HUMAN-COMPUTER INTERFACE LEZIONE 9: TEST DI USABILITA'

a.a 2021/2022

Prof. Mariacarla Staffa

VALUTAZIONE DEL SISTEMA

- ad ogni ciclo di iterazione si effettuano dei test di valutazione dell'ultimo prototipo prodotto
- la <u>valutazione</u> è un passo essenziale in una progettazione human-centred.

 All'inizio del progetto, l'obiettivo principale sarà la raccolta di indicazioni per orientare le attività di progettazione successive. Nelle fasi più avanzate, con la disponibilità di prototipi più completi, sarà possibile quantificare il livello di raggiungimento degli obiettivi dell'utente e dell'organizzazione



modello di proqettazione e sviluppo iterativi

VERIFICHE E CONVALIDE

I termini generici "valutazione" o "test" possono denotare due attività molto diverse:

- il controllo che il prodotto sia congruente con quanto espresso nei documenti di specifica dei requisiti. Per questo tipo di test si usa normalmente il termine verifica (verification);
- il controllo che il prodotto soddisfi effettivamente le esigenze per le quali è stato concepito. Per questo tipo di test si usa, invece, il termine convalida (validation).

TECNICHE DI VALUTAZIONE DELL'USABILITA'

Si possono dividere in due grandi categorie:

- Valutazioni attraverso l'analisi di esperti di usabilità ("ispezioni")
 - Valutazioni euristiche
 - Walkthrough cognitivi
 - · ...
- Valutazioni con la partecipazione degli utenti
 - Test di usabilità
 - Interrogazioni
 - · ...

VALUTAZIONI BASATE SU EURISTICHE

Euristico (dal greco eurisco, trovo): in matematica, qualsiasi procedimento non rigoroso che consente di prevedere o rendere plausibile un risultato, il quale in un secondo tempo dovrà essere controllato e convalidato per via rigorosa

(Vocabolario Treccani)

- Il sistema viene esaminato verificandone la conformità a specifiche "regole d'oro" (dette, appunto, "euristiche"), derivanti da principi e linee guida generalmente accettati
- Le euristiche possono essere diverse, più o meno dettagliate.
- Si preferiscono euristiche costituite da pochi principi molto generali,
 piuttosto che linee guida dettagliate, di difficile utilizzo

LE EURISTICHE DI NIELSEN (1994)

1. Visibilità dello stato del sistema

2. Corrispondenza fra il mondo reale e il sistema

3. Libertà e controllo da parte degli utenti

4. Consistenza e standard

5. Prevenzione degli errori

6. Riconoscere piuttosto che ricordare

7. Flessibilità ed efficienza d'uso

8. Design minimalista ed estetico.

9. Aiutare gli utenti a riconoscere gli errori, diagnosticarli e correggerli

10. Guida e documentazione

LE EURISTICHE DI NIELSEN (1994)

- 1. Visibilità dello stato del sistema: Il sistema dovrebbe sempre informare gli utenti su ciò che sta accadendo, mediante feedback appropriati in un tempo ragionevole.
- 2. Corrispondenza fra il mondo reale e il sistema: Il sistema dovrebbe parlare il linguaggio dell'utente, con parole, frasi e concetti familiari all'utente, piuttosto che termini orientati al sistema. Sequire le convenzioni del mondo reale, facendo apparire le informazioni secondo un ordine logico e naturale.
- 3. Libertà e controllo da parte degli utenti: Gli utenti spesso selezionano delle funzioni del sistema per errore e hanno bisogno di una "uscita di emergenza" segnalata con chiarezza per uscire da uno stato non desiderato senza dover passare attraverso un lungo dialogo. Fornire all'utente le funzioni di undo e redo.
- 4. Consistenza e standard: Gli utenti non dovrebbero aver bisogno di chiedersi se parole, situazioni o azioni differenti hanno lo stesso significato. Seguire le convenzioni della piattaforma di calcolo utilizzata.
- 5. Prevenzione degli errori: Ancora meglio di buoni messaggi di errore è un'attenta progettazione che eviti innanzitutto l'insorgere del problema. Eliminare le situazioni che possono provocare errori da parte dell'utente, e chiedergli conferma prima di eseguire le azioni richieste.

