

특집

소프트웨어의 사용편의성 평가

한 성 호[†] 곽 지 영^{††} 홍 상 우^{†††}

◆ 목 차 ◆

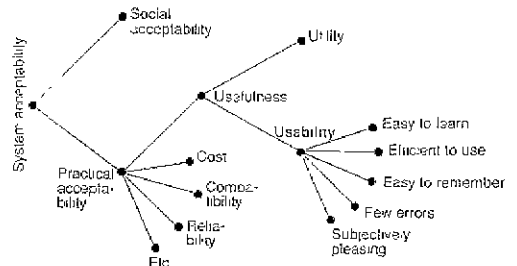
- | | |
|----------------|---------------|
| 1. 서 론 | 3. 적용 사례 |
| 2. 사용편의성 평가 체계 | 4. 결론 및 추후 연구 |

1. 서 론

과거에는 컴퓨터가 소수의 전문가에 의해서만 사용되어 사용편의성(Usability)이 열악하더라도 다양한 기능을 갖는 소프트웨어가 선호되는 것으로 인식되어 왔다. 그러나, 최근에는 사용자 층이 일반인이나 어린이에게까지 확대되었고, 사용편의성에 대한 인식이 높아지면서 사용자 중심으로 잘 설계된 소프트웨어가 아니면 시장성을 잃게 되는 실정이다. 이와 같이 사용편의성이 무엇보다 중요한 구매요건으로 등장하게 되면서 각 개발사에서는 사용편의성 향상을 위한 많은 노력이 이루어지고 있다. 외국회사의 경우 평균적으로 소프트웨어 개발비용의 48%~100%를 사용자 인터페이스 개발에 투자하고 있다(Myers and Rosson, 1992).

이처럼 소프트웨어 구매결정의 중요한 요인이 된 사용편의성이란 주어진 환경에서 사용자가 얼

마나 빨리, 효율적이고, 쉽게, 주어진 작업을 수행할 수 있는가 하는 정도를 의미한다. 따라서, 사용편의성이란 기존의 시스템 중심의 설계에서 탈피한 사용자 중심의 설계개념의 도입이라 할 수 있다. [그림 1]은 제품에 대한 소비자의 수용성(Acceptability)의 관점에서 사용편의성 조건들을 보여주고 있다. 결과적으로 사용편의성을 구성하는 학습의 용이성(Easy to learn), 사용상의 효율성(Efficient to use), 기억의 용이성(Easy to remember), 작업오류의 최소화(Few errors), 그리고, 제품에 대한 만족도(Subjectively pleasing) 등이 제품에 대한 수용성에 총체적인 영향을 미치는 것으로 파악되고 있다.



(그림 1) 사용편의성의 구성요소 (Nielsen, 1993)

† 정회원 : 포항공과대학교 산업공학과 교수

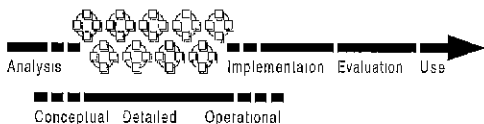
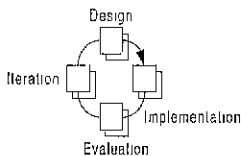
†† 정회원 : 포항공과대학교 산업공학과 박사과정

††† 정회원 : 포항공과대학교 산업공학과 석사과정

사용편의성이 낮은 제품은 다음과 같은 5가지의 직접적인 원인에 기인한다고 할 수 있다(Rubin, 1994).

