

Corso di Laurea in INFORMATICA

Interazione Uomo Macchina

Docente
M. F. Costabile
costabile@di.uniba.it

Metodi di valutazione di usabilità'

Questi lucidi sono stati preparati da Maria Francesca Costabile, Università degli Studi di Bari, per uso didattico. Essi contengono materiale originale di proprietà dell'Università degli Studi di Bari e/o figure di proprietà di altri autori, società e organizzazioni di cui è riportato il riferimento. Tutto o parte del materiale può essere fotocopiato per uso personale o didattico ma non può essere distribuito per uso commerciale. Qualunque altro uso richiede una specifica autorizzazione da parte dell'Università degli Studi di Bari e degli altri autori coinvolti.

Dipartimento di Informatica -
Università di Bari

1

1

Stili di Valutazione

- La valutazione di un sistema può aver luogo:
 - in **laboratori di usabilità**
 - nell'**ambiente** in cui gli utenti reali utilizzeranno il sistema
- Può essere effettuata:
 - da **esperti**
 - in collaborazione con **utenti reali**

2

Approcci alla Valutazione

1. **Test di usabilità**
2. **Studi sul campo**
3. **Valutazione Analitica**
 - **gli utenti non sono presenti**

Sharp et al., Interaction Design, Wiley, 2007

3

Alcuni Metodi di Valutazione

- **Valutazione basata sull'osservazione:** coinvolge utenti reali, osservati mentre interagiscono col sistema
- **Raccolta di informazioni (survey):** permette di ottenere informazioni sull'usabilità dei sistemi dal punto di vista dell'utente finale
- **Metodi analitici:** basati su modelli formali, applicabili a specifiche del sistema
- **Valutazione da parte di esperti:** sfrutta la conoscenza di esperti in HCI (anche considerata metodo analitico)

Alcuni approcci usano stessi metodi: test con utenti e studi sul campo sono entrambi basati su osservazione di utenti ma le condizioni e i modi in cui sono usati sono diversi

4

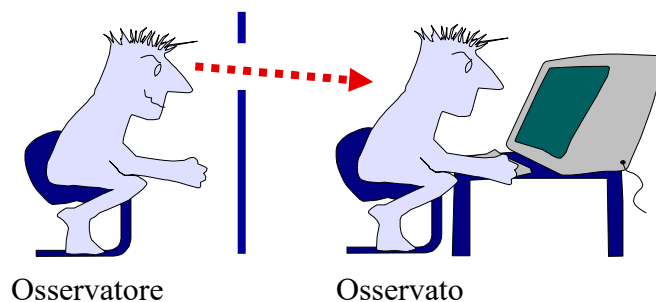
Scelta del Metodo di Valutazione

- **Fase del ciclo di sviluppo del sistema**
 - progetto *vs* implementazione
- **Stile di valutazione**
 - laboratorio di usabilità *vs* ambiente di utilizzo
- **Tipo di misure**
 - misure qualitative *vs* misure quantitative
- **Risorse richieste**

5

Test di usabilità

Utenti campione usano il sistema in un ambiente controllato, sotto osservazione da parte di esperti di usabilità che raccolgono dati, li analizzano e traggono conclusioni



R.Polillo - Ottobre 2010

6

6

Tipi di test di usabilità

- **Test di compito**

Agli utenti viene chiesto di svolgere **compiti specifici**, che permettano di esercitare le funzioni principali del Sistema

- **Test di scenario**

Agli utenti viene indicato un **obiettivo da raggiungere attraverso una serie di compiti elementari**, senza indicarli esplicitamente: l'utente dovrà quindi impostare una propria strategia di azioni

R.Polillo - Ottobre 2010

7

7

Esempio: test di compito (sito di e-commerce)

- **Task 1:** Registrarsi
- **Task 2:** Verificare se si può pagare con Visa e qual è l'importo minimo di un ordine
- **Task 3:** Verificare quali sono i tempi di consegna
- **Task 4:** Acquistare una scatola da 500 gr di tonno sott'olio
- **Task 5:** Cercare una confezione di sciroppo di acero
- **Task 6:** Verificare lo stato degli ordini effettuati
- **Task 7:** Verificare se esistono offerte speciali di pasta

R.Polillo - Ottobre 2010

8

8

Esempio: test di scenario (sito di e-commerce)

Scenario 1:

Domani sera hai due amici a cena, ma non hai tempo di andare al supermercato.
Decidi di fare la spesa on-line, pagando con la tua Visa.
Collegati al sito e ordina gli ingredienti per una cena veloce e poco costosa, ma simpatica.