LE EURISTICHE DI NIELSEN (1994)

- 6. Riconoscere piuttosto che ricordare: Minimizzare il ricorso alla memoria dell'utente, rendendo visibili gli oggetti, le azioni e le opzioni. L'utente non dovrebbe aver bisogno di ricordare delle informazioni, nel passare da una fase del dialogo a un'altra. Le istruzioni per l'uso del sistema dovrebbero essere visibili o facilmente recuperabili quando servono.
- 7. Flessibilità ed efficienza d'uso: Acceleratori invisibili all'utente novizio possono spesso rendere veloce l'interazione dell'utente esperto, in modo che il sistema possa soddisfare sia l'utente esperto sia quello inesperto. Permettere all'utente di personalizzare le azioni frequenti.
- 8. Design minimalista ed estetico: I dialoghi non dovrebbero contenere informazioni irrilevanti o necessarie solo di rado. Ogni informazione aggiuntiva in un dialogo compete con le unità di informazione rilevanti e diminuisce la loro visibilità relativa.
- 9. Aiutare gli utenti a riconoscere gli errori, diagnosticarli e correggerli: I messaggi di errore dovrebbero essere espressi in linguaggio semplice (senza codici), indicare il problema con precisione e suggerire una soluzione in modo costruttivo.
- 10. Guida e documentazione: Anche se è preferibile che il sistema sia utilizzabile senza documentazione, può essere necessario fornire aiuto e documentazione. Ogni tale informazione dovrebbe essere facilmente raggiungibile, focalizzata sul compito dell'utente, e dovrebbe elencare i passi concreti da fare, senza essere troppo ampia.

Principi del dialogo (ISO 9241)

Euristiche di Nielsen

Adeguatezza al compito Riconoscere piuttosto che ricordare Flessibilità ed efficienza d'uso Autodescrizione Visibilità dello stato del sistema Conformità alle aspettative Corrispondenza fra mondo reale e sistema Consistenza e standard Controllabilità Libertà e controllo da parte degli utenti Prevenzione degli errori Tolleranza verso gli errori Aiutare gli utenti a riconoscere gli errori, ... Adeguatezza alla individualizzazione Design minimalista ed estetico