- 사용자 중심이 아닌 제품의 기능에 중점을 둔 설계
- 소비자의 요구조건을 제품설계에 체계적으로 반영하지 못함
- 사용편의성에 대한 고려가 중요시 되지 않은 풍조
- 제품개발 참여 인원 간의 통합성 결여
- 인터페이스의 사용편의성 평가시 전문가적 Knowhow가 필요

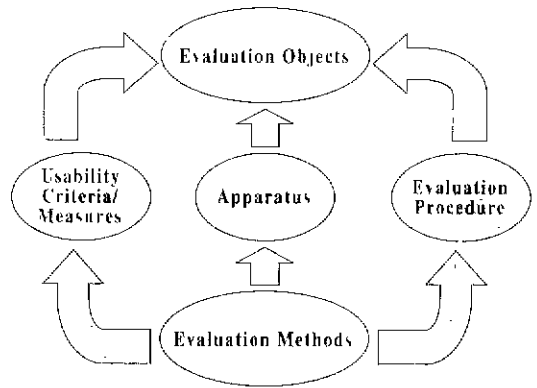
종합하면, 소프트웨어의 사용편의성 제고를 위해 무엇보다도 중요한 점은 설계자에 의한 임의적 설계로 만족하여서는 안되며, 개발단계별로 적절한 사용편의성 평가가 이루어져, 인터페이스를 지속적으로 개선해가야 한다는 것이다. [그림 2]와 같이, 제품설계단계에서는 제품의 Prototype을 제작하여 평가-수정하는 과정을 반복하게 되며, 개발이 거의 완료된 시기에는 사용자를 대상으로 하는 실험을 수행하거나 전문가 의견을 토대로 한 Inspection 등을 실시하여, 제품개발이 완료된 후에 심각한 문제점이 발견되는 것을 최대한 방지하여야 한다.



(그림 2) Prototype을 이용한 평가-수정 Cycle

2. 사용편의성 평가 체계

본 연구에서는 소프트웨어의 사용편의성 평가를 보다 체계적으로 수행하기 위해 사용편의성 평가체계를 구성하였다. 사용편의성 평가는 평가 기법과 평가대상, 평가 기준 등이 [그림 3]과 같은 상호 유기적인 연관관계를 갖는 것으로 나타낼 수 있다.



(그림 3) 사용편의성 평가 체계

[그림 3]에 나타난 바와 같이, 사용편의성의 평가는 평가 기법이 결정되면 그에 따른 평가 기준이 채택되고 일정한 평가장비나 도구를 이용하여 설정된 절차에 따라 특정 평가대상에 대한 평가가 실시되는 것으로 설명이 가능하다. 이러한 사용편의성 평가 체계에 요구되는 평가 대상, 평가 기법, 평가 기준 등은 사용편의성 평가에 종합적으로 영향을 미치게 되는 매우 중요한 구성요소이다. 본 연구에서 개발된 각 구성요소에 대한 정의와 체계에 대한 설명은 다음과 같다.

2.1 평가 기법(Evaluation Methods)

소프트웨어의 평가기법은 현재 약 30여종이 제안되고 있으며, 이들 기법들은 사용 목적이나 진행 절차, 적용 시기 등에서 큰 차이를 보이고 있

다. 본 연구에서는 이들 평가기법들을 크게 Usability Inspection과 Usability Testing, 그리고 이들 두 그룹에 해당되지 않는 기타 기법의 세 그룹으로 분류하였다.

2.1.1 사용편의성검사(Usability Inspection)

사용편의성 검사는 실험을 수행하지 않고 일반적인 제품에서 결함을 찾아내듯 소프트웨어를 검토하여 문제점을 발견해 내는 방법이다 (Nielsen & Mack, 1994). 사용편의성 검사에 해당되는 평가기법은 다음과 같다.

- 경험적 검사 (Heuristic Evaluation)
각 평가 대상들이 Usability Principle에 따라 적절히 설계되었는지의 여부를 사용편의성 전문가의 판단으로 평가하는 기법이다. 가장 적은 비용과 시간으로 평가 대상이 갖고 있는 대부분의 사용편의성 문제점들을 파악할 수 있어 범용적으로 사용된다.
- 설계 지침에 따른 검사 (Guidelines Review)
사용편의성에 대한 여러 설계 지침 (Guideline)을 이용하여, 각 평가 대상이 설계 지침에 위배됨이 없는지를 평가하는 기법이다.
- 복합적 검사 (Pluralistic Walkthrough)
사용자, 개발자 그리고 인간공학 전문가가 모여 사용편의성에 대한 여러 가지 쟁점을 논의하여 평가하는 방법이다.
- 일관성 검사 (Consistency Inspection)
소프트웨어 인터페이스에 대한 일관성을 집중적으로 검토하는 방법으로, 여러 소프트웨어가 하나의 제품을 형성하고 있는 경우에 주로 사용되는 기법이다.
- 표준 검사 (Standard Inspection)
인터페이스 전문가가 일정한 표준에 따라 인터페이스를 검사하는 기법이다.
- 인지적 검사 (Cognitive Walkthrough)
사용자의 문제 해결 과정을 가상적으로 미리

