

4. La evaluación en el modelo de proceso

4.1. Plan de evaluación

En un proceso de DCU, la evaluación es una parte esencial que debe comenzar tan pronto como sea posible. Cuanto más tardemos en realizar tareas de comprobación, más difíciles de implementar serán los cambios, errores y/o requisitos encontrados.

El MPIu+a⁷ no determina cuántas evaluaciones deben realizarse durante la realización de un proyecto, ni tampoco qué métodos deben utilizarse o cuántos usuarios o evaluadores tienen que intervenir. Lo que se plantea son unas ideas generales plasmadas en el plan de evaluación:

1) Antes de empezar una evaluación, debemos tener muy claros los objetivos de la evaluación. No debemos pensar en el sistema global, sino sólo en la parte que va a ser probada.

2) Cada evaluación debe tener un responsable (un integrante del equipo de desarrollo), que tendrá las responsabilidades siguientes:

- Moderar las sesiones.
- Reclutar a los usuarios y/o los evaluadores.
- Reservar el lugar en el que se producirá la evaluación (tanto si se trata del laboratorio como del puesto de trabajo).
- Dirigir las tareas que se deben evaluar (según los objetivos del punto anterior).
- Disponer de todos los recursos necesarios para la evaluación (prototipos, dispositivos necesarios, etc.)
- Tener claras las tareas que se deben probar.
- Tener claras qué técnicas de evaluación se utilizarán.

3) La evaluación se debe encajar tanto en lo que respecta a planificación como económicamente en el proyecto.

7. Este plan de evaluación está basado en el propuesto en la ISO 13407 (pág. 10).

4) Documentar los resultados de la evaluación (responsabilidad que también recae en el responsable) y distribuirlos entre los miembros del equipo de desarrollo que precisen de sus resultados (responsabilidad del jefe del proyecto).

4.2. Métodos de evaluación de la usabilidad

Siguiendo la clasificación por el tipo de evaluación vista anteriormente, veremos en este apartado los métodos más destacados de cada una de las tres categorías.

4.2.1. Inspección

El término *inspección* aplicado a la usabilidad aglutina un conjunto de métodos para evaluar la usabilidad en el que unos expertos conocidos como *evaluadores* analizan el grado de usabilidad de un sistema a partir de la inspección o examen detallado de su interfaz.

Hay distintos métodos que se enmarcan en la clasificación de evaluación por inspección. A continuación mostramos los más importantes: el heurístico, el recorrido cognitivo, el recorrido de la usabilidad plural y la inspección de estándares.

1) Heurística

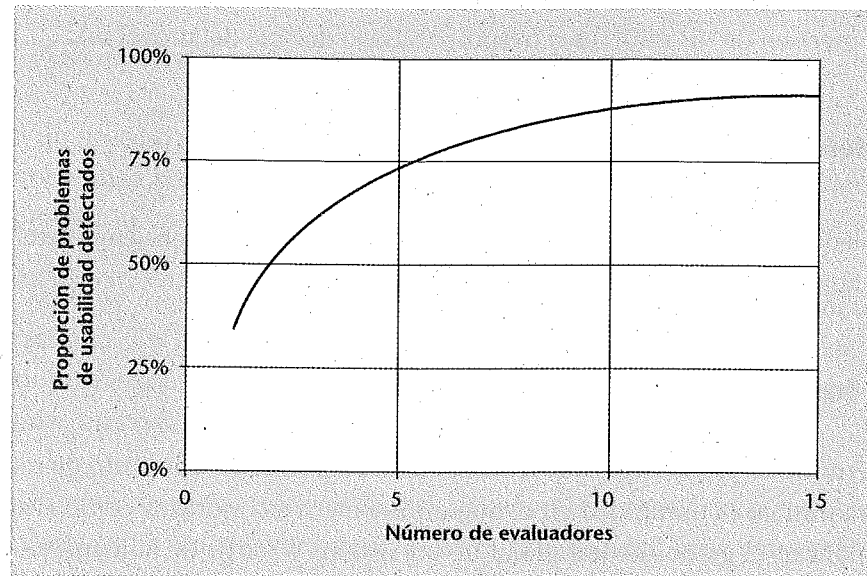
Método desarrollado por Molich y Nielsen (Molich y Nielsen, 1990) que consiste en analizar la conformidad de la interfaz con unos principios reconocidos de usabilidad (heurísticos) mediante la inspección de varios evaluadores expertos.

Puesto que es difícil que el desarrollador o un evaluador puedan encontrar todos los problemas de usabilidad en una interfaz, se utilizan a expertos para validar la interfaz a partir de unos criterios definidos.

En definitiva, en cualquier evaluación heurística es necesario implicar a varios evaluadores. Se recomienda normalmente utilizar de tres a cinco evaluadores, ya que se consideran suficientes y la inclusión de un mayor número de los mismos no garantiza una mejora en el resultado.

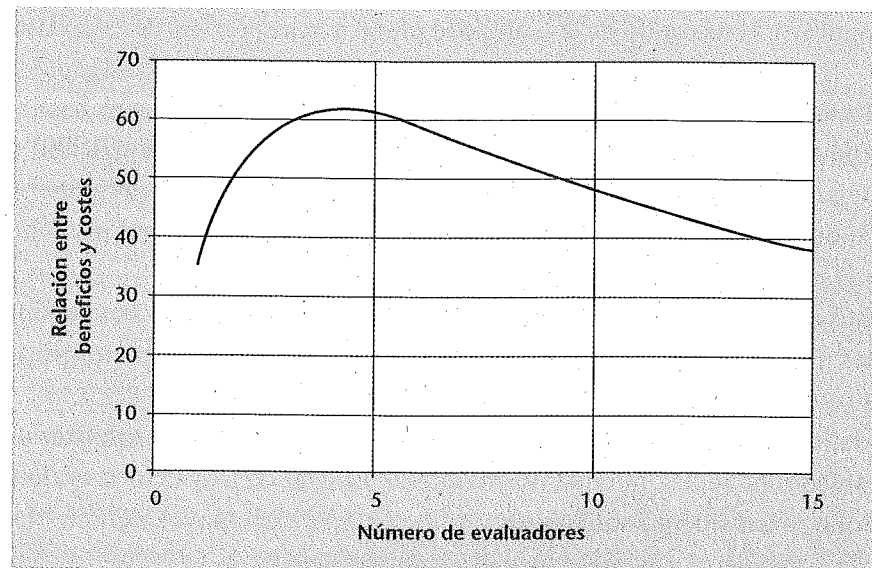
Este número ideal de evaluadores se obtiene de un estudio realizado por Nielsen y Landauer (1993), que queda resumido en las gráficas de las figuras 5.1 y 5.2.