R.Polillo - Ottobre 2010

9

9

Test di usabilità: logistica

Due filosofie:

- Usability lab (costi alti):
- Organizzazione informale (costi bassi)

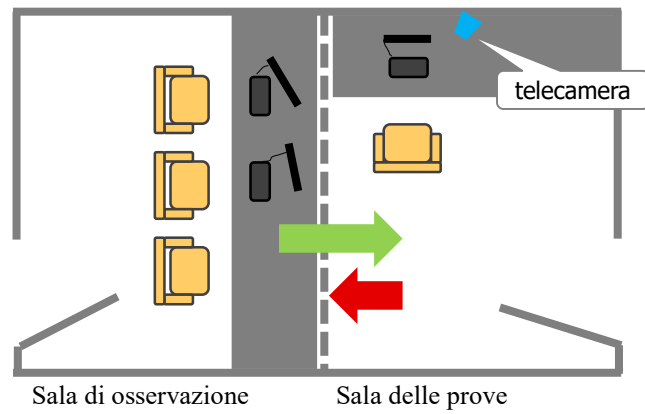
>>

R.Polillo - Ottobre 2010

10

10

Usability lab



R.Polillo - Ottobre 2010

11

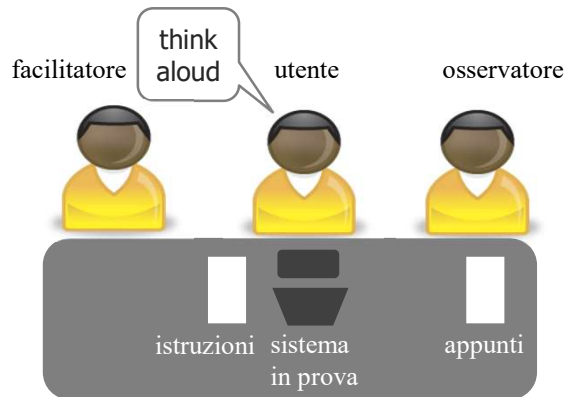
11



12

12

Organizzazione informale



R.Polillo - Ottobre 2010

13

13

Thinking aloud (o TA, pensare ad alta voce)...

- Protocollo di ricerca verbale nato nell'ambito delle scienze sociali
- Raccoglie dati qualitativi sui problemi incontrati dal partecipante durante il test
- Il partecipante esprime a voce alta i pensieri, i dubbi, le percezioni man mano che esegue i task
 - Azioni: "Ho scorso la pagina e sto leggendo la sezione XY che dice..."
"Ho cliccato il link home e sono arrivato all' homepage del sito in cui..."
 - Problemi: "Ho cliccato sul link home, ma il sistema riporta un errore XY"
"Credevo di trovare in questa pagina l'informazione XY, ma non riesco a trovarla"
- Limite: metodo di valutazione non naturale
 - Si richiede al partecipante di fare uno sforzo per verbalizzare tutto ciò che pensa