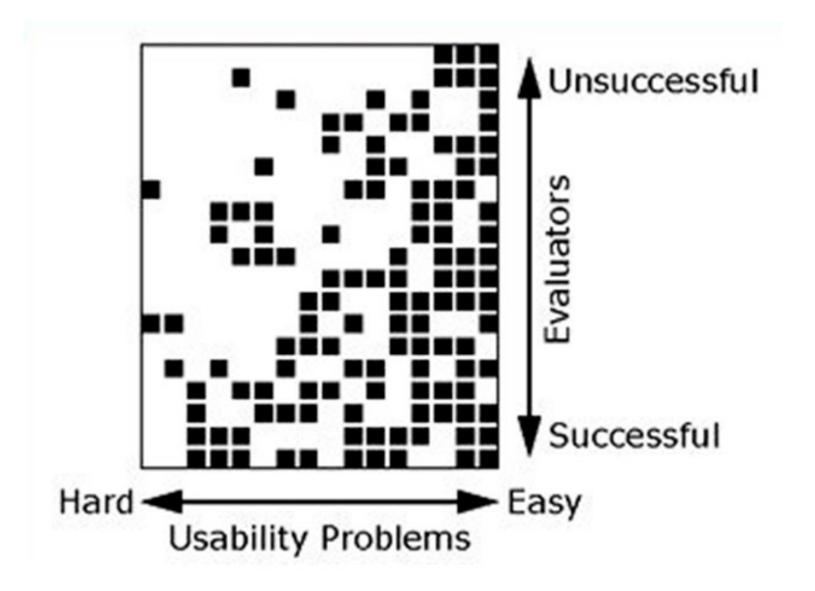
CONFRONTO FRA I PRINCIPI DEL DIALOGO DELL'ISO 9241 E LE EURISTICHE DI NIFLSEN

VALUTAZIONI EURISTICHE: ASPETTI

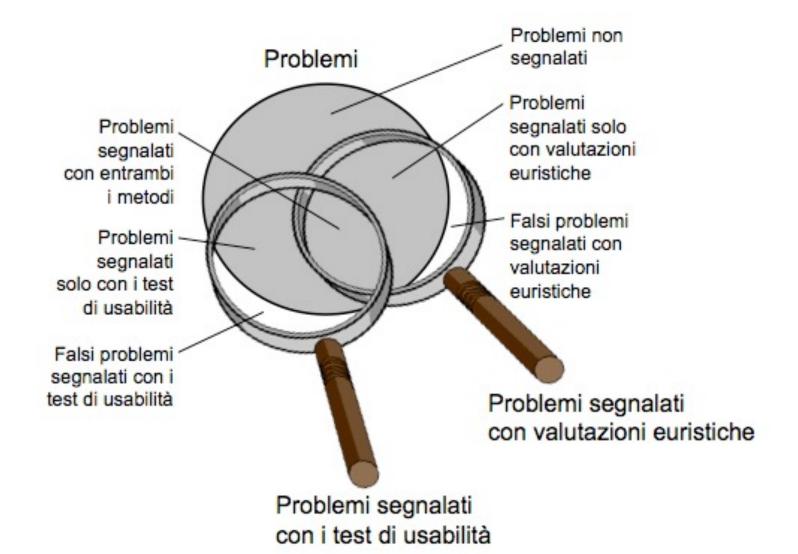
- Poco costose ("discount usability")
- Un valutatore non basta (valutatori diversi tendono a trovare problemi diversi)
- Tipicamente, ci si aspetta che 5 valutatori scoprano i due terzi dei problemi di usabilità
- I risultati sono tanto migliori quanto più i valutatori sono esperti (di usabilità e della applicazione)



ESEMPIO



VALUTAZIONE EURISTICHE VS TEST DI USABILITA'



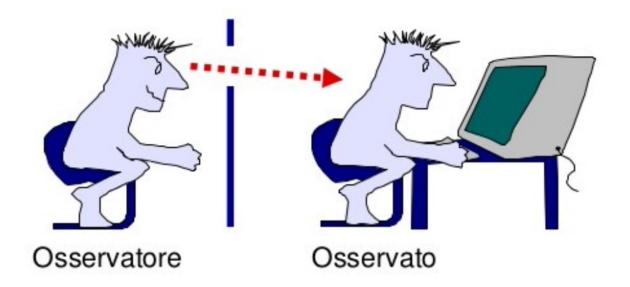


from: Carelman, Objets introuvables, 1969

TEST DI USABILITA'

- Utenti
- Osservatori
- Facilitatori
- Wizard of Oz

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



Si chiede all'utente di esprimere ad alta voce ciò a cui sta pensando:

THINKING ALOUD

- che cosa sta cercando di fare
- che cosa vede sullo schermo
- come pensa di dover proseguire
- quali dubbi e difficoltà sta provando

REGISTRAZIONI AUDIO-VIDEO

Registrazione di un test di usabilità per un'applicazione su iPhone



Registrazione di un test di usabilità con prototipo di carta

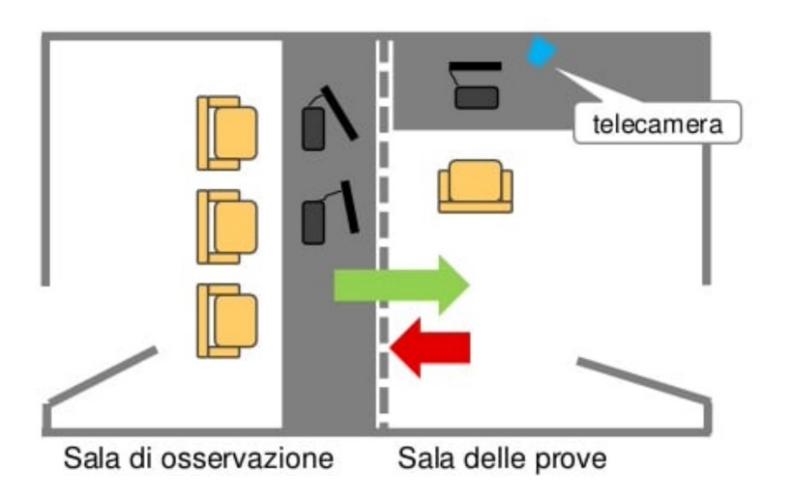




TEST DI USABILITÀ: LOGISTICA

- Due filosofie:
 - Usability Lab (costi alti)
 - Organizzazione Informale (costi bassi)

