

6

USER-CENTERED DESIGN E APPROCCIO METODOLOGICO ALLA VALUTAZIONE

In questo capitolo affronteremo il tema della valutazione dell'usabilità di un sito web secondo un criterio cosiddetto *user-centered*, ossia centrato sull'utente.

Secondo quest'approccio, la valutazione, per essere efficace, dovrebbe avvenire durante l'intero ciclo di vita di progettazione di un sistema interattivo, quindi nel nostro caso specifico di un sito web, e fornire spunti per modifiche e miglioramenti sia durante che a fine realizzazione. Inoltre, non dimentichiamo che il fatto di mettere in primo piano i bisogni e le caratteristiche degli utenti, analizzare sin da principio i compiti che dovranno svolgere sul sito e indirizzarli in maniera specifica, valutare ripetutamente i prototipi e le nuove soluzioni di interfaccia prendendo misure empiriche sulle performance degli utenti, sono tutti fattori che possono contribuire a evitare costosi errori di progettazione.

Tali principi si ispirano proprio alla teoria del design centrato sull'utente. Gould e Lewis [Gould e Lewis, 1983] originariamente formularono i principi cardine di tale teoria:

- focalizzarsi sin dall'inizio sugli utenti del sistema interattivo e sui compiti che andranno a svolgere;
- effettuare misurazioni empiriche sull'utilizzo del sistema;
- progettare l'interfaccia del sistema secondo un approccio iterativo che alterni progettazione, implementazione e valutazione (*design, implement, evaluate*).

Dal momento che riteniamo che il design centrato sull'utente e il processo di valutazione continuo e iterativo che ne consegue siano i fattori chiave per la realizzazione di un sito web usabile e a misura d'utente, in questo capitolo presenteremo le metodologie di valutazione seguendo l'ordine temporale in cui queste dovrebbero essere applicate: la fase di analisi dei requisiti, la fase preliminare di valutazione e la fase finale di valutazione. Sottolineiamo, però, che ogni singola metodologia verrà solo descritta brevemente, e ne saranno più che altro evidenziati gli ambiti applicativi e i vantaggi. Consigliamo ai lettori che

vogliono maggiori dettagli su questi argomenti la lettura di [Preece et al. 1994; Dix et al., 1998] e dei riferimenti presentati nelle singole sezioni.

Inoltre, per fornire al lettore qualche ulteriore suggerimento pratico, ogni metodo di valutazione sarà ulteriormente classificato secondo una classificazione ben nota nella letteratura di riferimento [Carroll e Rosson, 1987; Nielsen e Mack, 1994; Preece et al. 1994]:

- *metodi empirici*, che possono essere applicati quando gli utenti reali sono coinvolti nella valutazione e i dati empirici possono, di conseguenza, essere raccolti ed analizzati. I metodi empirici sono i migliori ai fini di una valutazione oggettiva dell'usabilità dell'interfaccia e sono in grado di rilevare il più alto numero di errori di cattiva progettazione, fornendo anche degli spunti per risolverli. Tuttavia, se sono condotti impropriamente possono risultare inutili o addirittura fuorvianti. Uno dei problemi in cui si può incorrere nell'applicazione di tali metodi è rappresentato dalla difficoltà nel reperire i soggetti partecipanti, che dovrebbero essere gli utenti reali del sistema, o meglio una fetta rappresentativa di essi;
- *metodi analitici*, che possono essere utilizzati quando gli utenti sono difficili da reperire o quando esistono riconosciuti e condivisi criteri di riferimento da applicare per valutare il sistema (come ad esempio regole euristiche, linee guida e principi di usabilità, ecc). I metodi analitici sono diventati negli ultimi anni molto popolari, in quanto rapidi, relativamente economici e ragionevolmente efficaci;
- *metodi formali* (anche conosciuti come i metodi teoretici o metodi basati su modelli), che derivano dall'ergonomia, dalla psicologia cognitiva e dall'ingegneria del software e sono basati sui modelli delle azioni dell'utente e sui suoi processi cognitivi. Questi metodi tendono a non essere usati da chi non ha conoscenze specifiche sugli ambiti dai quali derivano, e si rivelano particolarmente utili per la previsione delle performance dell'utente, della complessità del sistema e della sua comprensibilità.

6.1 Metodologie per la raccolta dei dati.

Prima di elencare e descrivere le metodologie e le tecniche utili sia ai fini di una progettazione centrata sull'utente sia per una corretta valutazione di un sito web, saranno presentate le tipologie di dati utili ai fini della valutazione e le modalità con le quali possono essere raccolti.

Gli esperti di valutazione possono scegliere tra diverse metodologie e strumenti per la raccolta di dati a seconda degli scopi e delle circostanze della valutazione stessa, che dipendono dal tipo di approccio e di conseguenza dalla metodologia di valutazione che si è deciso di portare avanti, dalla fase temporale e dagli eventuali

vincoli che ci possono essere. È anche possibile che più di un metodo di raccolta dei dati sia adottato in una singola metodologia di valutazione. Ad esempio, in un test di usabilità gli utenti sono osservati durante lo svolgimento dei compiti, e a fine test le loro opinioni vengono raccolte tramite un questionario.

6.1.1 *La raccolta delle opinioni degli utenti*

La raccolta delle opinioni degli utenti è una metodologia che viene usata per conoscere il punto di vista degli utenti di un sistema. La filosofia che sta alla base di queste tecniche è quella del *chiedere all'utente* [Dix et al., 1998]. La raccolta di opinioni degli utenti, che spesso è usata in congiunzione con altre tecniche di valutazione (come, ad esempio, alla fine di un test di usabilità o di un esperimento controllato) in molti casi si rivela essere uno strumento assai utile, in grado far emergere delle problematiche non considerate dai progettisti dell'interfaccia del sito.

Le opinioni possono essere raccolte durante le differenti fasi della progettazione: all'inizio, per definire i punti chiave dello sviluppo dell'interfaccia e del sistema, nella fase d'implementazione quando si testano i prototipi, e durante la fase finale, per raccogliere le reazioni degli utenti e il loro livello di soddisfazione, ad esempio dopo una sessione di test o in risposta al reale utilizzo del sito.

I questionari. I questionari sono caratterizzati da una serie predefinita di domande e di risposte, che a loro volta possono essere predefinite e in tal caso sono denominate *chiuse* (ad esempio possono essere scalari, multi-scelta, semi-aperte), oppure sono *aperte*. A causa di questa natura rigida, i questionari risultano essere meno flessibili rispetto alle interviste, ma possono essere amministrati e analizzati più facilmente: possono essere infatti trasmessi via posta tradizionale, via email, somministrati via telefono, oppure faccia a faccia. In questi ultimi casi l'influenza dell'intervistatore nel completamento di domande deve essere controllata rigorosamente per evitare meccanismi di interferenza, così come può accadere nelle interviste. Questi problemi non si hanno invece nel caso di questionari auto-compilati con i quali, tuttavia può risultare difficile mantenere la correttezza nel campionamento dei soggetti, (ad esempio molti soggetti non rispondono), e quindi la selezione casuale non è garantita (vedi 6.2.3.4).

I dati raccolti possono poi essere analizzati quantitativamente, attraverso l'utilizzo di statistiche descrittive o inferenziali (vedi 6.2.3.4), o anche qualitativamente (vedi ad esempio 6.3.1)

Poiché per poter generalizzare i risultati è richiesto un considerevole numero di questionari, in alternativa si possono usare ricerche già esistenti sulle caratteristiche

e i consumi della popolazione di riferimento (ad esempio gli stili di vita Eurisko¹).

I questionari possono essere usati per raccogliere informazioni soggettive (per esempio i dati socio-demografici dell'utente, i suoi interessi e le sue preferenze, ecc.), informazioni sulle caratteristiche della personalità degli utenti (attraverso test di attitudine, test conoscitivi e di personalità, ecc), spunti per la progettazione o la riprogettazione del sito, dati sul gradimento del sito e dei servizi-informazioni che offre o che dovrebbe offrire, ecc.

Per approfondimenti si veda [Dix et al., 1998; Corbetta, 1999].

Le interviste. Le interviste [Nielsen, 1993; Dix et al, 1998; Corbetta, 1999] possono essere usate per raccogliere e poi analizzare, prevalentemente in forma qualitativa, opinioni ed esperienze, preferenze esplicite degli utenti, le motivazioni di un certo comportamento, ecc. Le interviste sono più flessibili dei questionari, poiché l'intervistatore può gestire in maniera più libera le domande, ponendo anche questioni non strettamente preventivate e/o spiegare domande che l'intervistato non comprende appieno.

Per la loro natura più libera, le interviste sono meno adatte rispetto ai questionari per analisi di tipo quantitativo e sono maggiormente indicate per studi di tipo esplorativo/induttivo, quando il ricercatore non sa esattamente quello che sta cercando e non ha un'ipotesi predefinita da valutare (si veda la sezione 6.3 sulle metodologie qualitative).

Normalmente le interviste sono registrate o trascritte dall'intervistatore.

Le interviste possono essere strutturate, semi-strutturate o non strutturate:

- nelle *interviste strutturate* lo stesso insieme di domande è posto nella medesima sequenza ad ogni intervistato. L'intervista strutturata è più facile da effettuare e da analizzare data la sua natura più vincolante;
- nelle *interviste semi-strutturate* l'intervistatore deve seguire uno script che riporta gli argomenti dell'intervista. Le domande ed il loro ordine non sono prefissati. L'intervista semi-strutturata risulta essere più adattabile al contesto ed alle risposte inattese;
- nell'*intervista non strutturata* la forma ed il contenuto non sono prefissati e possono variare da soggetto a soggetto. L'intervista non strutturata è completamente flessibile e l'unico vincolo che deve rispettare riguarda il fatto che deve toccare gli argomenti rilevanti ai fini della ricerca per cui è utilizzata.

Quando le interviste sono applicate insieme ad altre tecniche di valutazione (ad esempio test di usabilità, valutazione interpretativa, inchiesta contestuale, ecc)

¹ <http://www.eurisko.it>

possono risultare utili per chiarire gli eventi. Ad esempio, possono servire a raccogliere le impressioni dopo che l'utente ha completato una sessione sperimentale, o per segnalare i problemi reali di uso.

6.1.2 *Metodi per l'osservazione del comportamento dell'utente.*

Questa famiglia dei metodi è basata sull'osservazione diretta o indiretta dell'utente. L'osservazione dell'utente è un metodo estremamente efficace, ma può presentare diversi inconvenienti. I principali svantaggi riguardano le possibili interferenze che possono essere causate dall'utente (ad esempio il cosiddetto Effetto di Hawthorne, in cui soggetti riferiscono miglioramenti perché pensano sia ciò che l'investigatore vuole sentire), o dallo sperimentatore (il quale con il suo atteggiamento può influenzare le prestazioni dell'utente: effetto di Rosenthal e di Pigmalione).