Figura 5.1. Problemas de usabilidad detectados en una interfaz en relación con el número de evaluadores



La curva representa la media de seis estudios de caso de evaluación heurística.

Figura 5.2. Número de evaluadores en relación en el beneficio aportado a un proyecto



La curva de la figura 5.2 muestra que el número óptimo de evaluadores para este ejemplo concreto son cuatro. Esto va en la línea de la tesis que indica que el número ideal de evaluadores es de tres a cinco. En el ejemplo, una evaluación heurística con cuatro evaluadores costaría 6.400 \$ y encontraría problemas de utilidad para obtener 395.000 \$.

Procedimiento

Cada evaluador realiza individualmente una revisión de la interfaz. Una vez concluidas las evaluaciones individuales, los evaluadores comunican y sintetizan sus resultados. Este procedimiento es importante para asegurar evaluaciones independientes e imparciales de cada evaluador. Los resultados de la evaluación se pueden registrar como informes escritos de cada evaluador o haciendo que los evaluadores comuniquen verbalmente sus comentarios a un observador⁸ mientras inspeccionan la interfaz.

Los informes escritos tienen la ventaja de presentar un expediente formal de la evaluación, pero requieren un esfuerzo adicional para los evaluadores y la necesidad de leerlo y diseñar un documento que integre el trabajo de todos los evaluadores por parte del encargado de la evaluación.

Esto supone, además, que los resultados de la evaluación están disponibles inmediatamente después de haber realizado la evaluación, puesto que un único observador necesita entender y ordenar un conjunto de notas personales, no el conjunto de los informes escritos por otros. Además, el observador puede asistir a los evaluadores en el funcionamiento de la interfaz en caso de que se encuentren problemas tales como un prototipo inestable, y puede ayudar a los evaluadores si tienen poco conocimiento del entorno en el que están trabajando y necesitan que se les explique determinados aspectos de la interfaz.

La aplicación del método se centra en la validación de las "diez reglas heurísticas de usabilidad de Nielsen" –conjunto revisado de reglas heurísticas de usabilidad a partir del análisis de 249 problemas de usabilidad (Nielsen y Mack, 1994).⁹

Diez reglas heurísticas de usabilidad de Nielsen

El conjunto revisado de reglas heurísticas de usabilidad de Jacob Nielsen es el que muestra la tabla 5.2.

8. Recurrir a un observador agrega un coste en los gastos indirectos de cada sesión de la evaluación, pero reduce la carga de trabajo de los evaluadores y ofrece un punto de vista subjetivo a las conclusiones.

9. Las reglas propuestas por Nielsen son las comentadas porque son las más utilizadas. Existen, sin embargo, otros conjuntos de reglas o incluso podríamos crear uno particular cuya evaluación se realizaría de manera similar.

Tabla 5.2. Las diez reglas heurísticas de J. Nielsen

1.	Visibilidad del estado del sistema	El sistema debe mantener siempre a los usuarios informados del estado del sistema, con una realimentación apropiada y en un tiempo razonable.
2.	Utilización del lenguaje de los usuarios	El sistema tiene que hablar el lenguaje de los usuarios, con palabras, frases y conceptos familiares, en lugar de que los términos estén orientados al sistema. Se pretende utilizar convenciones del mundo real haciendo que la información aparezca en un orden natural y lógico.
3.	Control y libertad para el usuario	Los usuarios eligen a veces funciones del sistema por error, y necesitan con frecuencia una salida de emergencia claramente marcada, es decir, salir del estado indeseado sin tener que pasar por un diálogo extendido. Es importante que el usuario pueda deshacer y rehacer cuando lo necesite.
4.	Consistencia y estándares	Los usuarios no deben preguntarse si las distintas palabras, situaciones o acciones significan la misma cosa. En general es útil seguir las normas y convenciones del entorno sobre el que se está implementando el sistema.
5.	Prevención de errores	Mejor que generar buenos mensajes de error, resulta más efectivo e importante prevenir su aparición.
6.	Minimizar la carga de memoria	El usuario no debería tener que recordar la información de una parte de diálogo a la otra. Es mejor mantener objetos, acciones y las opciones visibles que memorizar.
7.	Flexibilidad y eficiencia de uso	Las instrucciones para el uso del sistema deben ser visibles o fácilmente accesibles siempre que se necesiten. Los aceleradores no vistos por el usuario principiante mejoran la interacción para el usuario experto, de tal manera que el sistema puede servir para usuarios inexpertos y experimentados. Es importante que el sistema permita personalizar acciones frecuentes.
8.	Diálogos estéticos y diseño minimalista	Los diálogos deben contener la información realmente necesaria (eliminaremos la retórica y aquello que sólo se necesite ocasionalmente). Cada unidad adicional de la información en un diálogo compite con las unidades relevantes de la información y disminuye su visibilidad relativa.
9.	Ayuda a los usuarios a reconocer, diagnosticar y recuperarse de los errores	Los mensajes de error se deben expresar en un lenguaje claro (no con códigos extraños), se tiene que indicar exactamente el problema y deben ser constructivos.
10.	Ayuda y documentación	Aunque es mejor si el sistema se puede usar sin documentación, puede ser necesario disponer de ayuda y documentación. Ésta debe ser fácil de buscar, centrada en las tareas del usuario, se debe tener información de las etapas que hay que realizar y no tiene que ser muy extensa.

Fuente: Nielsen, 1994

Una sesión de evaluación heurística acostumbra a durar de una a dos horas. En caso de que se realice el test de una interfaz muy compleja, se puede dividir en varias sesiones que aborden diferentes aspectos de la interfaz.

El resultado de una evaluación heurística es una lista de problemas de usabilidad que el evaluador ha encontrado en el diseño de la interfaz del sistema evaluado.

2) Recorrido cognitivo (*cognitive walkthrough*)

Este método de inspección de la usabilidad se centra en evaluar la facilidad de aprendizaje de un prototipo de un sistema, básicamente por exploración, y esto está motivado por la observación de que muchos usuarios prefieren aprender software a partir de explorar sus posibilidades (Wharton y otros, 1994).

En el recorrido cognitivo, los revisores evalúan una propuesta de interfaz en el contexto de una o más de sus tareas.

El material que se considera ideal para realizar esta evaluación es un diseño más o menos detallado de la interfaz (bien puede ser en forma de prototipo de papel o en un prototipo de software), un escenario, el análisis de las tareas que hay que realizar, un estudio de la población de usuarios y el contexto de uso.

1) Ámbito y limitaciones del método. El recorrido cognitivo, al ser un método basado en la exploración, permite la adquisición de hábitos que pueden ocultar deficiencias en cuanto a usabilidad, puesto que el usuario consigue sus objetivos aunque sea de la manera más ineficiente posible.