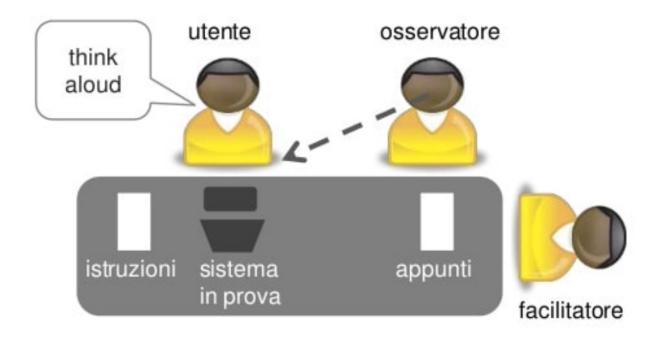
USABILITY LAB

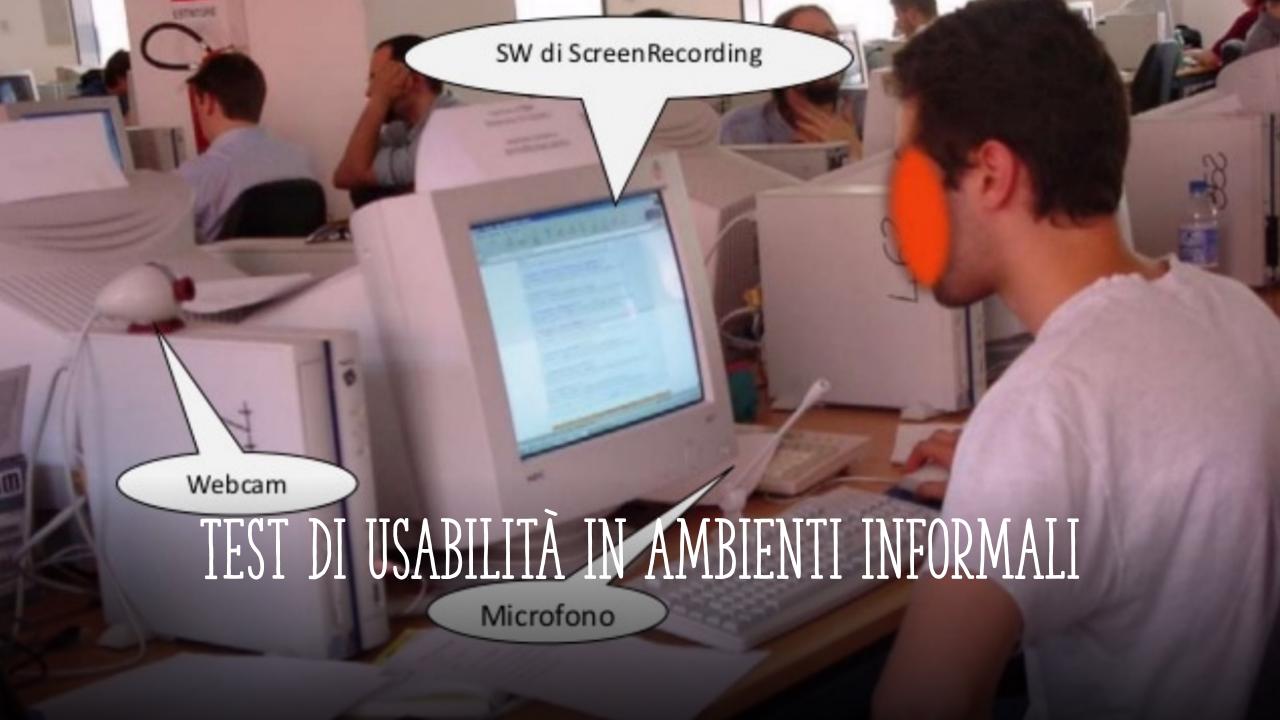






ORGANIZZAZIONE INFORMALE





TEST FORMATIVI E SOMMATIVI

l test di usabilità possono essere classificati, in funzione dei loro obiettivi, in due grandi categorie:

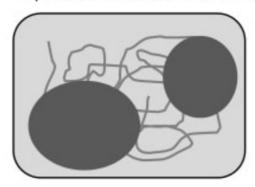
- <u>Test formativi (formative):</u> sono utilizzati durante il ciclo iterativo di progettazione, per sottoporre i vari prototipi a prove d'uso con gli utenti, allo scopo di identificarne i difetti e migliorarne l'usabilità. Si chiamano formativi perché, appunto, contribuiscono a "dare forma" al prodotto: il loro scopo è individuare il maggior numero possibile di problemi
- Test sommativi (summative): I test della seconda categoria si chiamano sommativi. Il termine deriva dal verbo "sommare", ed indica una valutazione più complessiva del prodotto, al di fuori o al termine del processo di progettazione e sviluppo. Sono test più completi di quelli formativi, che non hanno lo scopo di fornire indicazioni ai progettisti, ma di valutare in modo sistematico pregi e difetti del prodotto, o sue particolari caratteristiche.

TEST FOMRATIVI: «EFFETTO MASCHERAMENTO»

- Ogni problema di usabilità nel quale incappiamo monopolizza, per così dire, la nostra attenzione e ci impedisce spesso di vederne altri, soprattutto se di più lieve entità.
- Li scopriremo all'iterazione successiva, quando il problema iniziale sarà stato rimosso.
- Quindi si prova in fretta, si modifica rapidamente il prototipo eliminando i difetti più evidenti, e si prova ancora, e così via.

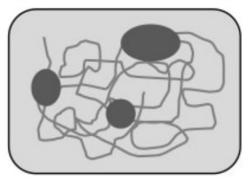
Iterazione n

I problemi incontrati mi impediscono di vederne altri:



Iterazione n+1

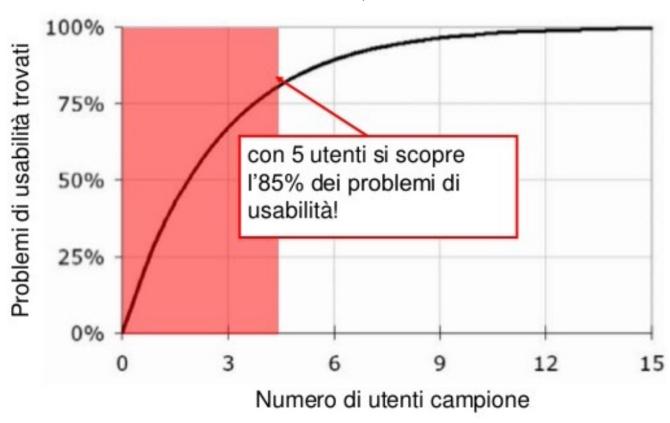
I problemi incontrati mi impediscono di vederne altri



Per il primo test di un prototipo iniziale di carta, 2–3 utenti sono in genere sufficienti.

LA REGOLA DI NIELSEN





TEST FOMRATIVI: 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"

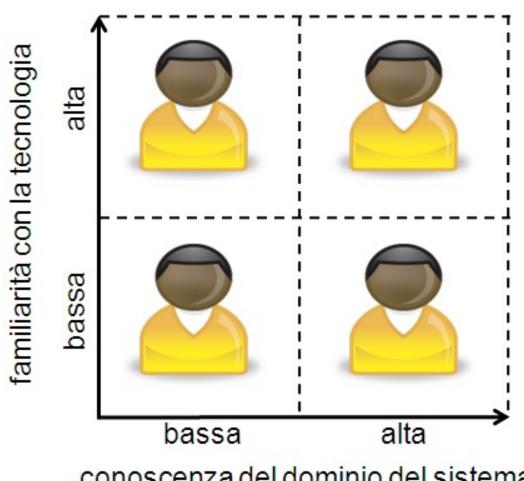
(Jakob Nielsen)