생각하여, 작업 수행 상의 문제점을 예측해 내기 위한 방법이다.

- 기능 검사 (Feature Inspection)
시스템의 여러 Feature가 사용자의 요구에 맞게 설계되었는지를 평가하는 방법이다.

2.1.2 사용편의성 실험 (Usability Testing)

사용편의성 실험은 사용자를 동원하여 직접 실험을 수행함으로써 실제 사용상의 문제점이나 결함을 발견해 내는 평가방법으로서 정형 실험 (Formal Testing)과 비정형 실험 (Informal Testing)으로 크게 구분된다. 정형 실험은 개발이 거의 완료된 시점에서 실험실 환경의 엄격한 통제하에 실시되는 방법이다. 비정형 실험은 문제점 파악이나, 타당성 검토, 비교 등의 목적으로 비교적 통제를 완화하여 이루어지는 것으로, 개발의 초기 단계에서부터 개발 완료 단계까지 고루 적용이 가능하다. 사용편의성 실험에 해당되는 평가기법을 나열하면 다음과 같다 (Nielsen, 1993; Rubin, 1994).

- 사고 발생법 (Thinking Aloud)
실험이 진행되는 동안 사용자가 자신의 생각과 느낌을 말로 표현하게 하는 방법으로 관찰자는 사용자가 현재 무엇을 의도하고 있는지, 그리고 사용자들이 왜 그 일을 하고 있는지를 파악할 수 있다.
- 관찰 (Observation)
가장 간단한 사용편의성 실험 방법으로 실제 현장에 소프트웨어를 설치해 두고 사용자를 관찰하여 사용편의성에 대한 정보와 작업 분석을 실시하는 방법이다.
- 집중 토의 (Focus Group Research)
약 6-9명의 사용자로 Focus Group을 형성하여 새로운 개념을 토론하거나 문제를 인지하기 위한 회의를 하는 방법이다. 주로 인터페이스가 구체적인 설계에 들어가기 전 프로토타입을 얼마동안 사용한 후에 사용자의 요구

사항과 사용상의 느낌 등을 평가하는데 사용될 수 있다.

- 사용 기록 평가 (Logging Actual Use)
컴퓨터가 자동적으로 시스템의 자세한 사용에 대한 통계를 수집하는 것으로 사용자의 키 조작이나 마우스 조작 등의 모든 행위를 자동으로 기록하여, 보다 상세한 데이터를 수집하기 위해 사용된다.
- 스케치 평가 (Paper & Pencil Evaluation)
컴퓨터를 이용한 프로토타입이 개발되기 전인 설계의 극히 초기단계에 연필과 종이를 이용하여 시스템을 개략적으로 표현해 보임으로써 평가를 수행하는 방법이다.
- 대표 작업 실험 (Benchmark Testing)
시스템에서 실제 사용자가 수행하게 될 대표적인 작업을 **Benchmark Task**로 선정하여 사용자에게 수행해 보도록 함으로써 시스템 사용상의 문제점을 파악하기 위한 평가 방법이다.
- 자유 작업 실험 (Hands-On Experiment)
사용자에게 시스템을 자유롭게 사용해 보도록 하여 사용자가 발견해낸 문제점을 보고하도록 하는 방법이다.