I protocolli verbali (think aloud protocols). I protocolli verbali e i cosiddetti protocolli think aloud sono basati sulla raccolta dei commenti verbali più o meno spontanei che l'utente fa durante una sessione sperimentale, o semplicemente mentre effettua un compito su un sito. Questi protocolli prevedono che all'utente venga chiesto esplicitamente di pensare ad alta voce mentre eseguirà un compito, al fine di registrare le sue reazioni spontanee.

Secondo Nielsen [Nielsen, 1993] i protocolli think aloud sono uno dei metodi più utili nel campo dell'ingegneria dell'usabilità in quanto gli utenti, nel momento in cui verbalizzano spontaneamente i propri pensieri ci aiutano a comprendere il loro reale punto di vista sull'interfaccia e sul funzionamento del sito. Il principale difetto di questi protocolli è che inficiano le misure di performance dell'utente, che non andrebbero considerate in compresenza di un think aloud.

L'osservazione dell'utente. Il metodo dell'osservazione prevede che l'utente sia osservato dallo sperimentatore durante la sessione sperimentale o ancora meglio nel suo naturale contesto d'uso, quando sta interagendo con il sito per portare a termine compiti "reali". L'osservazione dell'utente rientra tra le metodologie qualitative e verrà meglio approfondita nel paragrafo ad esse dedicato.

Aggiungiamo ancora qualche osservazione metodologica riguardante la sua effettiva messa in atto.

L'osservatore può prendere note, audio e/o video-registrare l'utente, usare software che registrano in un file video tutta l'interazione utente-sistema (*screen-capture software*), analizzare i file di log dell'interazione, ecc. Esistono inoltre dei software che analizzano il video e analizzano i pattern d'interazione.

In tutte le osservazioni sarebbe meglio non osservare direttamente gli utenti, ma registrare le loro prestazioni ed analizzare il materiale separatamente. Se

l'osservazione indiretta non è possibile, è importante cercare di non interferire con le azioni dell'utente.

L'osservazione dell'utente è una metodologia molto efficace, ma richiede tempo e denaro.

L'analisi dei file di log. Questo metodo prevede che i dati di log (generati da un web server, o da programmi dedicati) siano usati per tracciare le azioni dell'utente durante la sessione sperimentale o durante l'interazione reale tra utente e sistema.

L'analisi dei file di log è un metodo indiretto di osservazione. Quando viene usato per analizzare i dati sull'utilizzo reale di un sito, può essere considerato come uno dei metodi più affidabili.

6.2 Le fasi di valutazione

Come anticipato nell'introduzione di questo capitolo, le metodologie per valutare un sito web vengono ora presentate seguendo le loro fasi temporali di applicazione.

6.2.1 La fase di analisi dei requisiti

La fase di analisi dei requisiti avviene prima dell'implementazione di un'applicazione interattiva. Essa può essere definita come quel "processo per individuare che cosa un cliente (o meglio un utente) richiede da un sistema software" [Preece et al., 1994]. Durante questa fase può essere utile raccogliere i dati riguardo gli utenti tipo del sito web, nel nostro caso, come ad esempio le caratteristiche, il comportamento, le azioni, i bisogni, l'ambiente, e anche riguardo al dominio di applicazione, alle caratteristiche del sistema, ecc.

Tecniche come i questionari, le interviste, l'osservazione dell'utente, l'analisi (cognitiva) dei task, possono essere usate in questa fase per raccogliere questo tipo di informazione e gettare le basi per la progettazione del nuovo sito.

Secondo Preece l'analisi dei requisiti si articola nelle seguenti fasi:

- *l'analisi dei requisiti funzionali*, che specificano che cosa il sistema deve fare;
- *l'analisi dei dati*, dalla quale emerge la struttura dei dati e delle informazioni che il sistema deve presentare e gestire;
- *l'analisi dei requisiti di usabilità*, nella quale si fissa il livello accettabile delle prestazioni dell'utente e della sua soddisfazione mentre interagisce con il sistema.

Qui di seguito saranno ora elencate le principali metodologie che possono essere utilizzate durante la fase di analisi dei requisiti.

6.2.1.1 *L'analisi dei task.*

Le metodologie per analisi dei compiti (*task*) dell'utente comprendono metodi formali basati su una suddivisione analitica dei compiti che gli utenti potenziali devono poter svolgere quando interagiscono con il sito. Secondo Dix [Dix et al., 1998; Preece et al., 1994] l'analisi dei task può essere definita come quel processo che "analizza il modo in cui le persone svolgono il proprio lavoro: la cose che fanno, il modo in cui si comportano e le cose devono conoscere (...). L'analisi dei task riguarda sistemi esistenti e procedure; i suoi strumenti principali sono quelli dell'osservazione nelle sue varie forme".

Nella maggior parte dei casi, i compiti da analizzare sono scomposti in sotto-compiti. L'analisi gerarchica dei task, per esempio, scompone i compiti in una gerarchia di compiti e sotto-compiti ed utilizza i piani dell'utente per descriverne ordine e condizioni. Ad esempio, supponiamo di star progettando per un sito web la sezione in cui l'utente può gestire la sua posta elettronica e chiediamo quali compiti l'utente deve fare per inviare un'email:

- 1) deve registrarsi per entrare nel suo account di posta;
- 2) deve creare la nuova email;
- 3) scrivere l'email;
- 4) inviare l'email.

Ognuno di questi task può essere ulteriormente scomposto. Ad esempio il task 1 implica che l'utente inserisca email e password. Se non sono corrette le dovrà reinserire. Il task 2 implica invece che l'utente clicchi un pulsante per creare una nuova email, inserisca negli appositi campi (A, Cc, Bcc) l'email di uno o più destinatari (che potrebbero anche essere presenti in rubrica) scriva l'email, inserisca uno o più allegati, imposti delle opzioni nel messaggio (priorità, ricevuta di ritorno, ecc). Ad ognuno dei compiti che l'utente potrà eseguire dovranno corrispondere determinate funzionalità dell'interfaccia.

La raccolta dei dati utili all'analisi dei task può avvenire attraverso l'osservazione dell'utente, le interviste, i questionari, la documentazione esistente, l'analisi di prodotti concorrenti, ecc.

6.2.1.2 *I modelli cognitivi e i modelli socio-tecnici.*

La comprensione dei processi cognitivi che un utente mette in atto mentre svolge un compito e la conoscenza di cui necessita per farlo sono l'oggetto di analisi dei modelli cognitivi e i modelli socio-tecnici. Questi modelli formali, già utilizzati nello studio dell'interazione uomo-macchina, si pongono lo scopo di creare "una determinata rappresentazione dei processi cognitivi e mentali che gli utenti mettono

in atto nel momento in cui interagiscono con una determinata interfaccia; ovvero cercano di modellare aspetti della comprensione, della conoscenza, delle intenzioni e delle forme di ragionamento degli utenti” [Dix et al., 1998]; ed inoltre “cercano di fornire informazioni durante la fase di progettazione attraverso l’applicazione di teorie cognitive” [Preece et al., 1994]. Nello specifico, questi modelli cercano di descrivere non soltanto quali devono essere le opportune sequenze di comportamento degli utenti mentre svolgono una determinata azione, ma anche quello che gli utenti hanno bisogno di sapere per poter svolgere i compiti e come impiegano questa conoscenza sono nella reale esecuzione dei compiti [Dix et al., 1998].

Esempi di modelli cognitivi sono GOMS (Goals, Operators, Methods and Selection) e KLM (Keystroke Level Model).

I modelli socio-tecnici invece considerano non solo le prospettive tecniche ma anche quelle sociali, riconoscendo che la tecnologia è solo una parte del più ampio sistema organizzativo [Dix et, 1998]. L’enfasi di questi approcci è su soluzioni tecniche e sociali ai problemi. Per esempio il modello USTM/CUSTOM si focalizza nello stabilire quelli che sono i requisiti degli stakeholder, che sono qui definiti come tutti coloro che possono venire influenzati dal successo o dall’insuccesso del funzionamento del sistema (per esempio, viene considerato chi usa il sistema, chi riceve l’output e chi fornisce l’input, ecc).

Sia i modelli cognitivi e socio-tecnici che l’analisi dei task sono definiti giustamente da Dix [Dix et al., 1998] *modelli generativi* poiché sono in grado di fornire allo stesso tempo sia meccanismi di valutazione predittiva dell’interfaccia sia specifiche per la progettazione.

6.2.1.3 Il focus group.

Il *focus group* [Greenbaum, 1998], è una metodologia empirica che può essere utilizzata per raccogliere le opinioni e le esigenze degli utenti. Il focus group è normalmente organizzato come una discussione su argomenti specifici moderata da un “group leader”. Un tipico focus group comprende da 8 a 12 utenti potenziali del sito e dura circa due ore. Se gli utenti potenziali possono essere scomposti in gruppi ben identificabili (ad esempio studenti, docenti, personale tecnico-amministrativo) è consigliabile condurre diversi focus group con gruppi rappresentativi degli utenti tipo. Inoltre, a seconda del tipo di utenti coinvolti, i focus group possono essere usati per raccogliere requisiti funzionali, dati, requisiti di usabilità.

I focus group sono un’eccellente alternativa per raccogliere opinioni spontanee e idee, ma non forniscono informazioni su quello che gli utenti realmente fanno quando stanno interagendo con il sito.

I focus group possono anche essere usati nella valutazione partecipata (vedi.

6.3).

6.2.1.4 L'osservazione sistematica.

L'osservazione sistematica è una metodologia empirica che ha lo scopo di quantificare il comportamento dei soggetti osservati, in particolare quei comportamenti ricorrenti che si possono osservare in un contesto reale [Bakeman e Gottman, 1986]. Lo scopo è quello di definire anticipatamente dei **codici comportamentali** e chiedere agli osservatori di registrare la frequenza con cui questi codici predefiniti si presentano durante l'osservazione (ad esempio dei codici potrebbero essere: quante volte l'utente stampa una pagina web; quante volte nella navigazione utilizza motori di ricerca; quanto spesso usa il pulsante "Indietro", ecc).

Uno dei principali problemi legati a questa metodologia riguarda la formazione degli osservatori, che deve essere coerente e omogenea, in quanto essi devono essere messi in grado di riconoscere le stesse sequenze di comportamento.

L'osservazione degli utenti può essere realizzata utilizzando tecniche sequenziali o tecniche non sequenziali. Nelle analisi condotte con **tecniche non sequenziali** i soggetti sono osservati per intervalli temporali fissi in momenti diversi. Ad esempio, gli utenti possono essere osservati per un minuto ogni giorno mentre interagiscono con la intranet aziendale, ma ad ore diverse della giornata (mattino presto, tardo pomeriggio, prima di pranzo, ecc). Questo metodo è anche chiamato **metodo degli intervalli di tempo ripetuti**. L'osservazione non sequenziale può essere usata, ad esempio, per rispondere a domande su come gli utenti distribuiscono il loro tempo tra le diverse attività e in base a questo raccogliere spunti per l'analisi dei requisiti.