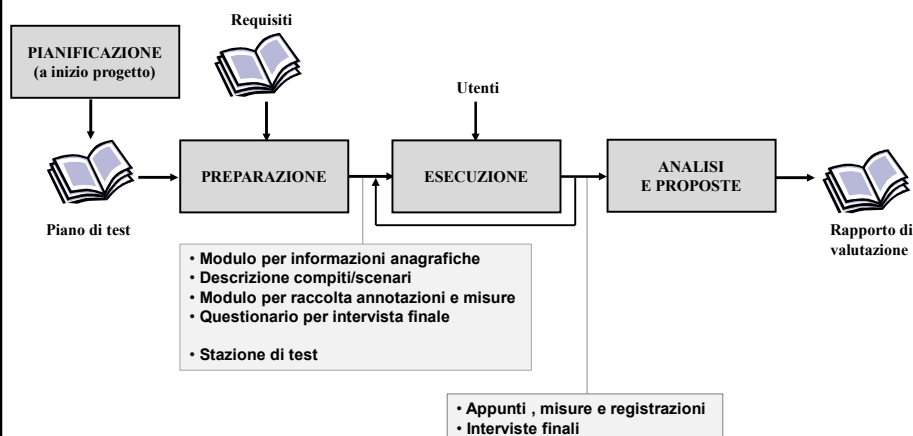
14

...Thinking aloud (o TA, pensare ad alta voce)

- Il conduttore (o facilitatore) deve sollecitare il partecipante a verbalizzare tutte le sue azioni
 - Che cosa sta cercando di fare
 - Che cosa vede sullo schermo
 - Come pensa di dover proseguire
 - Quali dubbi e difficoltà sta provando
- Il conduttore deve evitare di:
 - Intervenire nel merito dei problemi verbalizzati
 - Porre domande dirette che possono guidare il partecipante
 - Esprimere delusione o gioia per i comportamenti del partecipante
- Alcuni partecipanti hanno difficoltà a pensare a voce alta, dunque con questi è bene non insistere nell'incoraggiamento

15

Organizzare un test di usabilità



R.Polillo - Ottobre 2010

16

16

Preparazione del test

- Definizione obiettivi, tipo di test, misure da raccogliere
- Definizione numero e tipologia degli utenti campione
- **Definizione compiti e/o scenari d'uso**
- Individuazione utenti campione
- Preparazione materiali e ambiente di prova

R.Polillo - Ottobre 2010

17

17

Quali misure?

Tipicamente:

- il **tempo** richiesto da un determinato compito
- la percentuale di compiti portati a termine con successo (**tasso di successo**)
- la **soddisfazione dell'utente**

R.Polillo - Ottobre 2010

18

18

Alcune metriche da ISO 9241

Some Metrics from ISO 9241

Usability objective	Effectiveness measures	Efficiency measures	Satisfaction measures
Suitability for the task	Percentage of goals achieved	Time to complete a task	Rating scale for satisfaction
Appropriate for trained users	Number of power features used an expert user	Relative efficiency compared with power features	Rating scale for expert satisfaction
Learnability	Percentage of functions learned	Time to learn criterion	Rating scale for ease of learning
Error tolerance	Percentage of errors corrected successfully	Time spent on correcting errors	Rating scale for error handling

19

Tasso di successo: esempio

	Compito 1	Compito 2	Compito 3	Compito 4	Compito 5	Compito 6
Utente 1	F	F	S	F	F	S
Utente 2	F	F	P	F	P	F
Utente 3	S	F	S	S	P	S
Utente 4	S	F	S	F	P	S

Legenda: S=successo F=fallimento P=successo parziale

Tasso di successo : $(9 + (4 \cdot 0.5)) / 24 = 46\%$

successi

successi
parziali

numero
prove

R.Polillo - Ottobre 2010

20

20

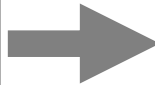
Quanti utenti?

Dipende dagli obiettivi del test e dalla complessità del sistema

Esempio: per un sito web, tipicamente

Check-up rapido, durante lo sviluppo:

- 5-7 utenti
- 5-7 compiti ciascuno (20-40 minuti per ciascun utente)



Valutazione approfondita di un sito complesso:

- 10-15 utenti
- 1 – 1,5 h per ciascun utente

R.Polillo - Ottobre 2010

21

21

Durante lo sviluppo iterativo

“Dopo il quinto utente, sprecherete il vostro tempo osservando più volte gli stessi risultati, senza imparare molto di nuovo

Dopo avere osservato 5 utenti, eliminate i problemi individuati, e provate di nuovo con 5 utenti”