2.1.3 기타

사용편의성 검사와 사용편의성 실험 이외에 해당 소프트웨어를 직접 사용하고 있는 사용자의 반응을 조사하여 소프트웨어를 평가할 수 있으며, 이에 해당하는 평가 기법을 나열하면 다음과 같다 (Nielsen, 1993; Rubin,1994; Shneiderman,1992).

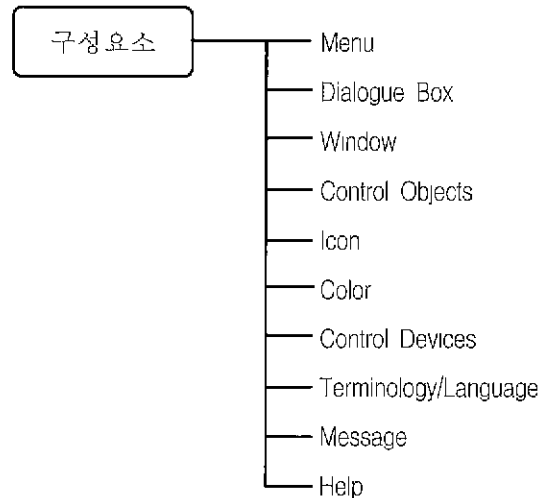
- 의견 조사 (Survey)
 - Questionnaire
 - Interview
- 사용자 의견 피드백 (User Feedback)
 - Online or Telephone Consultant
 - Online Suggestion Box
 - Online Bulletin Board

- User Newsletters & Conferences

- 사용편의성 감사 (Usability Audit)
- 필드 조사 (Field Study)
- 추후 조사 (Followup Study)

2.2 평가 대상(Evaluation Objects)

본 연구에서는 사용편의성을 위한 평가대상을 [그림 4]와 같이 10가지로 분류하였다. 이들은 인터페이스를 구성하는 기본적인 구성 요소(Interface Object)들로서, 어떤 종류의 소프트웨어에도 적용이 가능한 분류체계라 할 수 있다.



(그림 4) 구성 요소 분류체계

그러나, 이들은 소프트웨어 인터페이스의 개략적인 구성요소에 불과하며, 실제로 평가를 실시하게 되는 평가대상은 각 구성요소의 특성(Property)에 해당한다. 따라서, 본 연구에서는 10가지 기본 구성요소에 대한 40여 개의 특성을 평가 대상으로 선정하였다. 이때, 각 구성요소의 외형, 즉, 정적인 측면 뿐 아니라, 사용자와의 상호작용(Interaction)과 관련된 동적인 요소까지 모두 고려되어야 한다. 구성 요소 별로 평가되어야 하는 평가 대상이 정적인 측면과 동적인 측면

으로 분류되어 (표 1)에 정리되어 있다.

(표 1) 구성 요소 별 평가 대상

구성 요소	구분	평가 대상
Menu	Static	Naming
		Grouping
		Layout
		Structure
		Style
		Shortcut
	Dynamic	Activation
		Navigation
		Scrolling
Dialogue Box	Static	Layout
		Grouping
		Highlight
	Dynamic	Default Choice
Window	Static	Layout
		Background
		Title
		Border
		Solid Line
		Navigation
	Dynamic	Function
		Shortcut
		Selection
		Feedback
Icon	Static	Layout
		Metachor
		Size
		Representation
Dynamic	Selection	
	Activation	
	Feedback	
Color	Static	Contrast
		Quality
		Number
		Representation
Control Device	Static	Button Assign
	Dynamic	Action
Terminology	Static	Abbreviation
		Representation
Message	Static	Content
		Representation
		Modality
Help	Static	Content
	Dynamic	Activation
		Search

2.3 평가 기준 (Evaluation Criteria)

평가 기법과 평가 대상이 선정되면, 이들을 평가하기 위한 구체적인 기준이 제시되어야 한다. 소프트웨어의 사용편의성 평가에 사용되는 평가 기준은 정량적 평가 기준과 정성적 평가 기준으

로 구분된다.

