

Introduzione

di Paolo Paolini

Ma in definitiva, perché occuparsi di usabilità? È davvero così importante?

La mia esperienza, professionale più ancora che scientifica, mi dice che oggi l'usabilità è uno dei fondamenti culturali attorno a cui basare la progettazione. Se nell'epoca delle ".com" il mandato era "realizzate il sito in tre mesi", oggi la domanda è "il sito è efficace?", oppure "i visitatori riescono a usarlo nel modo che si voleva?".

L'usabilità, cioè la possibilità concreta di utilizzare al meglio un'applicazione interattiva, è l'elemento base per capire se i costi di sviluppo sono ricompensati da una qualità adeguata della fruizione e della *user experience*.

Il mio rapporto personale con l'usabilità è iniziato intorno al 1995: Francesca Costabile (attualmente docente ordinario presso l'Università di Bari) mi invitò a occuparmi, all'interno di un progetto per IBM Italia (per il Consorzio Corinto) dei problemi d'usabilità di strumenti per lo sviluppo di programmi Java. All'epoca ero docente alla Facoltà di Ingegneria di Lecce, e avevamo appena creato un Laboratorio (che poi si sarebbe chiamato SET-Lab) che ancora stava cercando una precisa identità culturale. Con Mario Bochicchio, Roberto Paiano e con un gruppo di lavoro del Politecnico di Milano (Franca Garzotto e Maristella Matera) cominciammo ad addentrarci in queste problematiche per noi nuove.

* Docente presso il Politecnico di Milano e l'Università della Svizzera italiana (Lugano).

La nostra prima impressione fu che il settore – e quindi l’insieme di metodologie per la valutazione dell’usabilità – fosse relativamente poco “ingegnerizzato” e insoddisfacente per una serie di motivi.

- I metodi proposti sembravano essere poco sistematici e poco adatti a un lavoro in squadra: nostro obiettivo era invece quello di formulare un metodo che potesse assicurare a una realtà industriale ampia e variegata (come IBM) una certa sistematicità. In particolare, ci si poneva l’obiettivo di rispondere a due semplici domande: come si può suddividere il lavoro di valutazione tra più *team* di valutazione? Ci si può aspettare una ragionevole “ripetibilità” dei risultati, nel senso che affidando a due squadre diverse la stessa valutazione si ottengono valutazioni simili (o perlomeno facilmente confrontabili)?
- Quasi tutte le metodologie sembravano porre una fiducia quasi assoluta nel fatto che mettendo un potenziale utente di fronte a uno strumento, e osservando il suo comportamento, si potessero ottenere valutazioni ottimali.

La nostra esperienza di utenti di strumenti complessi (quali quelli per lo sviluppo di *software*, per esempio) ci diceva invece come l’utente finale fosse in grado di capire direttamente la qualità (e anche l’usabilità) di applicazioni complesse solo dopo un utilizzo prolungato, di fronte a problemi reali in situazioni reali. La capacità di un generico utente di evidenziare i problemi d’usabilità (di uno strumento complesso) dopo poche ore di utilizzo in un laboratorio, ci sembrava dubbia. Non che discostessimo l’importanza delle opinioni dell’utente finale: avevamo dubbi che queste si potessero ottenere in breve tempo, in un laboratorio, in un uso simulato e non reale.

- I metodi di valutazione esistenti (soprattutto, come detto sopra, basati sull’osservazione delle reazioni di un potenziale utente) sembravano spesso incapaci di discriminare con chiarezza i sintomi dei problemi dalle loro cause, oppure i vari sintomi tra di loro.

La nostra esperienza – di realizzatori di applicazioni e di utenti di sistemi complessi – ci diceva che di fronte a uno stesso problema (per esempio: “l’utente non trova il bottone su cui cliccare”) spesso le cause possono essere ben diverse; per esempio: “è sbagliata l’icona associata al bottone”, oppure: “è sbagliata la collocazione del bottone all’interno della pagina”, oppure: “la scelta dei colori è errata”, oppure: “l’intera concezione della pagina è errata”, oppure: “l’impianto interattivo-navigazionale è sbagliato”, oppure, e peggio: “la difficoltà è

dovuta a una confusione cognitiva che si è generata alla pagina precedente”, e così via...