Nell'**osservazione sequenziale** invece ogni soggetto è osservato per un certo periodo di tempo e, in base a quanto emerge, vengono assegnati i codici comportamentali che meglio caratterizzano momenti contigui di attività. Per esempio ogni soggetto potrebbe essere osservato per 100 minuti, e poi l'osservatore, una volta analizzato il filmato, potrebbe decidere quale dei codici comportamentali meglio caratterizza ogni intervallo di 15 minuti. Le tecniche sequenziali si adattano meglio a rispondere a domande su come il comportamento si distribuisce nel tempo minuto per minuto.

6.2.2 La fase di valutazione preliminare

La fase preliminare di valutazione può essere portata avanti durante lo sviluppo del sito. In questa fase è di fondamentale importanza effettuare una o più valutazioni per evitare costose e complesse ri-progettazioni del sito una volta che è già terminato. La fase di valutazione preliminare può essere basata su metodi analitici

(in tal caso si rientra in quella che in letteratura viene normalmente definita come *valutazione predittiva*²) o metodi empirici (in tal caso si utilizzano metodologie proprie della *valutazione formativa*³).

6.2.2.1 La valutazione euristica.

Un'euristica può essere definita come una linea guida di riferimento o una regola pratica (ossia derivata dall'esperienza d'uso, o da Precedenti valutazioni) che può essere utilizzata sia come punto di riferimento in fase di progettazione sia per criticare esistenti soluzioni di interfaccia.

La valutazione euristica [Nielsen e Molich, 1990] è una metodologia di valutazione analitica nella quale un ristretto numero di valutatori esamina un'interfaccia alla ricerca di problemi che violino i principi e le linee guida di buona progettazione. Nielsen e Molich consigliano di coinvolgere da tre a cinque valutatori. Questi ultimi possono essere sia esperti di usabilità che non esperti. Gli esperti di ingegneria di usabilità si sono dimostrati più efficaci nel trovare i problemi rispetto agli esperti di ingegneria del software. Sempre secondo Nielsen e Molich gli esperti di usabilità che hanno anche familiarità con gli aspetti più tecnologici e informatici sono i valutatori migliori. Se i valutatori non sono propriamente degli esperti possono comunque provare ad effettuare una valutazione di tipo euristico. L'importante è che, manuali alla mano, si mettano a verificare che i principi fondamentali dell'interazione uomo-macchina e le linee guida di usabilità siano rispettate e correttamente applicate.

La valutazione euristica può anche essere effettuata durante la fase finale di valutazione, ovvero a sito finito. Tuttavia, al fine di evitare complesse spese di riprogettazione sarebbe meglio effettuarla in fase iniziale, e preferire le metodologie empiriche sommative (ad esempio i test di usabilità) nella fase finale.

Al fine di aiutare i valutatori, Nielsen e Molich suggerirono una lista di euristiche, derivanti da linee guida e principi ben consolidati dell'interazione uomo-macchina, di cui un progettista deve tenere conto:

- utilizzare un dialogo semplice e naturale;

²I metodi predittivi hanno lo scopo di fare previsioni basate sulla valutazione di esperti circa quelle che saranno le performance di un sistema interattivo e di conseguenza prevenire degli errori di progettazione. I metodi predittivi solitamente si rifanno a corpi di linee guida e principi condivisi di buona progettazione.

³I metodi formativi hanno l'obiettivo di valutare le prime scelte di progettazione ed ottenere suggerimenti per rivedere l'interfaccia in un ciclo continuo di progettazione-valutazione-riprogettazione.

- parlare il linguaggio dell'utente;
- minimizzare il carico cognitivo della memoria a breve a termine;
- essere consistenti;
- fornire feedback;
- fornire delle chiare vie di uscita;
- fornire delle scorciatoie;
- fornire messaggi d'errore comprensibili;
- prevenire gli errori.

Pur trattandosi di principi pensati per un'interfaccia web, essi risultano comunque validi nella progettazione delle interfacce in generale e sono in un certo senso completati dai requisiti generali dell'usabilità del software proposti da J. Nielsen nel 1993 [Nielsen, 1993]:

- *facilità di apprendimento*: puntare allo “zero learning time” ossia ad un tempo di apprendimento zero dovuto alla facilità dell'interfaccia;
- *efficienza d'uso*: definisce la capacità di soddisfare pienamente gli scopi per i quali il software e l'interfaccia sono stati progettati. è un parametro misurabile solo tramite la valutazione empirica del sistema;
- *facilità di comprensione*: favorire il riconoscimento piuttosto che il recupero dell'informazione nella memoria;
- *reversibilità degli errori*: ridurre gli errori e il loro impatto sul risultato finale;
- *soddisfazione nell'uso*: rendere piacevole e confortevole l'interazione.

Le euristiche di Nielsen e Molich sono poi state successivamente rielaborate dallo stesso Nielsen dopo studi successivi⁴:

- visibilità dello stato del sistema;
- corrispondenza fra il sistema e il mondo reale;
- libertà e controllo dell'utente;
- consistenza e standard;
- prevenzione degli errori;
- riconoscimento anziché ricordo;
- flessibilità e efficienza d'uso;
- estetica e design minimali;
- aiuti al riconoscimento e risoluzione degli errori;
- aiuti e documentazione.

⁴ http://www.useit.com/papers/heuristic/heuristic_list.html

6.2.2.2 *Il cognitive walkthrough.*

Il cognitive walkthrough è un metodo analitico di valutazione nel quale l'esperto, o gli esperti, svolgono il ruolo dell'utente, ovvero simulano quello che l'utente potrebbe fare sul sito, al fine di scoprire eventuali errori di usabilità nella progettazione dell'interfaccia. Il principio ispiratore di questa metodologia è quello dell'"apprendere attraverso l'esperienza". È stato infatti dimostrato che molti utenti preferiscono imparare ad utilizzare un sistema interattivo attraverso l'esperienza, ovvero esplorandone le sue funzionalità attraverso il reale utilizzo, e non dopo avere studiato il manuale per sapere come il sistema funziona [Dix et al., 1998]. Di conseguenza, secondo la metodologia del walkthrough cognitivo, partendo dall'assunzione che l'utente apprende ad usare l'interfaccia attraverso l'esplorazione dalla stessa, uno o più esperti di usabilità simulano l'interazione e i compiti fondamentali che l'utente potenziale può fare seguendo un determinato script. Dopodiché gli esperti devono rispondere ad una serie di domande concernenti le decisioni e le scelte che gli utenti devono prendere mentre utilizzano l'interfaccia (le domande possono riguardare, per esempio, questioni come quanto è facile identificare le conseguenze delle azioni fatte sul sito, valutare i progressi verso il raggiungimento di un determinato obiettivo, ecc). Alla fine le risposte che mettono in luce dei problemi finiscono nella lista delle cose da rivedere.

Il cognitive walkthrough può anche essere usato dopo una valutazione sperimentale, o un test di usabilità (post-task walkthrough). In questa modalità ai soggetti viene chiesto di riflettere e commentare le loro azioni. Ad esempio, il valutatore potrebbe loro chiedere il perché di certe scelte e azioni, anche riproponendo la stessa registrazione delle performance dell'utente.

6.2.2.3 *Il Wizard of Oz.*

Nella metodologia di valutazione empirica chiamata Wizard of OZ [Nielsen, 1993; Preece et al. 1994] si fa interagire l'utente con il sistema, che è in realtà guidato e quindi simulato dallo sperimentatore, il quale in tempo reale modifica le risposte del sistema in base agli input trasmessi dall'utente. L'utente, tuttavia, non è consapevole del "trucco" e interagisce con quest'emulazione del sistema ignaro del fatto che lo sperimentatore (il mago) stia guidando le risposte.

6.2.2.4 *I prototipi (la valutazione dei).*

Un prototipo di un sito web può essere definito come un artefatto che simula o anima alcune, ma non tutte, le caratteristiche del sito finale [Dix et al., 1998].

I prototipi si possono dividere in due categorie principali:

- **prototipi statici**, che generalmente consistono in una serie di videate dell'interfaccia viste dal computer o stampate o infine direttamente

disegnate su carta. La tecnica dei prototipi statici può anche essere utilizzata per illustrare uno scenario di possibili azioni dell'utente. Si possono infatti simulare in maniera statica le interfacce di partenza e i risultati delle azioni dell'utente e simulare passo-passo lo scenario di quello che succederà chiedendo poi all'utente di eseguire una determinata sequenza di azioni;

- **prototipi interattivi**, basati su software, che possono consistere in un'iniziale e parziale implementazione del sistema, oppure possono essere realizzati con software specifici che sono in grado di simulare l'aspetto e la resa finale (il cosiddetto look and feel) dell'interfaccia del sito. I prototipi interattivi possono essere:
 - **orizzontali**, quando si sviluppano in ampiezza e contengono una visione di superficie di quello che sarà il sito finale, ovvero implementano solo il primo livello dell'albero dei nodi del sito;
 - **verticali**, quando si sviluppano in profondità andando ad includere un limitato numero di percorsi tra tutti i possibili nodi del sito, ovvero implementano in profondità alcune delle sezioni del sito;
 - **basati su scenari**, quando implementano in maniera completa alcuni dei task fondamentali che l'utente potrà effettuare sul sito. Ad esempio, in un sito di commercio elettronico potrebbe essere implementato solo lo scenario che porta alla ricerca e all'acquisto di un prodotto, mentre le restanti parti potrebbero ancora essere in fase embrionale.

I test di usabilità dei prototipi sono piuttosto comuni e si rivelano utili, in quanto consentono agli sviluppatori di effettuare cambiamenti prima che sia troppo tardi. Uno studio condotto da Nielsen [Nielsen, 1990] ha dimostrato che i prototipi interattivi sono più efficaci nell'individuare i problemi generali di usabilità dell'interfaccia piuttosto che quelli basati su carta. In ogni caso la valutazione dei prototipi si rivela ottima quando si devono testare delle possibili soluzioni di interfaccia in fase formativa, ma non può e non deve sostituire il test di usabilità del sito finito.

Da notare che i prototipi basati su carta sono invece molto utili quando si deve simulare l'utilizzo di un'interfaccia non basata su PC. Sono infatti molto usati per testare le prime soluzioni di progettazione di siti web per ambiente *mobile* o tv. Per il un esempio di paper prototyping in ambiente *mobile* si veda [Weiss, 2002].