2.3.1 정량적 평가 기준

정량적인 평가(Qualitative Evaluation)에서는 작업수행속도나 정확도 등과 같은 대표적인 몇 가지 측면에 대한 정량적인 측정치를 파악하여 이를 기준으로 평가가 이루어진다. 정량적인 평가를 위해 주로 사용되는 평가기준을 나열하면 다음과 같다 (Hix and Hartson, 1993).

- 주어진 작업 수행에 소요되는 시간
- 주어진 시간 내에 완수된 작업의 수
- 성공적인 작업수행 횟수와 오류의 비율
- 오류를 복구하는데 소요된 시간
- 사용자가 범한 오류의 수
- 잇달아 발생하는 오류의 수
- 사용자가 사용한 적이 없었던 명령어나 시스템 모양, 특징의 수
- 실험 후에 사용자가 기억해낼 수 있는 시스템 모양, 특징의 수
- 설명서와 도움말 기능의 사용빈도와 이들을 사용하는데 소요되는 시간
- 설명서와 도움말이 사용자의 문제를 해결하여 주는 빈도
- 시스템에 대한 긍정적인 평가와 부정적인 평가의 비율
- 사용자가 작업수행 도중에 느낀 좌절이나 혼란의 표현 횟수
- 다른 경쟁사의 제품이 사용하기에 더 좋다고 말하는 사용자의 수
- 사용자가 해결할 수 없었던 문제의 수
- 한 작업을 수행하는데 있어서 여러 가지 방법이 있을 경우, 비효율적 전략을 사용하는 사용자와 효율적인 전략을 사용하는 사용자의 비율
- 사용자가 시스템과 상호작용하지않은 시간
- 사용자가 수행하던 작업으로부터 벗어나게

되는 횟수

2.3.2 정성적인 평가 기준

소프트웨어에 대한 정성적 평가에는 현재 제시되어 있는 설계원칙이나 설계지침 (General Principles and Guidelines) 등이 평가기준으로 활용될 수 있다. 파악된 정성적 평가 기준은 다음과 같다.

- 사용자 오류 방지 (Prevent User Errors)
- 작업의 최적화 (Optimize User Operations)
- 사용자 통제 기능 강화 (Keep the Locus of Control with the User)
- 일관성 (Consistency)
- 단순성 (Simplicity)
- WYSIWYG
- 직접성 (Directness)
- 관용성 (Forgiveness)
- 즉각적 피드백 (Immediate Feedback)
- 상세성 (Informativeness & Specificity)
- 미적 통합성 (Aesthetic Integrity)
- 적응성 (Adaptability)

(표 2) 평가 Matrix 구성 사례

		Consistency	Simplicity	Directness	Forgiveness
Menu	Style	○	○	X	X
	Layout	○	○	X	X
	Structure	○	○	X	X
	Naming	○	○	X	X
	Grouping	○	○	X	X
	Navigation	○	○	○	○
	Scrolling	○	○	○	X
	Selection	○	○	○	○
	Shortcut	○	○	X	X
Window	Border	○	X	X	X
	Layout	○	○	X	X

다음으로, 각 평가 대상별로 우선적으로 적용되어야 할 평가기준을 찾아내기 위해 평가기준에 대한 우선 순위를 책정하였다. 마지막으로 **Brainstorming**을 통해 각 평가대상에 대해 반드시 적용되어야 할 평가기준을 선별하여 [표 3]과 같은 최종 평가 Matrix를 작성하였다. [표 3]에서 검게 표시된 부분은 해당 평가대상의 평가기준으로 고려할 필요가 없음을 나타내며, 회색 표시된 부분에만 대해서만 검토가 필요함을 의미한다.

2.4 평가 Matrix 구성

앞에서 언급한 여러 평가기법 중에서 주로 전문가 의견에 의해 평가되는 사용편의성 검사 평가 기법을 이용할 경우, 평가 내용의 체계성이 없거나 평가에서 누락되는 항목이 생길 가능성이 크다. 따라서, 본 연구에서는 평가 대상과 각 평가 대상의 사용편의성 평가에 적용될 수 있는 정성적 평가 기준들을 평가 Matrix 형태로 재구성하여, 평가의 Checklist로서 활용하였다.