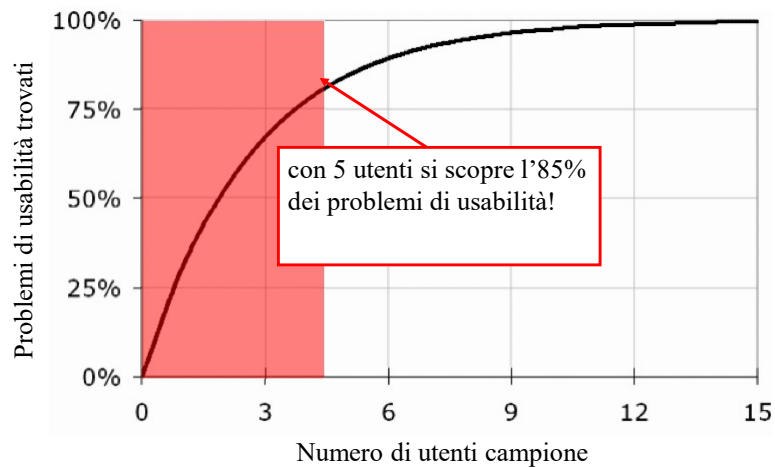
(Jakob Nielsen)

R.Polillo - Ottobre 2010

22

22

La regola di Nielsen



R.Polillo - Ottobre 2010

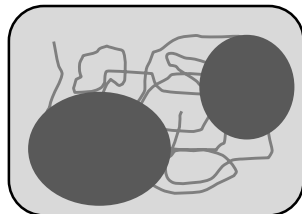
23

23

L'effetto mascheramento

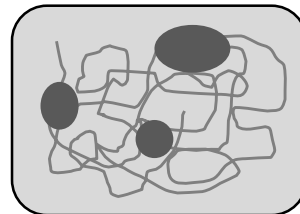
Iterazione n

I problemi incontrati mi impediscono di vederne altri:



Iterazione n+1

I problemi incontrati mi impediscono di vederne altri:



R.Polillo - Ottobre 2010

24

Quali compiti o scenari?

- E' una decisione critica
- Compromesso fra copertura delle situazioni possibili e tempo/risorse
- Basarsi sulle priorità espresse nei requisiti

R.Polillo - Ottobre 2010

25

25

Preparazione materiali e ambiente di prova

- **Scheda utente** (esperienza, conoscenza del sistema, ...)
- **Descrizione scritta** dei compiti/scenari, da dare agli utenti
- **Modulo di raccolta misure e osservazioni**, per l'osservatore (uno per ogni utente e compito/scenario)
- **Questionario** per l'intervista finale agli utenti

R.Polillo - Ottobre 2010

26

26

Esempio: Scheda utente

1. Dati personali

Nome: _____

Età: _____

Titolo di studio: _____

Professione: _____

2. Livello di conoscenza di Internet

Giudichi di avere una esperienza d'uso del web:
scarsa – media – buona – ottima

In media quante ore alla settimana usi il web?
meno di 1 – tra 1 e 5 – più di 5 – più di 10

Hai mai partecipato ad un forum o ad una chat su Internet?
Sì – No

3. Livello di conoscenza del sito

Hai già utilizzato il sito in esame?
No – Sì, meno di 3 volte – Sì, più di 3 volte – Sì, più di 10 volte

Se sì, quali operazioni hai effettuato?

Hai mai utilizzato siti simili?

No – Sì, qualche volta – Sì, spesso

R. Polillo - Ottobre 2010

27

Esecuzione del test

- Briefing agli utenti
- Conduzione del test
- Intervista individuale (o questionario, p.e. SUS) o focus group finale

>>

R. Polillo - Ottobre 2010

28

28

Briefing agli utenti

- Mettere gli utenti a proprio agio, per ridurre al massimo lo stress da esame
- Spiegare che lo scopo è di provare il sistema, non l'utente
- Spiegare quali registrazioni verranno fatte, e la politica relativa alla privacy
- Fornire agli utenti l'elenco scritto dei compiti/scenari

R.Polillo - Ottobre 2010

29

29

Il ruolo del facilitatore

Sì:

- a che cosa stai pensando?
- continua a parlare
- non scoraggiarti, tenta ancora

No:

- a che cosa serve quel bottone?
- perchè hai cliccato lì?

