부자연스런 단계들, 과도한 지식과 경험이 요구되는 단계들을 체크하기 위하여, 작업 수행시 사용되는 시스템의 특징들(features)을 기술하는 방법
consistency inspection	G	다중 프로젝트들을 대표하는 설계자들을 이용하여 한 인터페이스가 그들 자신의 디자인들처럼 같은 방식으로 운영되는지 검사하는 방법
standards inspection	S	각 인터페이스가 인터페이스 표준들을 잘 따르는지 검사하기 위해 관련 전문가를 이용하는 방법

S: a single evaluator at a time

G: group inspection methods

C: combine individual and group inspections

<표 3> 사용자에게 의한 평가 방법들의 요약

사용자에게 의한 평가 방법들의 분류	평가 방법들의 예
시스템과 상호작용할때 사용자의 수행도 (performance) 측정	Performance measures (time, error, efficiency)
시스템에 대한 사용자 태도(user's attitude)	평점 척도(rating scales) 주관적 측정, SD 법
시스템을 사용할때 필요한 정신적 노력(mental effort) 또는 시스템 사용시 발생하는 stress의 측정	Physiological analysis 주관적 평가, TLX

3. 사용성 평가 방법들의 비교

Jeffries et al. (1991)은 휴리스틱 평가, 소프트웨어 가이드라인, 인지적 시연과 사용성 테스트 등 네 가지 방법들의 상대적인 장점을 논의하고 각 방법들의 개선 안을 제시하였다 [3]. Karat et al. (1992)는 사용성 문제들을 발견하는 상대적 효율성의 관점에서 실증적 테스트(empirical testing)과 팀 시연(team walkthrough) 방법들을 비교하고, 실증적 테스트를 통해 시연 방법에서 발견하지 못한 상대적으로 중요한 문제들을 확인할 수 있음을 보였다 [4].

<표 4> 사용성 방법들의 요약 (Summary of the usability methods)

사용성 방법	수명 주기 단계	필요한 사용자수	장 점	단 점
heuristic evaluation	초기 디자인, 반복 설계의 "내부 사이클"	필요 없음	낮은 비용. 각 사용성 문제점들을 발견하고 전문 사용자의 의견을 수렴.	실 사용자들과 연관이 없으므로, 실 사용자의 요구들과 관련된 주요 문제점들을 발견 못함.
performance measures	경쟁적 분석, 최종 테스트	적어도 10명	명백한 값을 얻을 수 있고, 결과들을 비교하기 쉽다.	높은 비용. 각 사용성 문제점들을 발견할 수 없다.
thinking aloud	반복 설계, formative evaluation	3-5명	사용자의 잘못된 개념들을 지적할 수 있고, 비용이 저렴하다.	사용자들의 작업수행이 부자연스럽고, 전문 사용자의 경우 언어화하기 어렵다.
observation	직무 분석, 추후 연구들	3명 이상	사용자의 실제 직무들을 밝힐 수 있고, 기능들과 특징 등을 제시.	셋업하기 어렵고, 실험자 통제가 곤란.
questionnaires	직무 분석, 추후 연구들	적어도 30명	주관적인 사용자 선호도를 발견할 수 있고, 반복이 용이.	이해 부족을 막기 위해 예비 작업이 필요.
interviews	직무 분석	5 명	유연하고 상세한 사용자 의견과 태도.	시간 소비적이고, 분석과 비교가 어려움.
focus groups	직무 분석, 사용자 참여	그룹당 6-9명	자연스러운 반응들과 활발한 그룹 의견	분석하기 어렵고, 낮은 타당성.
logging actual use	최종 테스트, 추후 연구들	적어도 20명	빈번히 또는 사용되지 않는 특징들을 발견할 수 있고, 연속적으로 실행 가능.	거대한 양의 자료를 분석하기 위한 분석 프로그램이 필요하고, 사용자의 프라이버시 침해.
user feedback	추후 연구들	100명 이상	사용자 요건들과 관련점들의 변화를 파악.	의견을 얻는데 필요한 특별 조직이 구성되어야 함.

Nielsen et al. (1993)은 혜택-비용 분석을 통해서 휴리스틱 방법, 정형화된 방법 (GOMS), 사용자 테스트 방법들을 비교하였다 [7]. 그들은 사용자 테스트 방법이 저렴한 휴리스틱 방법에 비해 4.9배나 비용이 소요되지만 더 나은 수행도 예측을 제공한다고 하였다. 이러한 연구 결과들을 종합해 볼 때 사용성 평가 방법들간에는 각각의 장단점이 있으므로 평가 전략의 선택에 신중을 기해야 함을 알 수 있다. <표 4>는 제품의 사용자 인터페이스 사용성 평가시 이용되는 방법들을 요약하고 장단점을 비교한 것이다. 각 방법들은 수명 주기의 서로 다른 단계에 응용되며, 방법들의 장점 및 단점들이 부분적으로 각각의 방법에 대응될 수 있기 때문에, 사용성 방법들은 상호 보완적인 경향이 있다. 따라서 사용성을 평가할 때 하나의 방법만을 고수하는 것은 별로 바람직하지 않다.