A pesar de que esta técnica es idónea en la etapa de diseño, puede también ser aplicada durante la implementación y en el lanzamiento.

2) Pasos que hay que seguir. Para cada tarea dispondremos, mediante los documentos del análisis de tareas, de una secuencia de acciones que el usuario tiene que realizar satisfactoriamente para completar la tarea designada.

En cada sesión de evaluación, el evaluador explicará previamente a los usuarios: a) el tipo de interacciones que ofrece típicamente la interfaz, b) qué es lo que va a intentar realizar y c) qué acciones están disponibles.

Si el diseño de la interfaz es bueno, las intenciones del usuario provocarán que se seleccione la acción apropiada, a la vez que la interfaz debe presentar una realimentación que indique cómo están progresando para completar la tarea.

Así pues, hay que seguir los pasos siguientes:

a) **Definición de los datos iniciales del recorrido.** Antes de empezar el análisis del recorrido, se debe estar de acuerdo en estos cuatro aspectos:

- *¿Quiénes serán los usuarios del sistema?* La descripción de los usuarios debe incluir la experiencia específica acumulada o el conocimiento técnico, ya que son factores que pueden influir cuando intentan ocuparse de la nueva interfaz.¹⁰
- *¿Qué tarea o tareas serán analizadas?* En general, el análisis se debe limitar a una colección razonable, pero representativa, de tareas de prueba. La selección de las tareas se basará en los resultados del análisis de requisitos. Estas tareas que se deben probar:
 - Serán tan concretas y realistas como sea posible.
 - Incluirán el contexto necesario (al igual que el contenido de las bases de datos que los usuarios se espera que usen).
 - Reflejarán las condiciones típicas bajo las cuales se desarrollará la tarea seleccionada.
- *¿Cuál es la secuencia correcta de acciones para cada tarea?* Para cada tarea, debe haber una descripción de cómo se espera que el usuario vea la tarea antes de aprender la interfaz. También tiene que haber una descripción de la secuencia de las acciones que permiten realizar la tarea con la definición actual de la interfaz.

Las acciones de ejemplo pueden ser del tipo "pulse la tecla Return" o "ponga el cursor sobre el menú Fichero", como también puede ser una secuencia de distintas acciones simples que un usuario típico puede ejecutar como "seleccione Guardar del menú Fichero". El nivel de granularidad de las acciones dependerá de la experiencia de los usuarios.

- *¿Cómo se define la interfaz?* La interfaz debe presentar todos los elementos interactivos disponibles para la consecución de cada tarea, así como su re-

10. Esta información se encuentra en el subapartado 2.3 del capítulo IV.

acción a cada acción disponible (si se trata de un rediseño, seguramente ya dispondremos de esta información).

b) **Recorrido de las tareas.** Mientras el usuario está realizando el recorrido, el evaluador se planteará las preguntas siguientes:

- *¿Pueden alcanzar los usuarios el objetivo correctamente?* Por ejemplo, si la tarea es imprimir un documento, y lo primero que deben hacer es seleccionar una impresora, ¿se darán cuenta de que tienen que seleccionar la impresora?
- *¿Se dan cuenta de que está disponible la acción correcta?* Esto se relaciona con la visibilidad y la comprensibilidad de las acciones en la interfaz.
- *¿Asocian la acción correcta con el efecto alcanzado?* Los usuarios utilizan con frecuencia la estrategia de seguimiento de etiqueta, que les conduce a seleccionar una acción si el texto de la etiqueta para esta acción corresponde con la descripción de la tarea.
- *¿Ven que se está progresando hacia el objetivo de la tarea?* Este aspecto está directamente relacionado con la realimentación del sistema después de que el usuario haya realizado la acción.

El evaluador construirá un escenario que facilite la progresión de las tareas, y cuando se fracase en un determinado paso se relacionará el incidente con una o más de las cuatro preguntas anteriores para determinar la razón por la que un usuario puede fallar.

Los usuarios pueden saber:

- *Qué efecto han alcanzado*
 - Porque es parte de su tarea original.
 - Porque tienen experiencia en el uso del sistema.
 - Porque el sistema les dice cómo deben hacerlo.
- *Que una acción está disponible*
 - Por experiencia.
 - Viendo algún dispositivo (como un botón).
 - Viendo una representación de una acción.

- *Que una acción es apropiada* para el efecto que están intentando alcanzar
 - Por experiencia.
 - Porque la interfaz proporciona una guía o una escritura de la etiqueta que conecta la acción con lo que él está intentando hacer.
 - Porque todas las otras acciones parecen falsas.
- *Que ha habido éxito* después de una acción
 - Por experiencia.
 - Reconociendo una conexión entre una respuesta de sistema y la acción que intentaba realizar.

c) **Recogida de información.** Es importante recoger información mientras se vaya produciendo, durante toda la evaluación, para analizarla después.¹¹

En lo que respecta a material escrito, disponemos de una información inicial que debemos completar con datos referentes a los éxitos y fracasos de los usuarios en la realización de las acciones necesarias para completar las tareas.

Figura 5.3. Ejemplo que muestra las primeras páginas de un documento preparado para hacer la evaluación mediante el recorrido cognitivo correspondiente a un proyecto web preparado para un congreso científico.

Actividad:	EVALUACIÓN de un prototipo de software de la web del Congreso
Tipo de evaluación:	Recorrido cognitivo (<i>cognitive walkthrough</i>)
Definición de los datos necesarios:	
¿Quiénes serán nuestros usuarios?	
a) Los usuarios del sitio web responderán mayoritariamente a los perfiles siguientes:	
	<ul style="list-style-type: none"> – Estudiante universitario – Profesor (universitario) – Profesional de la industria – Doctorando (estudiante de doctorado) – Investigador
b) Conocimientos alcanzados relacionados con el sistema	
	<ul style="list-style-type: none"> • Estudiante universitario: <ul style="list-style-type: none"> – Lo más probable es que no conozca la mecánica del funcionamiento de un congreso. – No sabe qué son las ponencias y los pósters.

11. Resulta de gran utilidad grabar en vídeo toda la sesión. Nos va a permitir revisar posteriormente conversaciones, acciones, etc. y no perder ni un solo detalle.