TEST SOMMATIVI:

- Sono test più completi di quelli formativi, che hanno lo scopo di valutare in modo sistematico pregi e difetti del prodotto
- Sono di solito condotti quando il sistema è completamente funzionante, per esempio per indicarne i punti deboli e valutare l'opportunità di un redesign migliorativo.
- Oppure per confrontarne le caratteristiche con quelle di sistemi concorrenti.
- Coinvolgono di solito un numero maggiore di utenti, per esempio 10-15, o anche di più.

TEST SOMMATIVI:

- I soggetti da utilizzare nei test dovranno sempre essere scelti con cura, affinché rappresentino utenti tipici.
- Per poter interpretare correttamente l'esito di ciascun test, chi lo conduce dovrà conoscere, per ciascun soggetto, il livello di esperienza nell'uso di sistemi analoghi a quello in esame. Il suo profilo dovrebbe essere classificato su due dimensioni:
 - il livello di conoscenza del dominio applicativo del sistema,
 - e il livello di familiarità con la tecnologia utilizzata



conoscenza del dominio del sistema

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

TIPI DI TEST DI USABILITA'

Test di compito

Agli utenti viene chiesto di svolgere compiti specifici, che permettano di esercitare le funzioni principali del sistema Es.: Registrarsi

Test di casi d'uso

Agli utenti si chiede di svolgere diversi casi d'uso Es.: Acquista una scatola di pomodori

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 Es.: Effettua gli acquisti per una cena di amici domani sera

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

ESMPIO: 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.

ASPETTATEVI DELLE SORPRESE



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

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	Р	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*0.5)) / 24 = 46%

successi

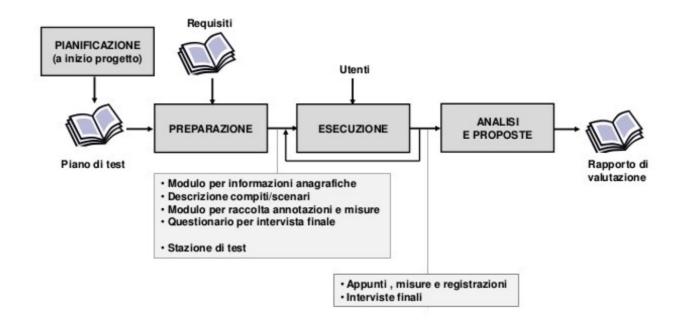
successi parziali

numero prove

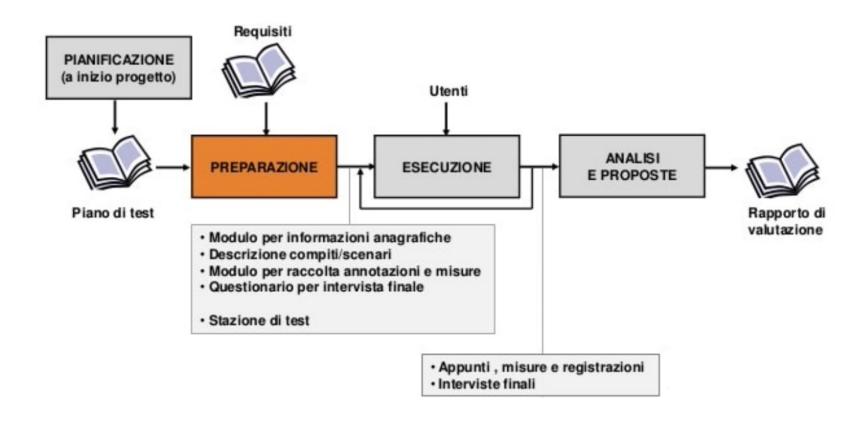
QUALI COMPITI O SCENARI?