NB: Interferenza cognitiva del think-aloud

R.Polillo - Ottobre 2010

30

30

Intervista o focus group finale

- Per discutere:
 - punti di forza e di debolezza dell'applicazione
 - aspetti da migliorare
 - aspetti graditi e sgraditi
- Utilizzare un questionario già noto (p.e. SUS) per dare una misura quantitativa (metrica) della soddisfazione dell'utente

R.Polillo - Ottobre 2010

31

31

Analisi dei risultati e proposte finali

- Analisi dettagliata dei dati e della registrazione
- Elenco dei singoli problemi, e loro gravità (es. **bloccanti / superabili / lievi**)
- Elenco degli interventi suggeriti, e loro priorità
- Stesura del rapporto di valutazione

R.Polillo - Ottobre 2010

32

32

Elenco dei problemi: esempio

	PROBLEMA IDENTIFICATO	PRIORITA'
1	Se si accede alla registrazione dalla Home Page, nella prima videata viene richiesta "la verifica del CAP". L'utente non comprende il significato dell'acronimo CAP (ritiene si tratti di un codice personale post-registrazione). Dopo l'intervento del facilitatore per segnalare il significato dell'acronimo, l'utente dichiara di non comprendere l'utilità di tale verifica.	A
2	Difficoltà nel comprendere il significato dei campi: [Domanda] e [Risposta]: non viene data alcuna informazione sul motivo di tale richiesta.	A
3	Non viene in alcun modo segnalato che il numero di caratteri che si possono inserire, sia per la [Domanda] che per la [Risposta], sono limitati. L'utente non ha la possibilità di accorgersi che entrambe le stringhe di testo inserite saranno troncate.	A
4	Iniziale smarrimento nella conferma della registrazione: l'utente si attendeva un comando "Invia" e non "Salva i dati password" (etichetta ricavata dal nome dell'immagine "Salva_dati_password.gif")	B
5	Viene dato l'obbligo di inserire due numeri telefonici creando frustrazione in chi non ha un secondo numero utile per gli scopi indicati: l'utente si mostra riluttante.	A
6	Al momento di inserire i dati per la consegna ad una terza persona, trovando reinseriti i propri dati, non si accorge del vero scopo di quella schermata, e aggiunge i suoi dati, lamentandosi inoltre che gli viene richiesto il CAP per la terza volta.	A
7	Identifica il simbolo di Page Up [^], posto a piè pagina, come un indicatore per muoversi sequenzialmente all'interno delle pagine (Forward, Back) anziché che per la funzione di scrolling nella pagina	M

(Compito: registrazione utente in un sito di e-commerce)

33

33

Elenco interventi suggeriti: esempio

CARRELLO – CASSA – SCONTRINO		
	RACCOMANDAZIONI	PRIORITA'
1	Lasciare sempre in vista i contenuti del carrello	1
2	Di fianco ad ogni prodotto del carrello inserire il comando "elimina dal carrello" oppure "elimina". Il comando "svuota il carrello" può restare in alto ad inizio lista.	1
3	I prodotti inseriti nel carrello saranno quelli che verranno conteggiati per la spesa. Non occorrerà selezionarli	1
4	Cambiare il "Totale Spesa" con il "Totale Carrello"	2
5	Sostituire il termine "conferma l'ordine" con il comando "Invia l'ordine", più chiaro e convenzionale in Internet	1
6	Trovare una modalità più chiara per scegliere la data e la fascia oraria di consegna; ad esempio, sottolineando con un link ogni possibilità di scelta	1
7	Dare informazioni sulle possibili modalità di pagamento ed offrire un link verso la pagina che contiene informazioni di dettaglio	1
8	Indicare i dati riassuntivi della spesa appena effettuata ed inviare messaggio di conferma alla casella e-mail del cliente	1
9	Rendere possibile la funzione di stampa dalla pagina contenente i dati riassuntivi	2
10	Eliminare i termini scontrino e cassa che risultano termini arbitrari in quanto non corrispondenti a delle funzioni reali ed utili per effettuare la spesa on-line.	2