Infine, è bene sottolineare che la valutazione di prototipi può essere usata nella cosiddetta **progettazione parallela** (*parallel design*), la quale consiste nella progettazione di diverse alternative di interfaccia e di funzionamento del sistema che saranno poi selezionate attraverso una qualche valutazione (ad esempio potrebbe essere fatto un test di usabilità sui diversi prototipi realizzati). L'alternativa che risulterà migliore verrà poi ulteriormente sviluppata e

implementata.

Anche la **progettazione razionale** (*design rationale*) e l'**analisi dello spazio delle alternative di progettazione** (*design space analysis*) possono essere utili nei casi di progettazione parallela. Il primo approccio [Dix et al, 1998] consiste nell'allegare descrizioni funzionali ed architetture relativamente alle scelte di interfaccia, per meglio spiegare il comportamento del sistema, mentre il secondo è un particolare approccio alla progettazione che incoraggia il designer dell'interfaccia a esplorare e ragionare sulle diverse alternative di progettazione.

6.2.2.5 Il card sorting.

Il *card sorting* [Rosenfeld e Morville, 2002] è una metodologia empirica utilizzata nell'ambito dello sviluppo dell'architettura d'informazione di un sito che ha lo scopo di esplorare come gli utenti di un sito classificano dei concetti, ed è molto utile per definire le strutture di navigazione di un sito. Come descritto nel Capitolo 2, le sezioni e le sotto-sezioni di navigazione di un sito si possono basare su schemi organizzativi ambigui. Il card sorting permette appunto di far valutare agli utenti target del sito questi schemi (card sorting chiuso) oppure di chiedere a loro stessi dei suggerimenti per la creazione degli schemi, attraverso una metodologia empirica (card sorting aperto).

Nello specifico, nel **card sorting chiuso** si chiede ad un gruppo formato da una quindicina di utenti⁵ di classificare dei concetti-unità informative in categorie già definite. Normalmente le categorie corrispondono alla struttura gerarchica delle sezioni, mentre i concetti-unità informative a sotto-sezioni, pagine e/o servizi del sito. Ad esempio, si potrebbe prendere la struttura di navigazione globale di un sito (es., Politica, Sport, Economia, Spettacoli, Cultura, Esteri) e chiedere agli utenti di classificare concetti (Calcio, Cinema, Tassi d'interesse, ecc) e unità informative (recensione di un libro, notizie finanziarie, servizi prenotazione spettacoli, ecc) all'interno di queste categorie.

Invece, nel **card sorting aperto** si procede per induzione. Viene infatti chiesto ad un gruppo di utenti di trovare delle categorie omogenee all'interno delle quali raggruppare una serie di concetti-informazioni che vengono loro presentati. Riprendendo l'esempio di sopra, si potrebbero prendere delle unità informative (recensione di un libro, notizie finanziarie, servizi prenotazione spettacoli, ecc) e chiedere agli utenti di raggrupparle in categorie esclusive, esaustive e significative.

Il card sorting è una tecnica in grado di far aumentare l'usabilità finale di un sito web perché consente di creare un'architettura dell'informazione che riflette il punto di vista degli utenti. In questo modo diventa possibile sviluppare strutture che aumentano la probabilità che gli utenti riescano a trovare ciò che cercano.

⁵ <http://www.useit.com/alertbox/20040719.html>

6.2.3 La fase finale di valutazione

La fase finale di valutazione avviene a progetto finito ed ha lo scopo di valutare la qualità complessiva del sito web con il coinvolgimento degli utenti finali.

6.2.3.1 L'ingegneria dell'usabilità e le metriche di usabilità.

Prima di affrontare la metodologia più comune per valutare l'usabilità di un sito, ovvero il test di usabilità, è bene fare un passo indietro e descrivere brevemente l'approccio dell'ingegneria dell'usabilità [Nielsen, 1993; Whiteside et al, 1988], e le metriche ad esso correlate, le quali verranno riprese appunto dalle metodologie empiriche di testing.

Le tecniche dell'ingegneria dell'usabilità derivano direttamente dall'ingegneria del software. In particolare, viene ripreso il concetto di *ciclo di vita del software*, che descrive le attività che si avvicinano dall'iniziale formazione dei concetti fino all'effettiva messa in opera del sistema. Uno dei punti chiave del ciclo di vita del software (che viene preso in prestito dall'ingegneria dell'usabilità) è il concetto di valutazione, che deve avvenire durante tutte le fasi del ciclo, sin da quella iniziale.

Tuttavia la differenza fondamentale tra il concetto di valutazione nell'ingegneria del software (incentrata sui concetti di validazione e verifica) e quello nell'ingegneria dell'usabilità riguarda la centralità della prospettiva dell'utente, non supportata dalla prima, fondamentale nella seconda.

Per gli ingegneri dell'usabilità i requisiti di un sistema interattivo non possono essere stabiliti dall'inizio, senza avere osservato e valutato gli utenti per determinare come essi interagiscono con il sistema.

Whiteside et al. [Whiteside, Bennett and Holtzblatt, 1998] ci hanno fornito una lista di metriche, chiamate appunto **metriche di usabilità**, che possono essere usate per determinare il grado di usabilità di un sistema interattivo durante l'interazione utente-sistema:

- quantità di tempo impiegata per eseguire un compito;
- percentuale di compiti completati;
- percentuale di compiti completati per unità di tempo;
- rapporto tra successi e fallimenti;
- tempo sprecato negli errori;
- percentuale o numero di errori;
- percentuale o numero di concorrenti migliori;
- numero di comandi usati;
- frequenza nell'uso dei comandi di help e documentazione;
- percentuale di comandi favorevoli e sfavorevoli;
- numero di ripetizioni di comandi sbagliati;

- numero di volte in cui l'interfaccia ha ingannato l'utente;
- numero di buone o cattive caratteristiche dell'interfaccia;
- numero di comandi disponibili non invocati;
- quantità di tempo che gli utenti usano per risolvere un problema;
- numero di volte che l'utente ha perso il controllo del sistema;
- numero di volte in cui l'utente esprime frustrazione o soddisfazione;
- ...

Da come si può evincere da queste metriche, la valutazione dell'usabilità di un software è basata sulla misurazione delle azioni dell'utente, che non possono essere rilevate quando il sistema non è ancora stato implementato. Esse ci forniscono quindi una lista di misurazioni che andranno poi applicate durante un test empirico.

6.2.3.2 Il test di usabilità.

Per essere certi dell'usabilità di un sito web, così come di un altro sistema interattivo, la cosa migliore è eseguire un test di usabilità (per approfondimenti si veda [Dumas e Redish, 1999; Rubin, 1994]). Secondo [Dumas e Redish, 1999]

“l'usabilità si riferisce a come le persone usano i prodotti. Testare l'usabilità significa essere sicuri che le persone possano trovare e usare delle funzionalità che incontrino i loro bisogni.”

Sempre secondo i due autori un test di usabilità è caratterizzato dai seguenti elementi distintivi:

- l'obiettivo principale è migliorare l'usabilità di un sistema interattivo;
- i partecipanti al test devono eseguire compiti reali, ovvero i compiti per i quali il sistema è stato progettato e che gli utenti reali eseguiranno abitualmente. Ad esempio, nel caso di un forum sul web, le azioni che vanno immediatamente testate sono le attività principali che gli utenti vorranno fare sul sito: registrarsi, effettuare la login, inserire messaggi e discussioni, cercare altri utenti, ecc;
- le performance dei partecipanti sono registrate. A tal proposito si possono utilizzare dei software che registrano tutte le azioni che accadano sullo schermo del computer (screen capture software), dei software che tengono traccia delle azioni dell'utente (logging software), video-registrare gli utenti con videocamere, web cam, ecc;
- si analizzano i dati raccolti e in base ai risultati si fanno gli opportuni cambiamenti sull'interfaccia. Le metriche da prendere in considerazione nell'analisi dei dati sono appunto le metriche di usabilità appena elencate. A seconda dei compiti e degli obiettivi si raccolgono i dati utili a misurare le metriche significative. È importante sottolineare che in un test di usabilità i dati vengono poi analizzati attraverso l'utilizzo di statistiche descrittive, come

la media, la varianza, la deviazione standard, che hanno lo scopo di descrivere e riassumere una serie di dati.

Testare un sito con degli utenti è un attività empirica fondamentale, in quanto è l'unico modo per sapere se il sito è veramente usabile e rispecchia il modello mentale⁶ dell'utente tipo.

Nel continuare questa nostra descrizione dei test di usabilità prenderemo in prestito quando brillantemente sostenuto da Krug nel suo ottimo libro *Don't make me think* [2006], in quanto pienamente condiviso dagli autori.

Secondo Krug la garanzia di avere un "gran sito" web è appunto il test di usabilità: testare un sito anche con un solo utente è 100 volte meglio che non testarne nessuno, ed è meglio un utente all'inizio del progetto che 50 alla fine. Niente batte una reazione dal vivo!

Questo ci introduce dunque alla tematica concernente il numero di utenti da coinvolgere in un test.

In uno dei suoi famosi AlterBox, Nielsen [Nielsen, 2000] ci dimostra che, per ottenere dei buoni risultati, bastano anche solo 5 utenti, con i quali si può arrivare a scoprire circa l'80% dei problemi: oltre questo numero infatti il numero di problemi di usabilità scoperti cresce più lentamente fino ad arrivare al 100% dei problemi scovati con circa 15 utenti. Quello che Nielsen consiglia è di svolgere più test di usabilità, ognuno con 5 utenti: fare un test con 5 utenti, ri-progettare il sito a seconda dei problemi emersi, rifare in test con altri 5 utenti, ri-progettare, testare di nuovo e infine ri-progettare di nuovo! Dunque secondo Nielsen è meglio fare 3 test, ciascuno con 5 utenti, che un solo test con 15 utenti, in un processo continuo di valutazione e riprogettazione.

Come Nielsen, ci insegna e come osserviamo all'inizio la valutazione si configura dunque come un processo iterativo.

Quale tipo di utenti conviene scegliere per un test di usabilità?

Secondo Krug [Krug, 2006] l'importanza di reclutare utenti rappresentativi è sovrastimata, in quanto, a differenza di un esperimento controllato (vedi 6.2.3.4) lo scopo del test non è quello di provare o confutare una teoria, ma semplicemente di valutare se l'interfaccia funziona bene.

Mentre, secondo gli approcci classici, gli utenti del test dovrebbero sempre essere degli utenti rappresentativi del target a cui è indirizzato il sito, secondo Krug basta prendere chiunque (amici, colleghi, vicini di casa) e, al limite, affinare le

⁶ I modelli mentali sono rappresentazioni della realtà che le persone utilizzano per comprendere specifici fenomeni. Secondo Norman [Norman, 1997] le persone, quando interagiscono con gli artefatti tecnologici, formano dei modelli mentali di se stessi e degli oggetti con cui stanno interagendo. Questi modelli hanno un forte potere predittivo ed esplicativo che consente di comprendere come funziona l'interazione.

scelta successivamente: in fondo davanti ad un nuovo sito siamo tutti principianti ed in ogni caso non è buona idea progettare un sito che solo la nostra audience riesce ad usare, quindi è meglio semplificare per tutti.