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

COME CONDURRE UN TEST DI USABILITA'



ORGANIZZARE UN TEST DI USABILITA'



PREPARAZIONE DEL TEST

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

MATERIALI PER LO SVOLGIMENTO DEL TEST

- <u>un modulo per raccogliere le informazioni sugli utenti</u>: informazioni anagrafiche, livello di esperienza, competenze nell'ambito specifico del sistema;
- il testo con le <u>istruzioni per lo svolgimento delle prove</u> da consegnare agli utenti: la descrizione di compiti e scenari dovrebbe essere concisa, ma particolarmente chiara, per evitare chiarimenti e spiegazioni durante lo svolgimento della prova;
- <u>la modulistica che gli osservatori utilizzeranno per raccogliere le misure relative all'esecuzione di ciascun compito, e le loro annotazioni durante il test</u>. Questi moduli devono permettere di annotare rapidamente le informazioni utili, e quindi dovrebbero essere predisposti in funzione delle caratteristiche del test.
- un questionario per le interviste finali degli utenti

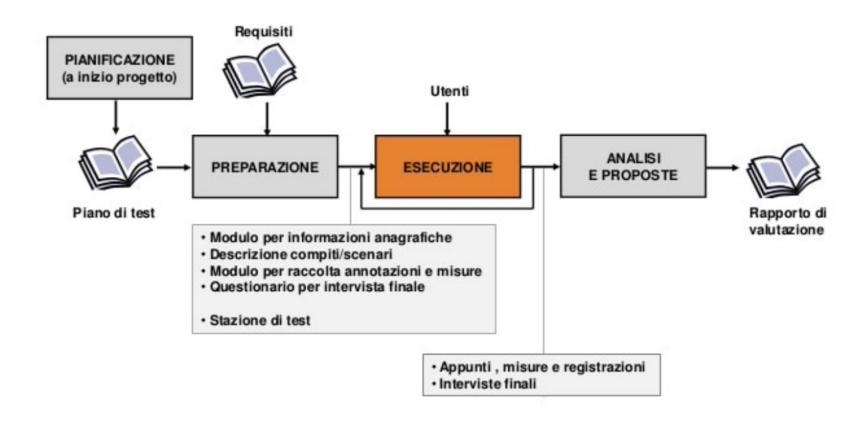
PREPARAZIONE MATERIALI E AMBIENTE DI PROVA

- 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

ESEMPIO

Dati personali	
Nome: _ Età:	
	-
Professione:	
2. Livello di cono	oscenza di Internet
Giudichi di avere i	una esperienza d'uso del web:
scar	sa – media – buona - ottima
In media quante o	re alla settimana usi il web?
men	o di 1 – tra 1 e 5 – più di 5 – più di 10
Hai mai partecipat Si –	to ad un forum o ad una chat su Internet? No
3. Livello di cono	scenza del sito
Hai già utilizzato il	sito in esame?
No-Sì, m	eno di 3 volte - Sì, più di 3 volte - Sì, più di 10 volte
Se sì, quali opera:	zioni hai effettuato?
Hai mai utilizzato :	siti simili?
No -	- Si,qualche volta - Si, spesso
Se s	si, quali?

ORGANIZZARE UN TEST DI USABILITÀ



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

IL RUOLO DELL'OSSERVATORE

Sì:

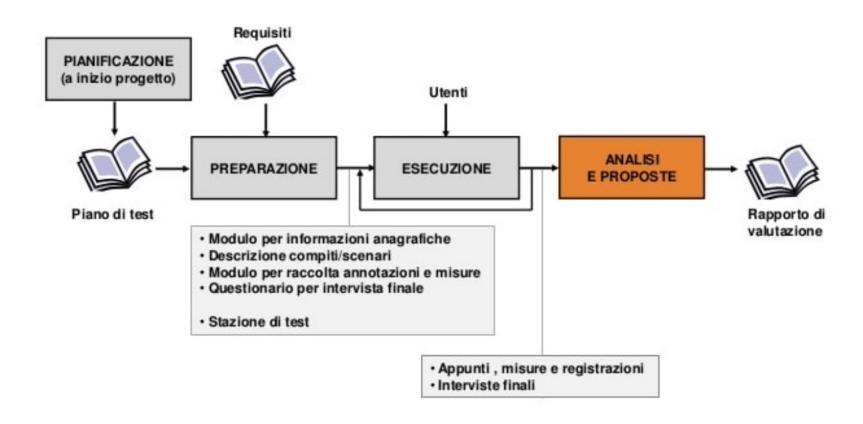
- a che cosa stai pensando?
- continua a parlare
- non scoraggiarti, tenta ancora