평가 Matrix는 먼저 평가 대상을 행에 배치하고, 평가기준을 열에 배치한 Matrix 형태로 구성되었다. 각 평가대상에 대해 해당 평가 기준의 적용 필요성을 판단하여, 적용이 필요한 Cell에 O표를, 불필요한 Cell에 X표를 각각 표시하여 [표 2]와 같은 Matrix를 구성하였다.

(표 3) 최종 평가 Matrix 사례

		Consistency	Simplicity	Directness	Forgiveness
Menu	Style				
	Layout				
	Structure				
	Naming				
	Grouping				
	Navigation				
	Scrolling				
	Selection				
	Shortcut				
Window	Border				
	Layout				

3. 적용 사례

본 연구에서 제시한 최종 평가 Matrix를 이용

하여 특정 소프트웨어를 대상으로 전문가 평가를 실시하였다. 또한, 전문가 평가 결과에 대한 타당성 확인과 결과의 보완을 위해 사용자를 이용한 실험을 동시에 실시하였다. 적용 대상이 된 소프트웨어는 Word Processor를 주된 기능으로 하며, 그래픽 처리, 데이터베이스를 위한 인터페이스 제작 기능을 부분적으로 갖춘 시스템이다.

3.1 Matrix를 이용한 전문가 평가

총 5명의 소프트웨어 사용편의성 전문가가 전점에서 언급한 평가 Matrix를 이용하여 대상 인터페이스에 대한 사용편의성 평가를 실시하였다. 먼저, 평가대상의 사용편의성 정도를 상, 중, 하로 Rating하였고, 그 중 '하'에 해당하는 Cell에 대하여 문제점을 기술하고 개선안을 제시하는 방법으로 평가를 진행하였다. 평가 Matrix를 이용한 평가 결과의 일부를 [표 4]에 제시하였다.

(표 4) 평가 Matrix를 이용한 결과 사례

		Consistency	Simplicity	Directness	Forgiveness
Menu	Style	상			
	Layout	중	중		
	Structure	하	히		
	Naming	중	상		
	Grouping	하	하		
	Navigation	심	상	상	
	Scrolling				
	Selection	하		상	
Window	Shortcut	중			
	Border	상			
	Layout	히	하		

평가 대상들 중 Help를 제외한 모든 대상에 대해 사용편의성 수준이 양호하지 않은 것으로 밝혀졌다. 특히, 메뉴의 경우, '하'의 평점을 받은 평가 기준이 가장 많았다. 평가 기준 측면에서는 "Directness", "Locus of Control with the User", "Immediate Feedback"은 비교적 잘 지켜지고 있었지만, 나머지 기준들은 잘 지켜지지 않았다.

전문가의견을 토대로 파악된 대상 제품 인터페이스의 문제점은 총 103개였으며, 각 대상별로 구분하여 보면 대화상자가 33개, 메뉴가 32개, 아이콘 19개 등으로 나타났다.

3.2 사용자를 이용한 실험 평가

사용자를 이용한 실험 평가는 아이콘 인지 실험, 대표작업을 이용한 실험, 자유 작업 실험 등으로 구성되었다.

아이콘 인지실험에서는 대상 소프트웨어에서 사용된 아이콘들을 사용자가 정확히 인지하는지 여부를 평가하기 위한 것으로, 총 87개의 아이콘 중 28%인 24개에 대해서는 적중률이 50%미만인 것으로 밝혀져 재설계 또는 변경이 요구되는 것으로 파악되었다. 또한, 사용자들의 응답결과는 설계자들에게 아이콘 인지도에 대한 중요한 참고자료로 활용되었다.

대표작업을 이용한 실험에서는 대상 소프트웨어 사용자가 수행하게 될 대표적인 작업을 대표 작업으로 선정하여 수행하게 하였다. 작업 수행 과정과 수행의 성공 여부를 추적하여 설계 적합성에 대한 평가가 가능하였다. 또한, 작업 수행 과정에서 사고 발생법 형식으로 사용상의 불편한 점이나 인터페이스의 문제점을 지적하도록 함으로써, 정성적 평가를 병행하였다.