Ovviamente ci sono delle eccezioni a questa regola. Gli utenti devono essere rappresentativi quando:

- il sito è usato da un preciso gruppo di utenti facili da reperire;
- l'audience è divisa in gruppi ben definiti, dagli interessi e dai bisogni divergenti;
- usare il sito richiede specifiche competenze.

Famosa è la distinzione di Krug tra test di usabilità tradizionale e test di usabilità a *buon mercato* (*discount usability test*). Se vi ha tempo è disponibilità un test tradizionale è da preferirsi. Altrimenti, anche un test in svendita può comunque produrre risultati sufficienti e aiutarci a raffinare la nostra interfaccia.

In un **test tradizionale**:

- gli utenti sono circa 8 e vengono scelti accuratamente tra gli utenti target del sito;
- il test viene svolto in un laboratorio di usabilità ;
- chi conduce il test è un professionista con esperienza nel campo della valutazione dell'usabilità;
- il test è pianificato e programmato con settimane di anticipo: lo script del test (i compiti che gli utenti devono fare) è discusso e rivisto con settimane di anticipo;
- quello che viene testato è il sito finito;
- durata: 10-15 giorni;
- dopo avere eseguito il test e analizzato i risultati, viene presentata una relazione approfondita di una ventina di pagine.

Invece, in un test di **usabilità a buon mercato**:

- si possono coinvolgere anche solo 3 o 4 utenti con l'unico requisito che abbiano familiarità con il web;
- il test può essere svolto in qualunque ufficio o sala conferenze;
- chi conduce il test è una persona paziente;
- il test è pianificato con un minimo di anticipo: basta decidere cosa mostrare
- il test non è eseguito a sito finito ma vengono eseguiti piccoli e continui test lungo tutto il processo di sviluppo;
- durata: 2-3 giorni;

- durante il test il valutatore prende una paginetta di appunti e se ne discute con i progettisti il giorno stesso o il giorno dopo.

Aggiungiamo inoltre che in un test tradizionale gli utenti andrebbero sempre remunerati, e che prima di entrambi i tipi di test sarebbe sempre bene effettuare un **pre-test** con un paio di utenti per essere certi che tutto funzioni per il meglio.

Nelle pagine che seguono è mostrato un esempio dei compiti del test di usabilità effettuato nel 2003 da uno degli autori per valutare la sezione del monitoraggio delle acque del sito web della Direzione delle Risorse Idriche della Regione Piemonte⁷.

Notate che, prima dell'elenco dei compiti che gli utenti devono effettuare durante test (che andrebbero sempre presentati in forma scritta, per evitare eventuali interferenze con il valutatore) viene mostrata una lettera di presentazione del test, che andrebbe sempre allegata allo script del test.

⁷ <http://www.regione.piemonte.it/acqua/index.htm>

Torino, 22 marzo 2004

Oggi testeremo il sito della Direzione delle Risorse idriche della Regione Piemonte per vedere come i reali utenti lo usano.

Questo è un test del sito e non un test sulle tue capacità. Tu non puoi fare nessun errore. Al limite gli errori sono nel sito!!

Vorremmo sapere esattamente che cosa pensi del sito, quindi non preoccuparti di offendere i nostri sentimenti: dobbiamo migliorare il sito e abbiamo bisogno di sapere cosa ne pensi veramente.

Il tuo test verrà registrato (con il tuo permesso) da un software installato su questo computer per poter poi ri-analizzare il modo in cui tu hai usato il sito e apportare i dovuti miglioramenti. Questo ci aiuta, in quanto non riusciremo a prendere nota di tutto e qualcosa potrà sfuggire.

Il test consiste in una serie di compiti da svolgere sul sito. Quando avrai terminato un compito potrai passare a quello successivo.

Se hai delle domande durante lo svolgimento dei compiti chiedi pure, ma solo se non sai più che fare...Ci interessa vedere come gli utenti usano il sito quando sono da soli, quindi ricorri a delle domande solo se stai per abbandonare lo svolgimento del compito.

Hai qualche domanda prima di iniziare?

Ti ringraziamo per la tua collaborazione

Cristina Gena
Dipartimento di Informatica
Università di Torino

Compiti del test

- 1) Cerca i valori dei parametri biologici del fiume Po dell'anno 2000. Scaricali sul desktop.
- 2) Cerca la carta piezometrica media della falda superficiale e visualizzala.
- 3) Cerca i dati di portata del fiume Po dell'anno 2000.
- 4) Cerca le caratteristiche della stazione di monitoraggio qualitativo di Torino del bacino del Po. Scarica la scheda sul tuo computer.
- 5) Cerca e scarica sul desktop la carta dello stato ecologico e della balneabilità dei laghi piemontesi.
- 6) Cerca e visualizza gli indici di stato chimico di un punto di monitoraggio sotterraneo presente a Torino.
- 7) Visualizza i dati di monitoraggio per la definizione dello stato ambientale - qualitativo - del lago di Avigliana.

Cerca informazioni dettagliate (testuali) sul monitoraggio da inserire in una ricerca a riguardo.



Figura 6.1. Interfaccia del sito testato (<http://acqua.di.unito.it/>), ora anche reperibile all'indirizzo: <http://gis.csi.it/acqua/>

Come risulterà chiaro più avanti, il test di usabilità ha molti aspetti in comune con un esperimento controllato (vedi 6.2.3.4). La differenza principale risiede nel fatto che, mentre un test di usabilità ha lo specifico obiettivo di valutare l'usabilità di un sito web o dell'interfaccia di un sistema interattivo, un esperimento controllato è basato su delle ipotesi più o meno complesse che devono essere verificate. In ogni caso, prima di iniziare un esperimento controllato, sarebbe bene eseguire prima un test di usabilità, al fine di evitare che i problemi di usabilità possano inficiare i risultati sperimentali.

Per riassumere, le **similarità** tra le due metodologie di valutazione sono che:

- entrambe si svolgono in laboratorio;
- i partecipanti sono degli utenti rappresentativi della popolazione di riferimento;
- vengono prese delle misurazioni soggettive ed oggettive;
- i dati vengono analizzati e presentati in un report scritto.

Mentre le **differenze**:

- gli obiettivi sono diversi: il test di usabilità ha lo scopo di migliorare l'interfaccia, l'esperimento vuole dimostrare la verità di una o più ipotesi;
- i problemi delle interface posso essere generati da molteplici fattori, mentre in un esperimento si cerca di isolare un numero limitato di variabili e di tenerle sotto osservazione;
- per l'analisi dei dati di un test sono sufficienti statistiche descrittive, mentre in un esperimento è richiesta l'applicazione di statistiche inferenziali;
- nel test di usabilità sono importanti anche fattori qualitativi come l'osservazione e la soddisfazione degli utenti.

6.2.3.3 La valutazione automatica dell'usabilità.

Negli ultimi anni c'è stato un interesse crescente verso i software che eseguono in maniera automatica delle valutazioni di usabilità. I metodo automatici per la valutazione dell'usabilità si suddividono in [Paternò, 2002]:

- metodi basati su **valutazione empirica**, nei quali vengono analizzati i file di log degli utenti generati da un server web,
- metodi basati su **valutazione analitica**, nei quali l'analisi viene fatta combinando criteri, linee guida e modelli.

Nel primo caso l'analisi dei dati sul reale utilizzo di un sito web viene

considerata come una soluzione ideale per scoprire reali problemi di interazione. L'analisi dei dati di utilizzo di un sito (per dettagli si vedano [Spiliopoulou, 2000; Mobasher, 2000; Pierrakos, 2003]) è un lungo processo che porta a conoscere il sito stesso secondo quella che è la prospettiva dei suoi utenti reali: nuovi pattern di utilizzo possono emergere dai dati (come ad esempio pagine sovente associate, che vengono visualizzate spesso una dopo l'altra, fatto del quale anche l'interfaccia dovrebbe tenere conto). Questo potrebbe essere il segnale che molti utenti navigano in maniera diversa da quello che era stato progettato.

Il processo di raccolta dei dati di utilizzo di un sito (*web usage mining process*) può portare alla scoperta di pattern sequenziali di utilizzo, raggruppamenti di pagine, raggruppamenti di utenti simili (o meglio aventi lo stesso comportamento sul sito). Una volta raccolti un numero sufficiente di dati e raggiunta qualche evidenza, anche confermata da analisi statistiche, il re-design del sito può essere effettuato in due modi [Perkowitz, 2000]:

- per **trasformazione**, migliorando la struttura del sito sulla base delle interazioni di tutti i visitatori;
- per **personalizzazione** (vedi sez. 4.8), adattando ai bisogni di ogni singolo visitatore la struttura del sito. Questa seconda opzione si può attuare solo se si riesce a riconoscere il visitatore quando accede al sito web.

Oltre a queste due alternative una terza soluzione può essere adottata: personalizzare il sito a seconda dei possibili gruppi di utenti emersi dall'analisi empirica dei dati. Ad esempio, il sito potrebbe proporre diverse strutture di navigazione per utenti occasionali o regolari, utenti poco esperti o esperti, ecc

I metodi analitici per la validazione automatica dell'usabilità comprendono invece strumenti automatici come ad esempio:

- Webxact⁸, che verifica che il codice della pagina sia conforme alle linee guida di accessibilità;
- WebSat⁹, che valuta l'usabilità analizzando il codice della pagina attraverso l'applicazione di linee guida di usabilità;
- Design Advisor che si basa su tecniche di eye-tracking¹⁰.

Esistono infine approcci misti che mettono insieme l'analisi dei file di log con

⁸ <http://webxact.watchfire.com/>

⁹ <http://zing.ncsl.nist.gov/WebTools/WebSAT/overview.html>

¹⁰ I sistemi di eye-tracking sono dei sistemi di tracciamento oculare per la valutazione dell'usabilità che registrano le direzioni dello sguardo dell'utente mentre sta interagendo con il sito.

le linee guida di usabilità e i modelli delle azioni dell'utente [Paternò, 2002].

6.2.3.4 *Gli esperimenti controllati.*

Questo tipo di valutazione empirica ha l'obiettivo di dimostrare la veridicità di una teoria attraverso l'osservazione in situazioni sperimentali.

Gli esperimenti controllati in HCI sono stati presi in prestito dalla psicologia cognitiva e sperimentale. Le origini della relazione tra psicologia e informatica posso essere riportate alla necessità sviluppare una connessione tra le due. In particolare, nel passato si è cercato di utilizzare teorie cognitive per capire cosa succede quando le persone interagiscono con i computer e di conseguenza questo può risultare utile per progettare correttamente l'interfaccia.