No:

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

NB: Interferenza cognitiva del think-aloud

ORGANIZZARE UN TEST DI USABILITA'



ANALISI DEI RISULTATI E PROPOSTE FINALI

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

ELENCO DEI PROBLEMI

	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.	
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.	А
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")	В
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.	А
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.	Α
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	М

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 carello" oppure "elimina". Il comando "svuota il carrello" può restare in alto ad inizio lista.	
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.	40-0-0

INDICE DI MASSIMA DEL RAPPORTO DI TEST

Obiettivi del test

Metodologia usata

Utenti

Compiti/scenari

Strumentazione

Modalità di svolgimento

Risultati del test

Analitici, con le misure

Raccomandazioni finali

Analitiche, con livelli di priorità

Allegati:

Eventuali questionari utilizzati

RAPPORTO DI VALUTAZIONE

Identificazione del documento

Riportare i nomi degli autori, la data e la versione del documento.

Sommario

Riportare una sintesi dello scopo del documento e delle sue conclusioni.

Prodotto valutato

Descrivere brevemente il prodotto o il prototipo sottoposto a test, con ogni informazione che lo identifichi con precisione. Indicare le aree funzionali sottoposte al test.

Obiettivi della valutazione

Descrivere gli obiettivi specifici perseguiti nella valutazione descritta nel documento.

Metodologia utilizzata

Specificare quanti utenti hanno partecipato al test, il loro livello di esperienza e le loro caratteristiche in relazione al prodotto in esame. Specificare i compiti o gli scenari assegnati, il contesto in cui si è svolto il test e la strumentazione utilizzata. Descrivere come è stato condotto il test e da chi, quanto tempo è durato (per ciascun utente e complessivamente), quali misure sono state raccolte, il ruolo degli osservatori e come sono stati analizzati i risultati.

Sintesi delle misure

Fornire una tabella di sintesi delle misure raccolte. Per esempio, i tempi di esecuzione e la percentuale dei vari compiti che sono stati portati a termine con successo, complessivamente e per ciascun utente. Aggiungere commenti ove opportuno.

Analisi dei risultati

Descrivere analiticamente i problemi incontrati da ciascun utente durante il test, compito per compito, allegando ove opportuno degli screen shot significativi e assegnando ad ogni problema un livello di gravità. Ogni problema sarà numerato, per un più facile riferimento. Descrivere in dettaglio, se significativi, reazioni e commenti degli utenti, registrati durante le prove. Questa è la sezione principale del documento, e dovrà contenere tutte le informazioni utili a formulare i possibili interventi per rimuovere i problemi descritti, senza che sia necessario tornare a esaminare il prodotto. Consisterà in genere di molte pagine.

Sintesi delle interviste agli utenti

Sintetizzare i risultati delle interviste effettuate a ciascun utente dopo l'esecuzione del test.

Raccomandazioni

Inserire la descrizione analitica degli interventi migliorativi proposti, raggruppati per livelli di priorità (per esempio: priorità 1: interventi indispensabili e urgenti; priorità 2: interventi necessari ma meno urgenti; priorità 3: interventi auspicati). Per ogni intervento proposto si farà riferimento al problema relativo, descritto nella sezione precedente. Gli interventi saranno numerati, per una facile tracciabilità.

Allegati

Allegare i moduli anagrafici compilati dagli utenti, la descrizione dei compiti/scenari data agli utenti prima del test, e tutti i questionari compilati nelle interviste finali. Allegare anche il materiale rilevante prodotto durante il test (per esempio, le riprese video).

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

VALUTAZIONE EURISTICHE VS TEST DI USABILITA'