- Como mucho, sabrá que se trata de un acontecimiento en el que se presentan trabajos avanzados.
- Puede estar interesado en algunos de los temas de los tutoriales.
- No sabe que se entregará un libro de actas con las ponencias.
- El profesor (universitario), el doctorando y/o investigador:
 - Sabe lo que son las presentaciones de ponencias.
 - Sabe lo que son los tutoriales.
 - Debería saber que se trata de un seminario de doctorado (pero, al ser una actividad no habitual –la mayoría de los congresos no tienen esta sección–, hay que insistir en su misión).
- El profesional de la industria:
 - Como norma general no conocerá la mecánica de funcionamiento de un congreso.
 - No sabe qué son las ponencias y los pósters.

c) Experiencia

- Estudiante universitario:
 - Seguramente nunca habrá asistido a un acontecimiento de este tipo.
 - Desconocerá que hay que inscribirse (pero puede intuirlo).
 - Sabe que si decide ir, debe buscar alojamiento.
- El profesor (universitario) y el doctorando y/o investigador:
 - Sabe que hay unas normativas que se caracterizan por el tipo de ponencia, el formato y las fechas de entrega y de revisión.
 - Sabe que para asistir a las conferencias hay que inscribirse.
 - Sabe que deberá buscar alojamiento si quiere asistir.
 - Sabe que se conseguirá libremente un libro de actas con las ponencias.
- El profesional de la industria:
 - Busca cosas muy concretas: la información debe ser muy clara y exacta, con objetivos tangibles.
 - Seguramente tampoco sabe que hay que inscribirse, etc. Está acostumbrado a que le digan un día, una hora y el tema que se tratará.

d) Descripción del prototipo

El prototipo para la evaluación es un prototipo de software del que se adjuntan dos imágenes:

- la primera es la página inicial y será la que verá todo el mundo en el momento de entrar en el sitio web del congreso.

Como elementos significativos de esta página inicial, tenemos el logotipo especialmente diseñado con motivo del acontecimiento en la esquina superior izquierda, cuyo estilo marca el estilo global del sitio web.

El logotipo está situado en una franja negra en el lado superior, la cual incluye, también, en esta página de inicio, el nombre y la edición del congreso, así como las fechas y el lugar donde se hace.

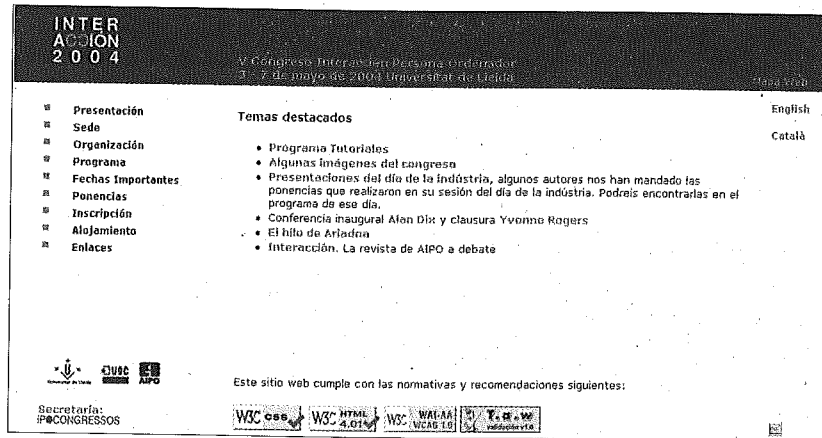
- El resto de la página es de color blanco, razón por la cual el sitio no se ve recargado y, en cambio, se refuerza la franja negra y los contenidos.

Debajo de esta franja negra encontramos:

- En el lado izquierdo, el menú de navegación y la información de los organizadores del acontecimiento (en forma de logotipo).

– Y en el centro, la información relevante de primer nivel. En el momento de esta evaluación, la información importante es la próxima fecha límite de presentación de ponencias y las novedades del congreso.

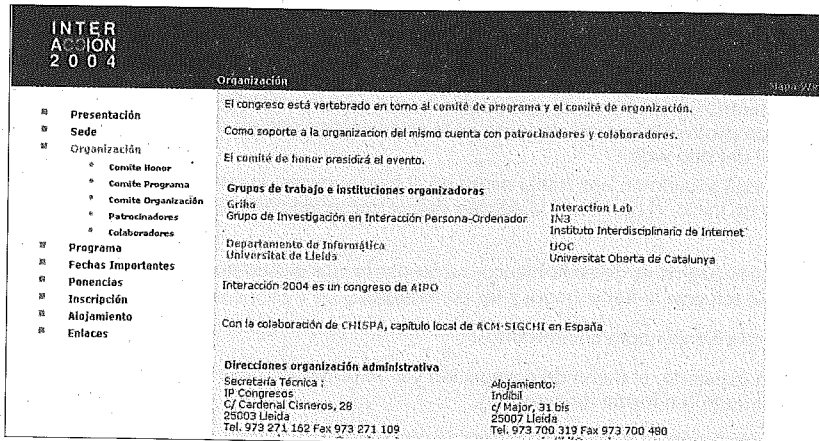
También podemos destacar los distintivos que aparecen en la parte inferior de esta sección, ya que el sitio web cumple las normativas que marcan los estándares de la W3C.



• Y la segunda corresponde a una "pantalla tipo" del resto de las páginas del sitio web. El lugar continúa presidido por la franja negra con el logotipo del congreso en su parte izquierda, pero ahora, en lugar del título del congreso, encontramos la información de navegación en forma de "migas de pan" (*breadcrumbs*), que facilita la ubicación del visitante dentro de la estructura del sitio web.

Debajo de la franja seguimos con el menú de navegación en el lado izquierdo y en el centro veremos la información relacionada con la opción en curso. En estas secciones, se refuerza la estructura de contenido con la ayuda de una franja vertical de color rosa-salmón que enmarca el contenido seleccionado reforzando la estructura.

Todos los enlaces están marcados en el mismo color más saturado y se subrayan al pasar el puntero por encima, lo cual refuerza el significado del enlace.



e) Tareas para hacer

• Tarea 1a: enviar una ponencia

Esto lo podrá hacer cualquier persona que quiera escribir un artículo relacionado con la temática.

– Acciones que se pueden hacer con el prototipo mencionado:

1. Informarse sobre los temas de las ponencias.
2. Si los temas parecen adecuados y se decide escribir y enviar un artículo (como es este caso).
 - a. Informarse de qué hay que hacer para enviar un artículo,
 - b. determinar el tipo de participación,
 - c. el formato que se utilizará y
 - d. la fecha límite.
3. Escribir el artículo.
4. Inscribirse.
5. Enviar el artículo.
6. Comprobar que se ha enviado correctamente.

• Tarea 1b: modificar una ponencia

Podrá hacer esta tarea cualquier persona que ya haya enviado alguna ponencia.