Come sostengono [Dix et al., 1998], dal momento che non esistono teorie psicologiche predittive, se si introducono nuove soluzioni di interfaccia (pensiamo soprattutto agli albori del web, e adesso con l'avvento del Web 2.0) queste devono essere testate e misurate in un esperimento al fine di sapere quale impatto potranno avere durante il loro reale utilizzo.

L'idea di base di un esperimento controllato è che al variare di un elemento in una situazione controllata il comportamento dell'utente possa variare ed essere misurato. Gli aspetti più importanti di un esperimento sono che

- i partecipanti devono essere credibili: devono essere utenti reali dell'applicazione che si sta andando a valutare;
- i compiti sperimentali devono essere credibili: i soggetti devono svolgere dei compiti che svolgeranno quando andranno a utilizzare realmente l'applicazione;
- i partecipanti devono essere osservati durante l'esperimento e la loro performance registrata (prendendo appunti, video-registrando, utilizzando software di loggia o screen capture software, ecc).

Gli esperimenti controllati dovrebbero avere luogo in un laboratorio ben equipaggiato con strumenti per registrazioni audio e video, specchi bidirezionali e computer attrezzati con opportuni software.

Il laboratorio presenta svantaggi e vantaggi: da un lato ci si trova in una situazione troppo artificiosa e innaturale, de-contestualizzata, molto lontana da quello che sarà il contesto reale di utilizzo. Dall'altro il laboratorio potrebbe essere l'unica opzione, come ad esempio nel caso in cui si vogliono creare delle situazioni non ancora esplorate [Dix et al. 1998].

In un esperimento base almeno due gruppi di soggetti sono sottoposti alle stesse condizioni sperimentali con una sola differenza nel trattamento, che è quella imputabile alla variabile indipendente. Ogni differenza osservata nel comportamento dei due gruppi può essere attribuita, e quindi causata, dal differente trattamento.

Secondo [Keppel et al., 1992] il processo schematico di un esperimento controllato può essere schematizzato nei seguenti passi.

Identificare l'argomento o la questione d'interesse. La maggior parte delle ricerche inizia con delle domande che il ricercatore vuole risolvere.

Studiare le teorie rilevanti e le ricerche correlate per vedere quello che gli altri hanno fatto e scritto sull'argomento e come hanno cercato di risolvere la questione.

Stabilire le ipotesi di ricerca da testare nell'esperimento.

Identificare le variabili dipendenti e quelle indipendenti. Le **variabili indipendenti** comprendono tutti quei trattamenti e condizioni che possono essere tenute sotto controllo dallo sperimentatore, ovvero esse possono essere manipolate o variate. A seconda della complessità dell'esperimento una o più variabili indipendenti possono essere considerate. Ad esempio, un esperimento controllato potrebbe voler valutare l'ipotesi che alcune tonalità di colori tendono a favorire gli utenti nello svolgimento dei compiti su un sito web. Nel caso più semplice si potrebbe testare l'efficacia di un solo gruppo di colori, mentre in casi più complessi si potrebbe valutare l'impatto di diversi tipi di colori (pastello, rossi, scuri, ecc).

Le **variabili dipendenti** comprendono i comportamenti osservati e misurati dopo la manipolazione della variabile indipendente. Esempi classici di variabili indipendenti sono

- il tempo necessario a portare a termine un compito,
- il numero e la qualità dei compiti portati a termine,
- gli schemi di interazione,
- la percentuale di apprendimento,
- il numero dei click effettuati per portare a termine un compito,
- l'utilizzo del bottone di back
- il ritorno alla home page
- ...

Altre variabili possono essere misurate indirettamente, come ad esempio:

- il carico cognitivo misurato attraverso la pressione sanguinea,
- la dilatazione delle pupille,
- il tracciamento della direzione dello sguardo dell'utente (eye-tracking),
- il numero e la durata degli sguardi
- il battito cardiaco,
- la temperatura della pelle,
- la pressione esercitata sul mouse,

- ...

Selezionare i soggetti. Anche se la popolazione di riferimento di un esperimento è ovviamente limitata, è spesso difficile, se non impossibile, raccogliere tutti i dati riguardanti i membri di quella popolazione. Gli psicologi normalmente fanno inferenze sulla popolazione di riferimento basandosi su informazioni raccolte da un campione di soggetti appartenenti alla popolazione stessa.

Quindi l'obiettivo delle tecniche di campionamento è proprio quello di raccogliere dati da campioni rappresentativi della popolazione di riferimento per fare inferenze sul comportamento di questi soggetti.

Il modo migliore per avere un campione rappresentativo è quello di **scegliere casualmente** i soggetti dalla popolazione target. In questo modo tutti gli appartenenti alla popolazione hanno la stessa probabilità di essere selezionati.

Il campionamento casuale assicura una grande **validità esterna**, che può essere definita come l'abilità di fare generalizzazioni su un'intera popolazione di riferimento a partire da un campione.

Tuttavia, molta ricerca in psicologia è fatta basandosi su campioni ottenuti attraverso una selezione non causale, come ad esempio nel campionamento dei soggetti disponibili.

È inoltre anche importante il numero dei soggetti coinvolti: un esperimento necessita di tanti soggetti quanti ne sono necessari per fornire dei valori significativi, e più soggetti si coinvolgono più la probabilità cresce. Per sapere quanti soggetti coinvolgere nell'esperimento si possono usare misure di sensibilità che verranno descritte nelle prossime pagine.

Condurre l'esperimento che consiste principalmente nel raccogliere i dati dopo avere stabilito i compiti che i soggetti devono svolgere per testare le ipotesi sperimentali.

In questa fase i soggetti sono assegnati ai differenti trattamenti.

Nella procedura più semplice chiamata ***between-subjects design*** (tra i soggetti) un *gruppo sperimentale* di soggetti è assegnato al trattamento, mentre un altro gruppo di soggetti, chiamato *gruppo di controllo* è assegnato ad una condizione in cui quello specifico trattamento è assente. Per tornare all'esempio di prima, un gruppo di soggetti (gruppo sperimentale) esegue i compiti previsti dall'esperimento utilizzando un'interfaccia web con sfondo pastello, mentre un altro gruppo di soggetti (gruppo di controllo) esegue gli stessi compiti con un'interfaccia avente sfondo bianco. Normalmente questa procedura ha lo scopo di vedere se ci sono differenze tra le due condizioni sperimentali (gli utenti del gruppo sperimentale riescono meglio nello svolgimento dei compiti?).

Se volessimo testare anche le interfacce web con sfondo scuro non faremmo

altro che aggiungere un altro gruppo. Di che tipo? Sperimentale!

All'altro estremo troviamo la procedura conosciuta come *within-subjects design* (**dentro i soggetti**) nella quale ogni soggetto è assegnato a tutti i trattamenti. Riprendendo l'esempio Precedente ogni soggetto dell'esperimento completa i task utilizzando sia l'interfaccia con sfondo pastello che quella con sfondo bianco. Il within-subjects design assicura una maggiore validità esterna, ma può causare effetti indesiderati dovuti al possibile apprendimento del funzionamento dell'interfaccia, effetti che devono essere bilanciati nel disegno dei compiti dell'esperimento.

Utilizzare le statistiche descrittive per descrivere i dati. Questo tipo di misurazioni (come la media, la mediana, la deviazione standard) consentono di descrivere o riassumere una serie di dati. Tuttavia, se si vuole sapere se l'esperimento ha prodotto risultati significativi e se si vogliono fare delle ulteriori inferenze, queste statistiche non sono sufficienti.

Utilizzare le statistiche inferenziali per valutare le ipotesi. Queste metriche sono utili per fare inferenze estendibili anche ai soggetti che non hanno partecipato all'esperimento, ovvero all'intera popolazione di riferimento, e quindi per valutare la **significatività** dell'esperimento che può essere definita come la probabilità (**p**) che lo sperimentatore rifiuterà l'**ipotesi nulla**¹¹ quando invece è corretta.

I livelli accettati di significatività sono 0,05 ($p=5\%$, "significativo") e 0,01 ($p=1\%$, "molto significativo").

Le principali statistiche inferenziali utilizzate per analizzare i dati di un esperimento controllato sono:

- l'**ANOVA** (*the Analysis of Variance*), ovvero l'analisi della varianza che è usata per determinare se le differenze riscontrate nelle variabili dipendenti sono state causate dalle variabili indipendenti o da fattori casuali. A tal proposito viene calcolato un rapporto, chiamato **indice del trattamento** (*treatment index*), tra la variabilità dei due gruppi sperimentali (between-groups variability)¹² e la variabilità all'interno dei gruppi (within-group variability¹³):

¹¹ L'**ipotesi nulla** è quell'ipotesi valutata dall'esperimento. Solitamente, secondo l'ipotesi nulla non vi è una relazione tra il trattamento e la popolazione di interesse.

¹² Nel **between-subjects design** i soggetti sono assegnati casualmente a uno dei due gruppi e la variabilità (**between-groups variability**) è rappresentata dalla differenza tra le medie ottenute dai due gruppi.

¹³ La **within-group variability** prende in considerazione la variabilità tra soggetti che hanno ottenuto lo stesso trattamento.

$$\text{indice del trattamento} = \frac{\text{between - groups variability}}{\text{within - groups variability}}$$

L'indice del trattamento ci dice se le differenze tra i gruppi sono dovute non solo all'errore sperimentale:

$$\text{indice del trattamento} = \frac{(\text{effetti del trattamento}) + (\text{errore sperimentale})}{\text{errore sperimentale}}$$

Se il valore ottenuto è vicino a 1 non sono presenti effetti del trattamento e di conseguenza l'ipotesi nulla è vera; altrimenti, se il valore ottenuto è più grande di 1 possono esserci effetti dovuti al trattamento. Ma questo non basta. Dobbiamo sapere quanto il valore ottenuto deve essere più grande di 1 per poter respingere l'ipotesi nulla e accogliere come vera **l'ipotesi alternativa**¹⁴. L'indice del trattamento è anche conosciuto come *F* (**F ratio**), e può essere calcolato attraverso una procedura statistica chiamata **F Test**.

Se il valore di *F* è uguale o eccede un *valore critico* l'ipotesi nulla è respinta, altrimenti è considerata come vera. Il valore critico di *F* è stabilito da una tabella che considera tutte le possibili combinazioni tra *a* (il numero dei trattamenti, ovvero di variabili indipendenti manipolate nell'esperimento) e *n* (il numero dei partecipanti all'esperimento) in un insieme di soggetti relativamente piccolo [Keppel, 1991; Keppel et al., 1998].