한편 사용성 방법의 선택은 부분적으로 평가에 활용되는 사용자들의 수에 의존한다 [5]. 가능한 적은 수의 사용자들이 활용되어야 하는 경우에는 휴리스틱 평가 (heuristic evaluation), 독백 기록 (thinking aloud), 그리고 관찰 (observation) 방법이 바람직하다. 약간 많은 수의 사용자들이 활용되는 경우에는 수행도 측정 (performance measurement)과 초점 그룹 (focus groups)을 이용하는 것이 가능하고, 매우 많은 수의 사용자들이 활용되는 경우에는 설문지 (questionnaires), 상호작용 기록(interaction logging), 그리고 자연스러운 사용자 피드백 (user feedback)의 체계적 수집 방법 등이 고려될 수 있다. 물론, 방법들을 약간 변형하는 경우 활용되어야 할 사용자의 수도 변한다. 예를 들어, 상호 협동적 상호작용 (collaborative interaction)은 독백 기록 (thinking aloud) 방법보다 두 배의 사용자들을 필요로 한다. 그러나 사용성 평가시 사용자들에게 보다 자연스러운 작업 환경을 제시할 수 있는 장점이 있다. 한편 평가에 활용될 수 있는 사용성 전문가의 경험 수준 또한 방법의 선택에 큰 영향을 미친다.

일반적으로 사용성 평가시 하나의 방법만으로는 불충분하므로 여러 방법들을 결합해서 사용한다. 다양한 사용성 방법들을 결합하는 데는 여러 가지 가능한 조합 방법들이 있을 수 있으며, 각 프로젝트의 특성들에 따라 약간 다른 조합도 가능하다. 일반적으로 유용하다고 알려진 조합은 휴리스틱 평가와 독백 기록 (thinking aloud) 또는 사용자 테스트 (user testing)의 형태이다. 보통 처음에는 인터페이스에 존재하는 명확한 사용성 문제점들을 발견하고 제거하기 위하여 휴리스틱 평가를 수행한다. 인터페이스에 대한 재설계 (redesign)가 이루어진 후에는, 휴리스틱 평가에 의해 지적되지 않은 남아 있는 사용성 문제점들을 발견하고, 반복적 설계 단계에서 결과들을 점검하기 위하여 사용자 테스트 (user testing)을 수행한다. 이러한 두 가지 방법들은 서로 다른 사용성 문제점들을 발견하는 것으로 보여지며, 상호 보완적인 관계를 가진다 [3, 4]. 결론적으로 어떠한 평가 방법들을 이용할 것인가 하는 문제는 가용 시간과 주어진 예산과 같은 외부적 요인들에 의해 제약을 받는다. 또한 평가자들의 경험과 지식 수준, 제품의 수명주기의 단계, 사용자 특성 등과 같은 요인들도 평가 전략을 결정하는데 영향을 미친다.

4. 제품 수명주기를 고려한 사용성 평가 방법들

제품의 사용성을 보증하고 유용한 인터페이스를 설계하기 위해서는 제품의 수명주기에 따른 체계적인 사용성 노력이 필요하다. 인간공학적 수명주기 (human factors lifecycle)를 통해 제품의 설계 및 개발 단계로 쉽게 통합될 수 있는 실제적이고 체계적인 사용성 노력을 기울일 수 있어야 한다. <표 5>는 인간공학적 수명주기와 사용성 평가 방법들을 결합함으로써 체계적인 사용성 노력을 기울이기 위한 틀 (framework)을 나타낸다. 결론적으로 사용성 방법의 성공을 보장하기 위하여 위에서 언급한 사용성 방법을 적용하기 위한 방법 (metamethods)을 정하는 것이 중요하다. 예를 들어 방법을 사용할 때 무엇을 할 것인가에 대한 명백한 계획을 세우거나, 실증적인 사용자 검증에 대한 계획을 세울 필요가 있다. 가능한 프로젝트 초기에 수명주기 동안에 수행될 사용성 활동을 열거하는 전체적인 사용성 계획을 세우는 것이 좋을 것이다.