– Acciones que puede hacer el prototipo mencionado:

1. Acceder a la gestión de ponencias.
2. Entrar al inicio de sesión e introducir la contraseña.
3. Buscar la ponencia que se debe modificar.
4. Buscar la opción de modificar.
5. Comprobar que todavía estamos a tiempo de hacer el cambio.
6. Poner el nuevo artículo que sustituye el anterior.
7. Comprobar que se ha enviado correctamente.

• Tarea 1c: comprobar la valoración de una ponencia

.....

• Tarea 2: registrarse en el congreso

.....

• Tarea 3: buscar alojamiento

.....

3) Recorrido de la usabilidad plural

Método desarrollado en los laboratorios IBM (Bias, 1995) que comparte algunos rasgos con los recorridos tradicionales, pero que tiene algunas características propias que lo definen, y que son las siguientes.

1) Este método se hace con implicados, generalmente usuarios representativos, desarrolladores y expertos en usabilidad.

2) Las pruebas se realizan con prototipos de papel u otros materiales utilizados en escenarios. Cada participante dispone de una copia del escenario de la tarea con datos que se puedan manipular.

3) Todos los participantes deben asumir el papel de los usuarios y, por lo tanto, aparte de los usuarios representativos que ya lo son, los desarrolladores y los expertos en usabilidad también lo tienen que asumir.

4) Los participantes deben escribir en cada panel del prototipo la acción que tomarán para seguir la tarea que están realizando, y escribir las respuestas tan detalladamente como sea posible.

5) Una vez todos los participantes han escrito las acciones que tomarían cuando interactuaban con cada panel, comienza el debate. En primer lugar, deben hablar los usuarios representativos y una vez estos han expuesto completamente sus opiniones, hablan los desarrolladores y después los expertos en usabilidad.

4) Inspección de estándares

Para evaluar este método, se precisa de un evaluador que sea un experto en los estándares que hay que evaluar. El experto realiza una inspección minuciosa a la interfaz para comprobar que cumple en todo momento y globalmente todos los puntos definidos en el estándar.

Un estándar es un requisito, regla o recomendación basado en principios probados y en la práctica. Representa un acuerdo de un grupo de profesionales oficialmente autorizados en un ámbito local, nacional o internacional (Smith, 1996). Tal y como indica Don Norman, "los estándares son para siempre", porque una vez establecidos, simplifican y dominan las vidas de millones, incluso miles de millones de personas.¹²

El teclado *qwerty* parece algo que siempre ha estado con nosotros, al igual que el sistema métrico decimal. ¿Cuál es el lado correcto para conducir en la carretera? ¿El derecho o el izquierdo? Obviamente, no importa mientras todos hagamos lo mismo, aunque, ¿no sería mejor para los fabricantes de automóviles y los conductores que se hubiese convenido un estándar único para todo el mun-

12. D. Norman. "Advanced TV Standards: Into the Future with Jaunty Air and an Anchor Around our Necks".

do? El impacto de acordar y, sobre todo, cambiar la conducción de todo el mundo por el mismo lado es enorme y por este motivo (entre otras razones) parece un problema inabordable, a pesar de que los costes derivados de mantener este estándar son mucho más importantes.

Con el tiempo, los costes de producción descienden, mientras que los estándares permanecen. De aquí que decidir un estándar no es una tarea que pueda tomarse a la ligera, ya que de esta decisión dependerá el futuro y, si ésta no es la mejor, será un lastre perpetuo.

Tipos de estándares

1) Los *estándares de iure* son generados por un comité con estatus legal, y están avalados por el apoyo de un Gobierno o institución para producir estándares.

El proceso para que un estándar se convierta en estándar *de iure* es bastante complejo. Resumiendo, los pasos que hay que seguir son:

- Primero se confecciona un documento preliminar que se debe hacer público.
- Después, cualquier persona o empresa puede presentar enmiendas de los borradores del documento, las cuales deben ser comentadas y resueltas.
- Después de un cierto tiempo, a veces años, se consigue un consenso y se acepta el nuevo estándar.

En informática hay una serie de comités que han participado en la creación de estándares *de iure*, de los cuales los más destacados son: la Organización Internacional para Estándares (International Organization for Standardization, ISO), la Comisión Electrónica Internacional (International Electrotechnical Commission, IEC), el Instituto Nacional Americano para Estándares (American National Standards Institute, ANSI), la sección Standards Association del Instituto de Ingenieros Eléctricos y Electrónicos Americano (Institute of Electrical and Electronics Engineers, IEEE), el Comité Europeo para la Estandarización (Comité Européen de Normalisation, CEN) y el Consorcio para la World Wide Web (World Wide Web Consortium, W3C).

2) Los *estándares de facto* que nacen a partir de productos de la industria que tienen un gran éxito en el mercado, o bien a partir de desarrollos hechos por grupos de investigación de universidades y que tienen una gran difusión. Estos

productos o proyectos de investigación llegan a tener un uso muy generalizado, y se convierten, por lo tanto, en estándares *de facto* (por ejemplo, el sistema X-Windows). Su definición se encuentra en los manuales, libros o artículos. Son técnicamente muy valiosos y muy utilizados.

Realización del método

Si bien este método podría realizarse partiendo de prototipos de baja fidelidad, lo más efectivo es realizarlo a partir de prototipos de software o incluso mejor con una primera versión del sistema final en la que estén implementadas las partes que deben confrontarse con el estándar (que normalmente serán aspectos más relacionados con la interfaz que con las funcionalidades).

En fase de análisis de requisitos se define el estándar que el sistema seguirá (ya sea porque es una especificación del proyecto o uno escogido por sus características), y el experto –en este estándar– realiza una inspección minuciosa a la totalidad de la interfaz para comprobar que cumple en todo momento y globalmente todos los puntos definidos en el estándar (Wixon, 1994). Durante esta exploración, al experto no le importa la funcionalidad de las acciones que va realizando.

4.2.2. Indagación

El proceso de indagación trata de llegar al conocimiento de una cosa, ya sea pensando o por conjeturas y señales. En los métodos de evaluación realizados por indagación, hay un gran trabajo de hablar con los usuarios y observarlos detenidamente usando el sistema en trabajo real (no para un test de usabilidad) u obteniendo respuestas a preguntas orales o por escrito.

Los principales métodos son: la observación de campo, los grupos de discusión dirigidos, las entrevistas, los cuestionarios y las grabaciones de navegación.

1) Observación de campo

Para hacer una observación de campo (Nielsen, 1993), el trabajo que se realiza es visitar el lugar o lugares de trabajo donde se estén realizando las actividades objeto de nuestro estudio y donde encontraremos a usuarios representativos.¹³

13. P. Dourish (2001). *Where the Action Is*. Cambridge: MIT Press. El autor de este libro presenta la relación entre la antropología y la etnografía con la informática social.