자유 작업 실험에서는 사용자에게 20~30분간 대상 소프트웨어를 직접 사용해보도록 하여, 만족스럽지 못한 점이나 불편 사항을 지적하도록 하였다. 그 결과 총 70여가지의 문제점이 파악되었으며, 그 중 가장 많은 문제점이 지적된 평가대상은 메뉴(26개)인 것으로 파악되었다.

4. 결론

본 연구에서는 경험적 검사와 같은 사용편의성

검사 기법을 보다 체계적으로 수행할 수 있도록 하기 위한 사용편의성 평가 체계를 개발하였다. 우선 소프트웨어에서 사용편의성 평가가 필요한 평가 대상을 정의하였고, 평가 기준으로서 12개의 사용편의성 평가 원칙 (Usability Principle)을 정의하였다. 이들을 평가 Matrix의 형태로 구성하여 이를 실제 소프트웨어의 사용편의성 평가에 적용하였다. 그 결과, 불과 5명 정도의 전문가가 2시간 내에 100여가지의 문제점을 발견해 낼 수 있었다. 전문가 평가만으로 실제 사용자의 반응을 예측하는 데는 한계가 있으므로, 사용자를 이용한 실험을 병행한 결과 10여명의 피실험자로부터 약 70여가지의 문제점을 파악하였다. 이들을 전문가 평가에서 파악된 문제점과 비교한 결과, 대부분이 중복되는 것으로 나타났다.

결론적으로, 일정한 사용편의성 평가 체계에 의해 전문가 평가를 체계적으로 실시할 경우, 사용자를 이용한 실험 평가에 비해 효율적일 수 있다는 가능성을 암시하고 있다. 즉, 제한된 시간 내에 소수의 전문가에 의해 신속히 평가가 완료되어야 하는 경우, 사용자를 동원하지 않더라도 일정 수준 이상의 평가가 가능하다.

향후, 평가 대상, 즉, 소프트웨어의 구성 요소를 구체화하고 포함 범위를 확장할 경우, 보다 완성도 있는 평가 체계 구현이 가능할 것으로 기대된다.

참고문헌

[1] Hix, D., and Hartson, H. R., Developing User Interfaces: Ensuring Usability Through Product and Process, Wiley, 1993
 [2] Myers, B. A., and Rosson, M. B., Survey on User Interface Programming, Proceedings of CHI Conference on Human Factors in Computing System, 1992

[3] Nielsen, J., Usability Engineering, Academic Press, 1993.
 [4] Nielsen, J., and Mack, R. L., Usability Inspection Method, John Wiley & Sons, 1994
 [5] Rubin, J., Handbook of Usability Testing, John Wiley & Sons, 1994
 [6] Shneiderman, B., Designing the User Interface, Addison-Wesley, 1992



한 성 호

1983년 서울대학교 산업공학과 (학사)
 1985년 서울대학교 산업공학과 (석사)
 1991년 Virginia Polytechnic Institute & State University, Industrial & Systems Engineering (박사)
 1991-1992 IBM Tucson, Usability Lab (Research Assist)
 1992-현재 포항공과대학교 산업공학과 교수
 관심분야 : 인간공학, 감성공학, HCI, Usability Engineering, Human-Machine System Design and Evaluation



박 지 연

1993년 포항공과대학교 산업공학과 (학사)
 1995년 포항공과대학교 산업공학과 (석사)
 1995년-현재 포항공과대학교 산업공학과 박사과정 재학 중

관심분야 : Usability Engineering, Human-machine Systems Interface, Product Design and Evaluation Process, Usability of Consumer Electronic Products



홍 상 우

1996년 포항공과대학교 산업공학과 (학사)
 1998년 포항공과대학교 산업공학과 (석사졸업 예정)
 관심분야 : Human-Machine Interface, Usability of Consumer Electronic Products, Product Design and Evaluation