<표 5> 체계적인 사용성 노력을 위한 틀

Human Factors Lifecycle	Usability Engineering Effort	Methods for Usability Measurement and Evaluation
<input type="checkbox"/> Market Analysis	▷ Consider the larger context ▷ Know the user	• concept test, • interviews • psychological scales • factor analysis • comparative analysis
<input type="checkbox"/> Feasibility Study	▷ Competitive analysis	• usability specification
<input type="checkbox"/> Requirement Definition	▷ Setting usability goals	• HTA, GTN, OSD, GOMS, • TAG, CCT, link analysis etc. • Standards • expert review, checklist • heuristic evaluation
<input type="checkbox"/> Product Acceptance Analysis	▷ Participatory design	
<input type="checkbox"/> Task Analysis	▷ Coordinated design of the total interface ▷ Guidelines and heuristic analysis	
<input type="checkbox"/> Global Design		
<input type="checkbox"/> Prototype Construction	▷ Prototyping	• simulation trials • iterative laboratory experiments • thinking aloud • constructive interaction • walkthroughs with actual users • autologging • user's attitude, questionnaire • physiological analysis • TLX
<input type="checkbox"/> User Testing and Evaluation	▷ Empirical testing	
<input type="checkbox"/> System Implementation		
<input type="checkbox"/> Product Testing	▷ Empirical testing	
<input type="checkbox"/> User Testing		
<input type="checkbox"/> Update and Maintenance	▷ Iterative design	• impact analysis • design rationale • follow-up study
<input type="checkbox"/> Product Survey	▷ Collect feedback from field use	• field study • hot line

5. 결론

제품 설계 및 개발시 상호 경쟁적인 상황에서 두 번째 주자가 되어 시장 기회를 놓치는 것 보다는 오히려 적은 문제들을 포함하는 제품을 출하하는 것이 더 나을 수 있다. 이것은 주어진 조건들 하에서 가능한 빠르고 저렴한 평가 방법들이 제품의 사용성 평가에 이용되어야 한다는 것을 의미한다. 그러나, 어떠한 평가 방법들을 이용할 것인가 하는 문제는 가용 시간과 주어진 예산과 같은 외부적 요인들과 평가자나 디자인 팀의 경험과 지식 수준, 제품의 수명주기의 단계, 사용자 특성 등과 같은 여러 요인들에 의해 제약을 받는다. 본 연구에서는 효율적인 사용성 평가 방법들의 선택을 위해서 사용성 평가 방법들을 분류하고, 각 평가 방법들의 장점 및 단점을 비교하였다. 또한 제품 수명주기와 사용성 평가 방법들을 결합함으로써 체계적으로 제

품의 사용성 보증 (usability assurance)을 할 수 있는 틀 (framework)을 제시하였다. 이러한 결과는 주어진 예산과 시간 제약 하에서 제품 수명주기에 가장 적합한 평가 전략 수립에 도움을 줄 것으로 기대된다.

참 고 문 헌

- [1] Began, N., and Macleod, M., *Usability specification and measurement*, National Physical Laboratory, Teddington, UK, 1993
- [2] Billingsley, P. A., " Ergonomic Standards Go Beyond Hardware", *IEEE Tr. on Software*, pp.82-84, March, 1994
- [3] Jeffries, R., Miller, J. R., Wharton, C., and Uyeda, K. M., User Interface Evaluation in the Real World: A Comparison of Four Techniques, *In Proceedings of CHI'91*, ACM, pp.119-124, 1991
- [4] Karat, C., Campbell, R.,and Fiegel,T., Comparison of Empirical Testing and Walkthrough Methods in User Interface Evaluation, *In proceedings of CHI'92*, ACM, pp.397-404, 1992
- [5] Nielsen, J., *Usability Engineering*, Academic Press, 1993
- [6] Nielsen, J., Usability Inspection Methods, *In proceedings of CHI'94*, ACM, pp.413-414, 1994
- [7] Nielsen, J. and Phillips, V. L., Estimating the Relative Usability of Two Interfaces: Heuristic, Formal, and Empirical Methods Compared, *In proceedings of CHI'93*, ACM, pp.214-221, 1993
- [8] Perlman, G., Practical Usability Evaluation, *In Proceedings of CHI'94*, ACM, pp.407-408, 1994
- [9] Shackel, B., and Richardson, S., *Human Factors for Informatics Usability*, Cambridge Univ. Press, 1991
- [10] Virzi, R. A., Refining the Test Phase of Usability Evaluation: How Many Subjects Is Enough?, *Human Factors*, 34(4), pp.457-468, 1992
- [11] 임치환과 이민구, 정보기술 제품의 사용성 평가 방법, 공업경영학회지, 제16권 28호, 1993
- [12] 이민구와 임치환, 제품 수명주기에 의한 사용성 공학 모형의 적용, 공업경영학회지, 제17권 30호, 1994