El principal objetivo consiste en observarlos para entender cómo realizan sus tareas y qué clase de modelo mental tienen sobre éstas. Esta información será completada con preguntas y/o entrevistas personales.

La fase del ciclo de vida del modelo de proceso más apropiada para realizar este método es, sin duda, la fase del análisis de requisitos, momento en el que necesitamos conocer el entorno de trabajo, los objetos, las personas, la organización, los métodos, las tareas que se realizan, etc., para poder recoger requisitos iniciales y opciones abiertas para incorporar en el diseño preliminar.

Procedimiento

Las visitas de campo deben prepararse previamente (la improvisación induce a la indecisión del evaluador e inseguridad a los usuarios que serán observados).

- 1) Se elige una variedad de usuarios representativos del producto (de distintos puestos de trabajo).
- 2) Se elabora la lista de las preguntas que necesitamos que nos contesten y de los datos que queremos recoger.
- 3) Hay que utilizar el sitio de observación y el tiempo con eficacia. Hay que intentar recoger tantos datos como sea posible en los puntos de observación.
- 4) Se puede hacer el análisis de datos después de volver al despacho.

Una parte del método se hace preguntando o entrevistando a los usuarios de su trabajo y la manera en la que utilizan el producto o qué pasos siguen para realizar sus tareas.

Otra parte consiste en observar cómo las personas utilizan el producto en el día a día.

2) Grupo de discusión dirigido (*focus group*)

El grupo de discusión dirigido o *focus group* (Nielsen, 1993) es una técnica de recogida de datos en la que se reúnen de seis a nueve personas (generalmente a usuarios, y a veces a implicados) para discutir aspectos relacionados con el sistema. Un evaluador experto en usabilidad y/o accesibilidad (dependiendo de lo que deseamos evaluar) hace las veces de moderador. Éste preparará previamente

la lista de aspectos que hay que discutir y recogerá la información que necesita la discusión.

Esto puede permitir capturar reacciones espontáneas del usuario e ideas que evolucionan en el proceso dinámico del grupo.

El método se puede aplicar en cualquiera de las fases de ciclo de vida del proceso de software. Aun así, en la que más se ha aplicado (y por tanto, de la que se tiene mayor experiencia) es en la fase de lanzamiento.

Procedimiento

Los procedimientos generales para dirigir un grupo de discusión dirigido son:

- 1) Localizar a usuarios representativos (típicamente de seis a nueve por grupo de discusión dirigido) que quieran participar.
- 2) Seleccionar a un moderador.
- 3) Preparar una lista de temas para ser discutidos y objetivos para asumir a partir de los temas propuestos.
- 4) Controlar la discusión sin inhibir el flujo libre de ideas y comentarios.
- 5) Asegurar que todos los participantes contribuyen a la discusión.
- 6) Procurar que no haya un participante que domine la discusión.
- 7) Dejar que la discusión discurra libremente y no de manera estructurada, pero procurar que siga un esquema planeado.
- 8) Escribir un resumen de las opiniones que han prevalecido y comentarios críticos de la sesión, incluyendo cuotas representativas.

Aspectos que hay que considerar

Cuando se dirige un grupo de discusión dirigido, hay una serie de aspectos importantes que es necesario tener en cuenta:

- 1) Tener más de un grupo principal, puesto que el resultado de una sola sesión puede no ser representativo y una sola discusión pudo haberse centrado en

un subconjunto de las ediciones o de los aspectos de menor importancia del sistema.

2) Es preferible que el moderador tenga dotes dinamizadoras y comunicativas. No es tan simple como preparar las preguntas, porque el moderador necesita facilitar y dirigir la discusión en tiempo real.

3) Debe dudarse siempre de los datos recogidos, porque debido a su naturaleza desestructurada y de flujo libre tienden a tener una validez cuestionable y son muy difíciles de analizar.

3) Entrevistas

Entrevistar a los usuarios respecto de su experiencia en un sistema interactivo resulta una manera directa de recoger información. Las entrevistas pueden ser estructuradas o abiertas; en las primeras, el evaluador es más rígido en procurar el buen seguimiento del guión preestablecido, mientras que en las abiertas se da espacio a los contertulantes para que se expresen con más libertad.

Las cuestiones pueden variar con tal de que se adapten al contexto. Normalmente, en una entrevista se sigue una aproximación de arriba abajo.

Las entrevistas pueden ser efectivas para una evaluación de alto nivel, particularmente para extraer información sobre las preferencias del usuario, impresiones y actitudes. Puede ayudar a encontrar problemas no previstos en el diseño.

Para que la entrevista sea lo más efectiva posible, debe ser preparada con antelación, con todo un conjunto de preguntas básicas, teniendo presente que el revisor puede adaptar la entrevista al entrevistado concreto para obtener el máximo beneficio.

4) Cuestionarios

El cuestionario es menos flexible que la entrevista, pero puede llegar a un grupo más numeroso y se puede analizar con más rigor. Además, se puede utilizar varias veces en el proceso de diseño.¹⁴

La base del cuestionario es la recolección de información a partir de respuestas contestadas por los usuarios.

14. B. Shneiderman (1998). *Designing the User Interface: Strategies for Effective Human-Computer Interaction* (cap. 4; 3.^a ed.). Reading: Addison-Wesley.

Hay una serie de tipos de preguntas que se pueden incluir en el cuestionario:

1) *Pregunta de carácter general.* Preguntas que ayudan a establecer el perfil de usuario y su puesto dentro de la población en estudio. Incluye cuestiones como edad, sexo, ocupación, lugar de residencia y otras.

2) *Pregunta abierta.* Preguntas útiles para recoger información general subjetiva. Pueden dar sugerencias interesantes y encontrar errores no previstos.

3) *Pregunta de tipo escalar.* Permite preguntar al usuario sobre un punto específico en una escala numérica.

Ejemplo de pregunta de tipo escalar

¿Es comprensible el diseño de los iconos? _____ Poco

1	2	3	4	5
---	---	---	---	---

 Mucho

4) *Pregunta de opción múltiple.* En este caso, se ofrecen una serie de opciones y se pide responder a una o varias.

Ejemplo de pregunta de opción múltiple

¿Qué tipo de software has utilizado?

- ☐ Tratamiento de texto
- ☐ Hoja de cálculo
- ☐ Bases de datos
- ☐ Contabilidad

Son particularmente útiles para recoger información de la experiencia previa del usuario. Un caso especial es cuando se le dan opciones para contestar *sí* o *no*.

5) *Pregunta ordenada.* Se presentan una serie de opciones que hay que ordenar.