R.Polillo - Ottobre 2010

34

34

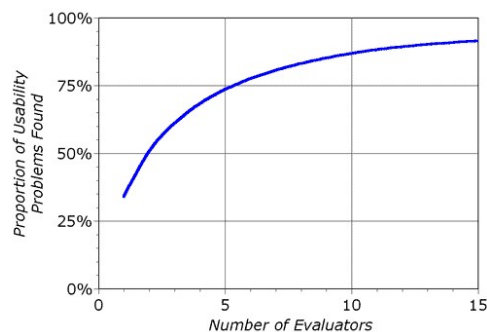
Ispezione di Usabilità

- Valutazione da parte di esperti
- Gli utenti finali non sono coinvolti
- Ispezione sistematica delle schermate di un prodotto sulla base di euristiche, check-list e standard
 - principi di usabilità
 - style guide aziendali
 - standard tecnologici
 - norme ISO/UNI

35

Valutazione Euristica

- Esame delle singole schermate da parte dell'esperto sulla base di principi e linee guida (euristiche), eseguendo alcuni task rappresentativi
 - verifica delle componenti statiche e dinamiche dell'interfaccia
- Correlazione dei risultati dell'analisi degli esperti e conclusioni
- E' un "discount usability method"
 - a very efficient method with a high benefits-costs ratio (Nielsen, 1993)



Persone diverse trovano problemi diversi

(Nielsen and Landauer, 1993)

36

Esempio di rapporto di valutazione

Con l'aiuto di tecniche di ispezione è stata effettuata una valutazione di usabilità di tale applicazione con l'intento di facilitare l'interazione dell'utente con il sistema. Tutti i problemi riscontrati e le possibili soluzioni sono stati riportati nel seguito. E' comunque auspicabile un incontro con i progettisti del sito per discutere sulle soluzioni proposte per migliorare l'usabilità del sito.

N.º Problema	Locazione	Problema	Principio Violato	Possibile Soluzione
1	Homepage	Il nome della compagnia sull'homepage non è sempre presente poiché animata.	Minimizzare il carico di memoria dell'utente	Il nome della compagnia dovrebbe essere sempre presente insieme al logo e non animato.
2	Homepage	Il nome della compagnia essendo quello della famiglia non dà l'idea di cosa si occupa e di cosa offre il sito.	Minimizzare il carico di memoria dell'utente	Aggiungere una linea esplicativa per far capire di cosa si occupa senza costringere a navigare per conoscere il sito.
3	Lato destro dell'homepage	La pubblicità inserita sul lato destro dell'homepage distrae l'utente, primo perché animata, secondo perché crede che sia un item importante (data anche la sua posizione), infine l'utente crede che sia cliccabile e invece non lo è.	Prevedibilità	Sarebbe opportuno inserirla in un luogo meno attraente rispetto agli item di più alta priorità.

37

Valutazione Euristica

Problem rating

- It is important to realize that an individual evaluation is an opinion and should be treated as such. The benefit comes when you combine your 3-5 evaluations together into a larger list, removing duplicates. After merging, have your evaluators rate each problem, those they found themselves and those found by others.
- Usually, you use a 5-point scale, with zero meaning "I do not think it is a problem," 4 indicating a "usability catastrophe," and less serious problems rated in between. Calculate the average rating for each problem and see what stands out as the most serious.
- If all of your evaluators agree that something is a catastrophe, then it should be high on your list to correct. Other errors will be ranked lower and can be tackled as time and resources permit. Besides finding problems, a heuristic evaluation helps you prioritize what to focus on next.