Esiste un altro test che ci consente di analizzare i risultati ottenuti da due diversi gruppi e quindi la differenza tra le due medie: il **t test**¹⁵, che è algebricamente uguale all'*F* test.

Il **Chi square test** (χ^2) invece è quella metrica statistica utilizzata per valutare i valori significativi che possono assumere dei dati categorici, ovvero quei dati che classificano il comportamento dei soggetti in categorie mutuamente esclusive che non possono venire ordinate (mentre l'*F* test e il *t* test possono essere utilizzati quando le variabili indipendenti che devono essere misurate assumono valori continui). Nel test del Chi quadro le frequenze osservate di comportamento sono paragonate con le frequenze

¹⁴ L'**ipotesi alternativa** è quell'ipotesi secondo cui vi è una relazione tra il trattamento e la popolazione di interesse.

¹⁵ Il *t* test è uno speciale caso di *F* test, in quanto i due test sono algebricamente equivalenti: $F = (t)^2$ e $t = \sqrt{F}$.

- attese di comportamento, basate su considerazioni teoriche o empiriche;
- le **misure di sensibilità**. La sensibilità può essere definita come la capacità di rilevare le differenze (tra due gruppi sperimentali) quando sono presenti. La sensibilità di un esperimento è calcolata attraverso l'omega quadro (ω^2 **effect size or treatment magnitude**) e il potere (**power of an experiment**). L'omega quadro misura la forza degli effetti del trattamento in un esperimento, fornendo la misura della grandezza del trattamento nei valori delle variabili dipendenti dovuti ai cambiamenti nella variabile dipendente come percentuale della variabilità totale. Nelle scienze sociali valori uguali a 0,01 / 0,06 / > 0,15 sono considerati rispettivamente come *piccoli, medi e grandi* effetti del trattamento. Il **potere di un esperimento** è l'abilità di riconoscere gli effetti del trattamento. Il potere si può aumentare riducendo la variabilità del trattamento, analizzando i fattori di controllo al posto di renderli casuali, riducendo la variabilità tra i soggetti (utilizzando ad esempio un within subjects design). Il potere può essere usato per stimare la dimensione del campione. Nelle scienze sociali il valore accettato del potere (power) è di 0,80, il che significa che ripetendo l'esperimento l'80% delle volte si otterrebbe lo stesso risultato. Progettare gli esperimenti con un elevato potere assicura non soltanto una grande ripetibilità dei risultati, ma aumenta anche le possibilità di trovare gli effetti desiderati;
- le **statistiche non parametriche**, sono statistiche che non fanno assunzioni circa la distribuzione dei valori e sono da preferirsi quando i dati non hanno una distribuzione uniforme. Esempi di test non parametrici sono: Mann-Whitney test, Kruskal-Wallis test, Wilcoxon test, and Friedman test;
- i **test post-hoc**, che sono misurazioni che vengono effettuate dopo che i dati sono già stati esaminati e sono usate per limitare gli errori. Il test di questo tipo maggiormente usato in psicologia e nelle scienze sociali è il Tukey's Honestly Significant Difference (Tukey's HSD) test.

Fornire le conclusioni sulle ipotesi della ricerca.

Preparare un report formale per pubblicazioni e presentazioni.

In un esperimento ideale dovrebbero variare solamente le variabili indipendenti, ma invece nella realtà ci sono anche altri fattori che variano e che possono influenzare le variabili dipendenti. Questi fattori sono chiamati **variabili di confusione** (*confounding variables* o *nuisance variables*) e possono causare seri problemi quando influenzano senza alcun controllo il comportamento dei soggetti partecipanti all'esperimento fino a rovinare l'esperimento, se non si riesce a distinguere tra gli effetti delle variabili indipendenti e quelli delle variabili di

confusione.

Ad esempio, se in un esperimento i soggetti appartenenti al gruppo di controllo vengono testati sempre di mattina, e quelli del gruppo sperimentale sempre di pomeriggio, queste differenze di orario potrebbero influenzare le performance dei soggetti a causa della stanchezza, delle condizioni ambientali e contestuali diverse, del traffico diverso che ci può essere su Internet la mattina e il pomeriggio, ecc.

Altri potenziali problemi possono essere generati da

- locazioni diverse,
- rumori,
- attrezzatura,
- l'influenza degli sperimentatori,
- ...

Un modo per controllare queste variabili di confusione è quello tenerle costanti, in maniera tale che esse possano avere la stessa influenza sulle diverse condizioni di trattamento.

Sfortunatamente però non tutte le potenziali fonti di confusione possono essere tenute sotto controllo (come ad esempio l'intelligenza dei soggetti, la loro velocità di lettura, ecc). Dunque l'effetto delle variabili di disturbo che non possono essere tenute sotto controllo viene neutralizzato assegnando casualmente i soggetti alle diverse condizioni sperimentali. Casualmente significa che ogni soggetto ha la stessa probabilità di essere assegnato a condizioni sperimentali differenti. Ad esempio non sarebbe corretto assegnare tutte i soggetti femminili ad un gruppo e quelli maschili ad un altro.

Infine è bene notare che le statistiche sopra riportate sono sovente rinforzate da osservazioni di tipo qualitativo al fine di valutare con una maggiore correttezza i risultati ottenuti nell'esperimento.

Nelle pagine che seguono è mostrato un esempio dei compiti di un esperimento controllato effettuato nel 2003 da uno degli autori per valutare due versioni (una personalizzata e l'altra non personalizzata) del sito del monitoraggio delle acque della Direzione delle Risorse Idriche della Regione Piemonte¹⁶. L'ipotesi che si voleva dimostrare era che il sito in versione personalizzati favorisse maggiormente gli utenti nello svolgimento dei loro compiti. A tal fine, un gruppo di soggetti venne fatto intergere con la versione personalizzata del sito, mentre un altro gruppo di soggetti con la versione non personalizzata. I due gruppi svolsero lo stesso set di compiti i cui risultati vennero paragonati per testare la validità dell'ipotesi iniziale.

Descrizioni maggiormente dettagliate dell'esperimento sono reperibili presso le fonti: [Gena and Ardissono, 2004; Gena and Ardissono, 2006].

¹⁶ <http://www.regione.piemonte.it/acqua/index.htm>

Compiti da svolgere sul sito

1. Vai alla ricerca avanzata dei dati qualitativi dei corsi d'acqua superficiali
2. Cerca
 - a. Gli indici di stato ambientale del fiume Po, nei comuni di Sanfront e Trino nell'anno 2002
 - b. Senza scaricare il file, cerca anche i valori dei parametri chimici e microbiologici per lo stesso fiume, comuni e anno
 - c. Senza scaricare il file, cerca anche i valori dei parametri biologici per lo stesso fiume, comuni e anno

Chiedi allo sperimentatore le nuove istruzioni

3. Vai alla ricerca avanzata dei dati quali-quantitativi misurati in continuo dei corsi d'acqua superficiali
4. Cerca
 - a. Nell'annale idrologico della Regione Piemonte i dati di portata del fiume Po, nella stazione di Torino-Murazzi, relativi all'anno 2000
 - b. Senza scaricare i file, vai alla ricerca avanzata nelle acque sotterranee e cerca i dati quantitativi nel comune di Torino, in piazza d'Armi
5. Torna alla ricerca avanzata dei dati quali-quantitativi misurati in continuo dei corsi d'acqua superficiali
6. Cerca
 - a. Nella Consultazione parametri idrometrici e chimico-fisici misurati in continuo i parametri idrometrici del fiume Po, stazione di Torino-Murazzi del giorno 1 gennaio 2000
 - b. Senza scaricare il file, cerca le caratteristiche della stazione di monitoraggio della stazione di Torino-Murazzi



Figura 6.2. Interfaccia del sito testato, reperibile all'indirizzo:
http://acqua.di.unito.it/acqua/m_qualitativo/cercaBdaBioRisultati.asp?frmCorso=PO&frmComune=SANFRONT,%20TRINO&frmAnno=2002

6.3 La valutazione qualitativa

I metodi qualitativi di valutazione non sono frequentemente impiegati nella valutazione dei siti web, per lo meno non quanto le metodologie più quantitative. Tuttavia, questi approcci sono legati al reale utilizzo delle applicazioni, in condizioni e contesti reali. Di conseguenza, essi possono offrire dei risultati significativi sia in termini di valutazione del sistema sia per l'analisi dei requisiti.

La distinzione tra ricerca qualitativa e quantitativa ha origine nelle scienze sociali. Un'immediata e significativa definizione di ricerca qualitativa ci viene da Anselm Strauss, uno dei fondatori della Grounded Theory. Per Strauss [Strauss and Corbin, 1998, pp 10-11]

“con il termine ricerca qualitativa indichiamo qualsiasi tipo di ricerca che produce delle soluzioni che arrivano non da procedure statistiche o altri mezzi quantitativi. Ci si può riferire alla ricerca sulle vite delle persone, alle esperienze reali, ai comportamenti, alle emozioni, ai sentimenti, così come al funzionamento delle organizzazioni, ai movimenti sociali, ai fenomeni culturali, e alle interazioni tra le nazioni. (...). Parlando di analisi qualitativa ci riferiamo non alla quantificazione di dati qualitativi ma piuttosto a processi non matematici di interpretazione, portati avanti con il proposito di scoprire concetti e relazioni nei dati grezzi e di conseguenza organizzarli in schemi teoretici di spiegazione”.

I metodi qualitativi di ricerca sovente utilizzano le investigazioni di tipo etnografico, anche definite come osservazioni partecipate. Nelle scienze sociali, e in particolare nelle ricerche sul campo, l'**osservazione partecipata** è un metodo di ricerca qualitativo che richiede il coinvolgimento diretto del ricercatore con il suo oggetto di studio [Corbetta, 1999]. Durante l'osservazione partecipata l'osservatore si immerge in un nuovo mondo con lo scopo di esplorare il punto di vista dei suoi abitanti.

L'utilizzo di materiale etnografico è abbastanza comune nella valutazione dei sistemi interattivi, soprattutto nel mondo anglosassone e nordeuropeo.

I sostenitori degli studi qualitativi hanno scoperto che gli approcci etnografici possono scoprire dei requisiti per i sistemi attraverso la dettagliata osservazione del modo di lavorare degli utenti e dell'ambiente di lavoro.

Al contrario, gli approcci tradizionali, come gli esperimenti in laboratorio, tendono ad essere scollegati dai dettagli del lavoro dal vivo. Secondo una prospettiva etnografica i metodi quantitativi, come gli esperimenti controllati e le analisi quantitative in genere, sono prive di significato in quanto sono de-

contestualizzate ed esaminate entro gli sterili confini di un laboratorio.

I ricercatori qualitativi sono alla ricerca di un coinvolgimento più diretto e desiderano avere una visione più ampia della relazione tra tecnologia e lavoro attraverso la comprensione di come le caratteristiche dei sistemi software sono parte delle pratiche di lavoro.