- Infine, il nostro dubbio principale riguardava la genericità del *feedback* dato ai progettisti. Data la nostra esperienza decennale, e in qualche caso pluridecennale, nel progettare, e quindi realizzare, applicazioni complesse – sia tradizionali che multimediali e online –, ci sembrava inadeguato che i progettisti ricevessero valutazioni del genere: “l’utente si perde”, “l’utente non sa come navigare”, “l’interfaccia è intimidente”, e così via. Come progettisti sapevamo quale risposta avremmo voluto avere da un’attività di valutazione: *che cosa non funziona, e per quale motivo non funziona!* Non che volessimo – a fronte di possibili errori – una controproposta di progettazione (compito che spetta ai progettisti e non ai valutatori): ci aspettavamo però che i progettisti fossero messi in grado di capire su quali aspetti lavorare (vedi punto precedente) sapendo, con ragionevole precisione, quali fossero i problemi da risolvere.

Con i requisiti di cui sopra il gruppo di lavoro (su commissione del Consorzio Corinto) mise a punto una prima proposta di metodo (poi chiamata SUE) i cui punti qualificanti potevano essere così riassunti.

- L’“ispezione” dell’applicazione da parte di valutatori esperti di applicazioni (e specificamente addestrati alla valutazione), era considerata imprescindibile: o come unica valutazione, o come guida per successive valutazioni con utenti finali. L’ottimizzazione del rapporto benefici/costi (l’ispezione è di gran lunga meno costosa che la prova con gli utenti), la difficoltà nel valutare a fondo un’applicazione complessa e la necessità di un *feedback* professionale, ci avevano convinti di questa scelta.
- I diversi piani di valutazione (per esempio interfaccia, contenuto, navigazione, interazione ecc.) dovevano essere tenuti rigorosamente distinti, e valutati separatamente. La nostra esperienza ci diceva anche che l’eccellenza in un piano di valutazione (per esempio: contenuti ottimi) non implicava minimamente la qualità su altri piani (per esempio: navigabilità scadente). La nostra consapevolezza della ripartizione del lavoro (progettuale e realizzativo) tra ruoli diversi, rendeva necessario ripartire più rigorosamente la valutazione nelle sue componenti.
- Per ciascun piano di valutazione bisognava individuare delle linee guida operative per i valutatori. Queste linee guida (ini-

ziammo da quelle per la navigazione) furono chiamate (probabilmente con scelta infelice) *abstract task*. In sostanza fornivamo al valutatore una lista precisa di operazioni da fare, e dei criteri per valutare, a fronte di ciascuna operazione, l'efficienza dell'applicazione.

Questa scelta aveva un triplice scopo: omogeneizzare meglio le valutazioni dei vari valutatori (che avrebbero svolto operazioni simili nel valutare), rendere più efficace la valutazione (fornendo istruzioni dettagliate a valutatori che potevano essere anche non particolarmente esperti) e rendere più efficiente la valutazione stessa (riducendo tempi e costi).

Dopo la prima esperienza con SUE l'attività di ricerca proseguì per qualche tempo presso il Politecnico di Milano, al laboratorio HOC – Hypermedia Open Center, del Dipartimento di Elettronica e Informazione, e quindi generò anche alcune esperienze a Lugano (Facoltà di Scienze della Comunicazione dell'Università della Svizzera italiana), dove era sorto TEC-Lab, il Technology Enhanced Communication Lab, specializzatosi nella ricerca a cavallo tra tecnologie e discipline connesse, in senso lato, con le scienze della comunicazione. In particolare una tesi (di Vittorio Del Percio) fu l'occasione per una serie di passi in avanti significativi.

Gli attributi (genericamente trattati in SUE) per esprimere i risultati di valutazione furono messi meglio a fuoco. In altre parole ci si rese conto del fatto che era necessario “standardizzare” il vocabolario utilizzato per specificare che cosa era andato bene e che cosa era andato male nell'ispezione.

Il tentativo di eseguire una certa operazione può portare (e spesso porta), infatti, a valutazioni tra di loro contraddittorie. Individuare, per esempio, quali siano le mostre temporanee disponibili in un certo museo in una certa data, potrebbe essere reso possibile con operazioni potenti (che con pochi *click* producono il risultato), ma difficili da capire e/o da memorizzare. Oppure, alternativamente, potrebbero richiedere molti *click*, ma in una sequenza facile da capire e ricordare.