Ejemplo de pregunta ordenada

Ordena la utilidad de cómo se ejecuta una acción:

1 la más útil; 2 la siguiente; etc.; 0 si no se utiliza.

- ☐ Por iconos
- ☐ Selección de menú
- ☐ Doble clic

Partes de un cuestionario

La actividad de la realización de cuestionarios puede estar vinculada a la consecución de ciertas tareas que el evaluador ha creído conveniente realizar (actividad combinada de varios métodos de evaluación) para medir aspectos interactivos del sistema. En estos casos, es recomendable dividir el cuestionario en tres partes:

1) *Pretarea:* las preguntas de esta sección suelen ser generales sobre ciertas habilidades del usuario (esta parte se acostumbra a aprovechar para recoger información útil sobre el perfil del usuario).

2) *Postarea:* esta sección se repetirá tantas veces como tareas tenga que resolver el usuario.

Ejemplo de postarea

Después de la realización de un juego en un sitio web orientado a un público muy joven, se pregunta lo siguiente (las respuestas son muy simples, debido a la corta edad de muchos de los usuarios):

1. ¿Has tardado mucho en poder empezar a jugar? ☐ Sí ☐ No
2. ¿Te ha gustado el juego? ☐ Sí ☐ No
3. ¿Has leído las instrucciones? ☐ Sí ☐ No
4. Si las has leído, ¿las has encontrado comprensibles? ☐ Sí ☐ No
5. ¿Tenías claro el objetivo que tenías que conseguir? ☐ Sí ☐ No

3) *Posttest:* esta sección recogerá aspectos generales acerca de la percepción general del usuario tras la consecución de las diferentes tareas planteadas.

Ejemplo de posttest

Tras la realización de todas las tareas planteadas por el evaluador del sitio web orientado a un público muy joven, se pide a cada usuario que conteste a estas preguntas:

1. El trabajo con este sitio web te ha resultado...
- ☐ Muy fácil ☐ Fácil ☐ Normal ☐ Difícil ☐ Muy difícil

2. Con las opciones del menú, saber lo que me puedo encontrar ha sido...

☐ Muy fácil ☐ Fácil ☐ Normal ☐ Difícil ☐ Muy difícil

3. ¿Identificas los juegos con los que puedes jugar a tu edad? ☐ Sí ☐ No

4. ¿Recomendarías el sitio a un amigo? ☐ Sí ☐ No

5) Registro de uso (*logging*)

La técnica de registro de uso o *logging*¹⁵ se hace a partir de "grabar" o "recoger" todas las actividades realizadas por el usuario con el sistema para su posterior análisis. Para esto, es necesaria una aplicación secundaria que realice automáticamente esta labor y que, además, pase totalmente desapercibida para el usuario.

El registro de uso implica tener en el ordenador una ampliación del sistema que recoja automáticamente estadísticas sobre el uso detallado del sistema. Es útil porque muestra cómo los usuarios realizan su trabajo real y porque es fácil recoger automáticamente datos de una gran cantidad de usuarios que trabajan bajo distintas circunstancias.

Típicamente, un registro de la interfaz contendrá una estadística sobre la frecuencia con la que cada usuario ha utilizado cada característica en el programa y la frecuencia con la que los distintos eventos de interés (tales como mensajes de error) han ocurrido. La estadística que muestra la frecuencia del uso de comandos y de otras características de sistema puede utilizarse para optimizar las tareas más frecuentes o para identificar las características que realmente se utilizan y aquellas que raramente se utilizan. La estadística que muestra la frecuencia de las distintas situaciones de error y el uso de la ayuda se puede utilizar para mejorar la usabilidad del sistema (reajustando las características que causan la mayor parte de los errores y la mayoría de los accesos a la ayuda en línea).

Procedimiento

El registro se realiza generalmente modificando los *drivers* del sistema, por ejemplo del ratón o del teclado, u otras partes del sistema que permitan el re-

15. El término *logging* proviene de analizar los ficheros de registro (ficheros con extensión log). Su misión es dejar rastro de todo lo que acontece en el contexto en el que se aplica. Nacieron con la imperiosa necesidad que tenían los implementadores de software de disponer de información acerca de lo que pasaba durante la ejecución de sus aplicaciones para depurarlas en la solución de errores e introducir mejoras; es decir, de disponer de información que ayudase a los implementadores de software.

gistro de las acciones del usuario o modificando la aplicación que estamos probando. Este último método es el preferido, ya que hace más fácil registrar acontecimientos de interés. Si los únicos datos disponibles son entrada de información y salida sin procesar, es mucho más difícil analizar los acontecimientos de gran interés para el uso del sistema, como por ejemplo situaciones del uso de alguna característica o de error. Si el sistema equipado se ejecuta en una unidad central o en sitios de trabajo con un espacio compartido del fichero, es fácil recoger datos de registro simplemente copiando los ficheros de diario de cada usuario en los intervalos regulares. Si no, puede ser necesario recoger datos de registro mediante correo electrónico automáticamente o pidiendo que los usuarios ejecuten periódicamente un programa que envíe el fichero por correo.

Características

1) Es un método muy económico.

Una de las principales ventajas de este método es que pueden analizarse las acciones de un número de usuarios muy elevado con prácticamente el mismo coste.

2) No se necesita la presencia física de los usuarios. Incluso si se aplica para analizar sitios web, no es necesario ni un espacio especial para la tarea.

3) Se puede realizar remotamente, lo cual permite evaluar un gran número de datos de infinidad de usuarios sin desplazarse a su lugar de procedencia.

4) Los datos suelen tener un formato estándar, lo cual facilita la comparación de datos según diferentes criterios (meses, días, semanas, países, etc.).

5) Los resultados se obtienen de manera instantánea. No es necesario esperar a un análisis especial de expertos para entender qué ha pasado, ni se necesitan transcripciones, ver cintas de vídeo, etc.

6) Permite tener al usuario en su entorno habitual. El *log* se recoge con el usuario en su ordenador sin que esté siendo observado, y esto ofrece datos más reales sobre el uso.

7) Muestras amplias. Normalmente estaremos hablando de miles de usuarios.

8) Muestreo de los usuarios a lo largo del tiempo, con lo cual se recoge su variabilidad.

Esta técnica se puede utilizar en las etapas de prueba de versiones beta, de despliegue o para el rediseño de aplicaciones existentes (caso para el cual es muy indicada).

¿Qué necesitamos para realizar un registro de uso de un sitio web?

Las herramientas básicas necesarias para realizar un análisis de registros o *logs* son, por una parte, el fichero que los contiene y, por otra, la aplicación que analiza este fichero.