38

Valutazione Euristica

Web evaluation

- Heuristic evaluation is well-suited for the Web because it can be easy, fast and inexpensive. Often the evaluators are specially-trained usability specialists, but you can also use people with just an hour of training. If your site is already "live" or on a staging area, then your evaluators can be spread out across the globe.
- Basically, heuristic evaluation involves identifying your heuristics, gathering opinions about the usability of your site, merging and rating the problems that were identified, and then trying to work toward solutions.
- Heuristics: the rules of thumb you use are the most important part of a heuristic evaluation. They set the stage for identifying problems and provide the vehicle for finding solutions.
- Jakob Nielsen originated heuristic evaluation in the early 1990s and has compiled the most comprehensive set of heuristics.

39

Valutazione Euristica

Heuristics for User Interface Design

- Nielsen's heuristics come from analyzing more than 200 usability problems and statistically reducing the set to the 10 most important rules:
- Visibility of system status
- Match between system and the real world
- User control and freedom
- Consistency and standards
- Error prevention
- Recognition rather than recall
- Flexibility and efficiency of use
- Aesthetic and minimalist design
- Help users recognize, diagnose, and recover from errors
- Help and documentation
- <http://usableweb.com/> (Inspection Methods)
- http://www.webreview.com/1997/10_10/strategists/10_10_97_2.shtml

40

Le 10 euristiche proposte da J. Nielsen

- Riviste e annotate da Giorgio Brajnik
 - domande specifiche che possono favorire un'analisi più approfondita dei problemi
 - Il valutatore deve però integrare la sua valutazione con altre domande, specifiche per contesto dell'analisi (sito, utenza, scopo dell'indagine, ...)
 - <http://users.dimi.uniud.it/~giorgio.brajnik/dida/psw/euristiche.html>

41

Valutazione Euristica Inconvenienti

- Effetto dell'esperienza del valutatore sulla qualità della valutazione
 - Valutatori novizi: valutatori "scarsi" (22% of problems)
 - Esperti di usabilità: 1.8 as good as the novices (41% of problems)
 - double experts (usability + domain): 1.5 as good as the usability experts (60% of problems)

(Nielsen, 1994)
- Heuristics are too general, and too little operational
 - not able to guide the evaluators during inspection

(Doubleday et al., 1997)
- Limited support in using a precise terminology
 - problem reports are too vague

42

Indirizzi sull'usabilità e su W3C

- [Http://www.useit.com/hotlist](http://www.useit.com/hotlist)
- [Http://www.usableweb.com](http://www.usableweb.com)
- [Http://www.useit.com/books](http://www.useit.com/books)
- [Http://www.w3.org](http://www.w3.org)
- <http://www.w3schools.com> (vari tutorials su HTML, XML, ...)

43

Altre tecniche basate sull'utente

- Inchieste (*inquiry*)
 - interviste individuali
 - questionari on-line
 - focus group

NB: Non possono sostituire i test di usabilità!
- Esperimenti di laboratorio statisticamente significativi

R.Polillo - Ottobre 2010

44

44

Esperimenti di laboratorio: esempi

- Leggibilità di testi sul video (dimensione font, tipo di font, colori, ottimali ...)
- Efficacia dei banner pubblicitari su pagine web
-

Richiedono una metodologia rigorosa che qui non trattiamo
(scelta del campione, definizione delle ipotesi, conduzione dell'esperimento, analisi dei risultati con i metodi della statistica)

R.Polillo - Ottobre 2010

45

45

Controllo di risposte fisiologiche

Speciali apparecchiature permettono di condurre test relativi alle risposte fisiologiche dell'utente durante l'interazione con il sistema

In particolare:

- apparecchiature per il tracciamento dei movimenti oculari
- apparecchiature per il rilevamento di reazioni emotive (attività cardiaca, sudorazione, attività elettrica nel cervello, tono muscolare)

R.Polillo - Ottobre 2010

46

46

Classificazione dei problemi

- Qualunque sia la tecnica di valutazione utilizzata, i problemi individuati dovranno essere classificati sulla base della loro importanza, ad esempio:

Problema irrilevante: può non essere risolto

Problema secondario: da risolvere con bassa priorità

Problema rilevante: da risolvere con alta priorità

Problema bloccante: deve necessariamente essere risolto prima che il sistema venga rilasciato