All'opposto, i sostenitori dei metodi quantitativi polemizzano sul fatto che anche nel caso degli studi sul campo la situazione non è del tutto naturale, dal momento che i soggetti sono in qualche modo influenzati dalla presenza dei valutatori e/o dagli equipaggiamenti per la registrazione.

Da un punto di vista più analitico, i metodi etnografici possono venire utilizzati per esaminare i processi e le pratiche lavorative [Dourish, 2001]. I **processi lavorativi** possono essere definiti come quelle procedure formalizzate o regolarizzate grazie alle quali il lavoro è portato avanti (procedure per l'autorizzazione dei pagamenti, per la fornitura degli ordini, ...). I processi lavorativi sono codificati nei manuali, nei sistemi informativi, ecc.

Invece, i ricercatori etnografici in HCI hanno spesso prestato attenzione alle **pratiche lavorative**, ovvero quei meccanismi informali di routine attraverso i quali questi processi sono messi in pratica e organizzati per affrontare le esigenze quotidiane. Il modo in cui le persone “deviano” dalle procedure formalizzate tende a riflettere meglio, o per lo meno in maniera più fruttuosa, l'adattamento dei processi alle specifiche circostanze nelle quali l'attività è portata avanti. Questo risulta essere particolarmente rilevante nello sviluppo di quei sistemi informativi nei quali i progettisti presumono che i processi lavorativi costituiscano la perfetta descrizione di quello che realmente accade. In questo caso processi sono codificati nei sistemi informativi senza tenere in conto di quello che realmente accade nella pratica, e questo può portare a degli errori di interfaccia.

Preece et al. [Preece et al., 1994] classificano le investigazioni etnografiche con il termine di **valutazione interpretativa**, che può essere riassunta con il termine “passare il tempo con gli utenti” e si basa sull'assunzione che i piccoli fattori che vanno oltre il comportamento visibile possono influenzare grandemente i risultati. Dal momento che il contesto del laboratorio non è una condizione paragonabile a quella del mondo reale solo l'osservazione degli utenti nel loro reale contesto può individuare la presenza di questi fattori.

Per Walsham [Walsham, 1993] “*l'interpretativismo*” riguarda gli approcci utili alla comprensione della realtà la cui conoscenza è necessaria alla costruzione sociale ed è una conoscenza soggettiva. Una valutazione interpretativa può essere utile in diversi stadi del ciclo di sviluppo, ma in particolare negli studi di fattibilità e nelle revisioni post-implementative.

La valutazione interpretativa si articola empiricamente in:

- **intervista contestuale**, che è un tipo di valutazione qualitativa basata su metodologie etnografiche. Secondo quest'approccio utenti e valutatori lavorano insieme all'interno dell'ambiente di lavoro degli utenti per identificare e comprendere i problemi di usabilità dell'interfaccia sottoposta a valutazione [Preece et al., 1994]. Normalmente la valutazione contestuale è organizzata come un'intervista semi-strutturata orientata a coprire tutti quegli aspetti interessanti dell'interazione utente-sistema nel reale contesto d'uso. Molto spesso l'intervista viene registrata per essere poi successivamente rielaborata insieme dall'intervistatore e dall'intervistato. Per approfondimenti [Beyer, 1998];
- **intervista cooperativa** [Monk et al., 1993], che include metodi dove l'utente è incoraggiato ad agire come un collaboratore nella valutazione al fine di individuare non solo i problemi di usabilità, ma anche la loro soluzione. I protocolli "think aloud" rientrano nei metodi cooperativi in quanto consentono all'utente di fare domande, commenti e suggerire alternative appropriate ai valutatori, e ai valutatori di dare suggerimenti all'utente;
- **valutazione partecipata**, che è più aperta e soggetta ad un maggiore controllo da parte degli utenti dell'intervista cooperativa ed è strettamente legata alle tecniche di progettazione partecipata, nelle quali gli utenti stessi sono coinvolti nelle fasi di progettazione. Nella valutazione partecipata si applicano tecniche come i focus group e la valutazione di prototipi. Per dettagli si veda [Greenbaum and Kyng, 1991];
- **l'etnografia**, che implica la raccolta di dati in reali situazioni di lavoro. Gli approcci etnografici in HCI riconoscono l'importanza di imparare in maniera più approfondita come la tecnologia è usata nella realtà. Questo implica [Monk et al., 1993]
 - la necessità di usare una serie di metodi quali l'osservazione intensiva, l'intervista in profondità, la partecipazione alle attività e agli eventi della realtà sotto osservazione;
 - una prospettiva olistica, nella quale ogni cosa (i sistemi informatici usati, le attività di rito, gli artefatti, i testi) è di vitale importanza ed è fondamentale immergersi nelle situazioni reali.

Svariate sono le fonti dei dati utili all'analisi etnografica e includono la registrazione audio-video, le fotografie, il prendere appunti, ecc

Considerando infine l'analisi della vita di ogni giorno non si può non citare la teoria dell'etnometodologia di Harold Garfinkel [Garfinkel, 1967] definita da [Preece et al., 1994] come un metodo che presuppone l'assenza di un modello a

priori dei processi cognitivi che vengono messi in atto mentre una persona svolge un'attività, e invece si limita ad analizzare il comportamento osservando gli eventi nel loro reale accadere. Gli etnometodologi richiamano l'attenzione sull'analisi di quei metodi di senso comune che le persone utilizzano facendo le azioni di ogni giorno. Le persone richiamano questi metodi come soluzioni pratiche ai problemi pratici per rendere il mondo sensato e comprensibile nel corso delle azioni di ogni giorno.

Per gli etnometodologi le azioni quotidiane sono organizzate da regole precise e implicite che hanno a che fare con l'interazione umana e che per essere scoperte implicano la rottura di queste regole implicite. Da qui nascono i metodi non convenzionali dell'etnometodologia, come ad esempio parlare troppo vicino ad una persona estranea, bere dal bicchiere di un'altra persona ad una festa. Il fatto che queste azioni causino stupore e disorientamento nella gente conferma l'esistenza di tali regole implicite.

L'etnometodologia è stata usata in HCI per studiare l'impatto delle nuove tecnologie in vari ambienti di lavoro con un focus sui dettagli del lavoro e sulle pratiche sociali attraverso le quali i soggetti possono influenzare lo sviluppo della tecnologia, su come la tecnologia viene sviluppata e usata, e su quello che si dice e si scrive a suo riguardo. Per dettagli si veda [Dourish, 2001].

6.3.1 *The Grounded Theory*

La Grounded Theory è una metodologia qualitativa di ricerca sviluppata da Anselm Corbin and Juliet Strass che la definiscono come

“una teoria derivata dai dati, raccolti in maniera sistematica e analizzati durante il processo di ricerca. In questa metodologia la raccolta dei dati, l'analisi e l'eventuale teoria stanno in stretta relazione. Un ricercatore non inizia un progetto con una predeterminata teoria in testa (...). Piuttosto, il ricercatore comincia con un'area di studio e lascia che la teoria emerga dai dati.” [Strauss and Corbin, 1998, p. 12]

Nell'introduzione del loro libro, Strauss and Corbin delineano le caratteristiche di un ricercatore qualitativo, che deve avere:

- capacità di tornare indietro e analizzare criticamente le situazioni,
- capacità di riconoscere i pregiudizi
- capacità di pensare in maniera astratta
- capacità di essere flessibili e aperti alle critiche utili
- sensibilità verso il mondo e le azioni di risponde
- un senso di assorbimento e devozione ai processi lavorativi
- flessibilità e apertura

- capacità di interpretare le situazioni ambigue

Tutte queste caratteristiche tuttavia non servono a nulla se il ricercatore non sviluppa un nuovo modo di rapportarsi ai dati e alle informazioni del mondo in cui vive.

Secondo l'approccio della Grounded Theory i dati sono raccolti utilizzando le stesse tecniche delle altre metodologie di ricerca. I dati possono essere sia quantitativi che qualitativi o una combinazione di entrambi.

Per Strauss and Corbin le principali componenti della Grounded Theory sono:

- i dati: le interviste, le osservazioni, i documenti, i film, le registrazioni;
- le procedure utilizzate per interpretare e organizzare i dati. Normalmente queste implicano la concettualizzazione, riduzione, rielaborazione e costruzione di relazioni tra i dati attraverso domande e paragoni;
- report orali e scritti.

In particolare il secondo punto è ulteriormente sviluppato e poi suddiviso in tre passi importanti che sono le componenti principali della teoria:

- l'*open coding*, ovvero il processo analitico attraverso il quale i concetti vengono identificati e le loro proprietà e dimensioni scoperte;
- l'*axial coding*, ovvero quel processo che implica la costruzione di relazioni tra categorie e sottocategorie di dati, chiamato assiale perché la codifica avviene intorno agli assi delle categorie, collegando le categorie alle loro proprietà e dimensioni;
- il *selective coding*, ovvero quel processo che implica l'integrazione e il raffinamento della teoria.

Come metodologia e serie di metodi, la Grounded Theory può essere applicata non solo nelle scienze sociali, ma anche in campi pratici come la comunicazione, l'architettura, l'antropologia, la psicologia e ovviamente l'informatica e lo studio di interfacce.

6.4 Quantitativi e qualitativi: come scegliere?

La scelta tra metodologie qualitative e quantitative di valutazione dipende dal punto di vista della valutazione: mentre la ricerca quantitativa cerca di spiegare la varianza delle variabili dipendenti causata dalla manipolazione di variabili indipendenti, e quindi viene definita come **basata sulle variabili**, nella ricerca qualitativa l'oggetto dello studio è l'individuo stesso, e quindi viene definita **basata sui casi**.

I ricercatori qualitativi sostengono che un soggetto non può essere considerato

come una somma di variabili e di conseguenza una più profonda conoscenza di un piccolo gruppo di soggetti è più utile di un esperimento controllato con un campione di utenti rappresentativi. Anche se gli obiettivi ultimi di entrambi gli approcci sono simili (entrambi vogliono scoprire teorie predittive in grado di generalizzare il comportamento degli individui), esse sono portate avanti in maniera del tutto diversa: mentre i ricercatori quantitativi cercano di spiegare le relazioni di causa-effetto tra le variabili e fare generalizzazioni sui risultati ottenuti dagli esperimenti (**approccio estensivo**), i ricercatori qualitativi vogliono comprendere i soggetti sotto studio interpretando il loro punto di vista e analizzando i fatti in profondità (**approccio intensivo**) al fine di proporre una nuova comprensione della realtà.

In conclusione, se il nostro scopo è quello di scoprire nuove teorie, un approccio qualitativo può portare a risultati più fruttuosi, mentre se vogliamo investigare le relazioni tra variabili conosciute, è meglio un approccio quantitativo.