Fu quindi evidente come una valutazione approfondita richiedesse, necessariamente, di dare un “voto” a ogni singolo attributo (per esempio: “facilità”, oppure “potenza”, oppure “auto-evidenza”, e così via) e che, all'interno di un gruppo di valutatori, la lista degli attributi da usare (e i criteri per valutarli) dovesse essere omogenea.

Il trattamento numerico delle valutazioni fu definito in dettaglio per la prima volta. L'obiettivo era duplice: fornire da un lato

ai manager dati sintetici di valutazione molto “comunicativi”; fornire, d’altro lato, ai progettisti valutazioni (esprese in forma numerica) dettagliate sulle singole operazioni e sui singoli attributi.

Inoltre verificammo, su casi concreti, la difficoltà nel “sommare” tra di loro i vari risultati. I “voti” assegnati ai vari attributi, per una stessa operazione, non potevano essere semplicemente tradotti in una media: per utenti casuali e poco esperti, per esempio, la facilità d’uso è molto più importante della “potenza”; l’opposto succede per utenti abituali ed esperti. Oppure, avendo assegnato i “voti” a un gruppo d’operazioni, non è possibile semplicemente farne la media. Alcune operazioni potrebbero essere, per lo meno per alcuni utenti, molto più importanti di altre, e di questo bisogna tener conto “pesando” i vari risultati.

Fu inoltre meglio definito il confine tra la valutazione di tipo generale, che si può effettuare senza conoscere gli scopi e le caratteristiche dettagliate dell’applicazione, dalla valutazione specializzata, basata cioè sulla conoscenza di dettaglio di chi sono coloro che fruiranno dell’applicazione.

La grafica, o alcuni aspetti della “navigabilità”, per esempio, possono essere valutati anche senza sapere a che cosa serva l’applicazione. Per altri aspetti – per esempio la qualità dei contenuti oppure la qualità dell’organizzazione dell’informazione – è necessario conoscere gli scopi dell’applicazione, il suo pubblico, il modo in cui verrà usata, e così via.

Da quest’attività di ricerca scaturì una serie d’iniziative congiunte (da parte del Politecnico di Milano e dell’Università della Svizzera italiana): diversi esperimenti di valutazione di applicazioni multimediali e web e diverse messe a punto del metodo. Un’esperienza particolarmente significativa è stata quella realizzata per conto della Commissione Europea (V Programma Quadro di ricerca) nell’ambito del progetto VNET5: una dozzina di progetti di ricerca, condotti in vari paesi europei, furono aiutati nel valutare l’usabilità dei loro strumenti e delle loro applicazioni. In altri progetti Europei (per esempio: UWA o OpenDrama) fu affinata la capacità, riconosciuta anche dai *partner* internazionali, di valutare l’usabilità in modo sistematico ed efficace.

Tutte queste varie attività consentirono di perfezionare il metodo – cui fu dato il nome di MiLE, dai nomi di Milano e Lugano – con alcuni miglioramenti particolarmente significativi:

- fu identificata la necessità di tener conto, in modo puntuale e dettagliato, dei requisiti per valutare la qualità/usabilità di aspetti “verticali” delle applicazioni, cioè di aspetti legati strettamente alle loro finalità;
- fu introdotto l’uso di scenari applicativi (ipotizzando un tipo d’utente che debba soddisfare una specifica necessità in una specifica situazione), scelti tra i più caratterizzanti, per valutare in modo più approfondito gli aspetti specialistici di un’applicazione.

Quanto sopra è stato possibile anche per lo sviluppo di una moderna metodologia, AWARE – Analysis of Web Application Requirements, anch’essa frutto della collaborazione tra Università della Svizzera italiana e Politecnico di Milano, per l’analisi e l’ingegnerizzazione dei requisiti.

Questo libro raccoglie la maggior parte dei risultati dell’attività pluriennale di ricerca e sperimentazione sopradescritta: lo scopo è quello di divulgare le tematiche, le problematiche e le metodologie per valutare la usabilità.

È necessario, infatti, che dai laboratori di ricerca escano, per il mondo professionale e delle aziende, indicazioni chiare e non ambigue, che migliorino la consapevolezza dei problemi e delle loro possibili soluzioni.

Migliorare la prassi quotidiana di professionisti, ricercatori e appassionati: questo in sintesi è l’obiettivo del libro.

Paolo Paolini