El fichero de registros es un fichero que automáticamente almacena y gestiona el servidor web. El único inconveniente con el que podemos encontrarnos es que para poder tener acceso a éste, debe ser el encargado del sistema web (administrador de webs o *webmaster*) el que nos lo proporcione o nos diga cómo acceder al mismo.

Para analizar este fichero, existen aplicaciones disponibles –las más conocidas de las cuales son Analog o WebTrends Log Analyzer–, que generan informes rápidos y exactos sobre un sitio web y ofrecen una valiosa información acerca de sus visitas.

Extraen toda la información del registro del servidor y generan un informe completo que incluye estadísticas (numéricas y gráficas) mensuales, diarias y horarias del origen de sus visitas (*referrers*), las páginas más visitadas, los archivos servidos, informe de errores y palabras claves utilizadas en los buscadores, navegadores y sistemas operativos.

Conceptos importantes

Para realizar un análisis de registros es imprescindible tener muy claro el significado en este contexto de los siguientes términos (su mal uso y/o interpretación puede generar conflictos en la interpretación de los resultados):

1) *Accesos*. Un acceso es cualquier solicitud hecha al servidor al que se está conectado. La solicitud puede incluir cualquier elemento: páginas HTML, imágenes, ficheros audio, etc. Cada línea válida en el registro del servidor es considerada un acceso. Esta cifra representa el total de solicitudes realizadas al servidor durante un periodo de tiempo determinado.

2) *Archivos*. Algunas solicitudes hechas al servidor requieren que éste envíe “alguna cosa” como respuesta al cliente, por ejemplo una página HTML o una imagen. Cuando pasa esto, esta respuesta es considerada un archivo.

3) *Página*. Cualquier documento HTML es considerado una página. También lo es cualquier tipo de salida que se muestre en forma HTML. No incluye otros documentos ubicados dentro del documento (como imágenes gráficas, clips de audio, etc.).

A pesar de esto, el concepto de página puede variar de un servidor a otro. Sin embargo, suele estar más o menos establecido que una página es cualquier fichero con .htm, .html, etc.

4) *Cliente*. Cada solicitud realizada al servidor procede de un cliente único que es identificado por su dirección IP. El número de clientes muestra el número de direcciones IP que hacen solicitudes al servidor durante el periodo de tiempo analizado. No es el número exacto de usuarios individuales que han visitado el servidor; resulta imposible de determinar utilizando únicamente los registros y el protocolo HTTP.

5) *Visita*. Se calcula a partir del tiempo transcurrido entre dos solicitudes hechas desde una dirección IP determinada.

Ejemplo típico de la información contenida en cada línea de un fichero de registro de un servidor web

```
196.40.43.218 - - [31/May/2002:00:11:52 +0200] "GET /faq.php HTTP/1.0" 200
25258 "http://www.google.com/search?q=comercio+electronico" "Mozilla/4.77 [en]
(Win98; U)"
```

La primera parte de la línea (196.40.43.218) es la dirección IP o dirección física del visitante. A continuación vienen:

- la fecha y la hora de visita ([31/May/2002:00:11:52 +0200]),
- el archivo solicitado (GET faq.php)
- el *referrer* (http://www.google.com/search?q=comercio+electronico/), que es la página de la que proviene el navegante, y
- finalmente podemos ver el navegador (Mozilla/4.77) y el sistema operativo (Win 98) utilizados para visualizar la página solicitada.

4.2.3. Test

En los métodos de usabilidad por test usuarios representativos trabajan en tareas utilizando el sistema o el prototipo, y los evaluadores utilizan los resultados para ver cómo la interfaz de usuario soporta a los usuarios con sus tareas.

Los principales métodos de evaluación por test son la medida de las prestaciones, el *thinking aloud* (o 'pensar en voz alta'), la interacción constructiva, el test retrospectivo, el método del conductor y la ordenación de tarjetas.

1) Medida de las prestaciones

Este método de evaluación está basado en la toma de medidas acerca del rendimiento u otro tipo de aspecto subjetivo que afecte a la usabilidad del sistema, para lo cual será necesario disponer ya sea del sistema ya implementado o de un prototipo que permita evaluar estos aspectos.

Entonces, hay una serie de características importantes que hay que tener en cuenta:

a) No debemos olvidar que el objetivo primordial es mejorar la usabilidad del producto; no debe confundirse con un test de funcionalidad, el cual tiene como objetivo garantizar que el producto funcione de acuerdo con las especificaciones.

b) Los participantes en el test (que evidentemente serán usuarios reales realizando tareas reales) se analizarán tanto en la manera en la que utilizan el producto como midiendo el tiempo que les lleva realizarlo.

c) A pesar de que también puede realizarse en el entorno del usuario, es muy apropiado realizar este método en un laboratorio de usabilidad.

Un aspecto primordial para la realización del test será la selección de tareas que los usuarios deberán realizar. Los siguientes puntos deberán tenerse presentes a la hora de escoger estas tareas:

- Tareas que demuestren problemas de usabilidad. El criterio más importante para seleccionar tareas es utilizar tareas que prueben los problemas potenciales de usabilidad del producto. Como en cualquier otro procedimiento de test, cuantos más problemas encontremos, mejor.
- Tareas sugeridas por la propia experiencia. Los desarrolladores siempre tienen algunas ideas respecto a dónde encontrar problemas. Saben qué

partes del producto fueron más difíciles de diseñar y cuáles son los problemas que se tienen que probar.

- Tareas derivadas de otros criterios, como por ejemplo las tareas que son difíciles de recuperar después de un error.
- Tareas que los usuarios harán con el producto. Se seleccionan tareas habituales en el día a día de los usuarios, para optimizar la usabilidad de los aspectos más cotidianos.

Procedimiento

Tractaremos en primer lugar de comprender qué se puede medir. Para esto, podemos recoger:

1) Medidas de rendimiento

Esto quiere decir explicar las acciones y los comportamientos. Este tipo de medidas son cuantitativas, y se pueden contar personas, cuántos errores se cometen, cuántas veces repiten el mismo error, etc.

Ejemplos típicos de medidas de rendimiento de un test de usabilidad

- Tiempo
 - para completar una tarea.
 - consumido en menús de navegación.
 - consumido en ayuda en línea.
 - para buscar información en un manual.
 - invertido para recuperarse de errores.
- Número
 - de opciones de menú erróneas.
 - de opciones incorrectas en cajas de diálogo.
 - de selección de iconos incorrectos.
 - de teclas de función mal seleccionadas.
 - de llamadas a la ayuda.
 - de pantallas de ayuda en línea.
 - de veces que se consulta el manual.
- Observaciones
 - de frustración.
 - de confusión.
 - de satisfacción.

2) Medidas subjetivas

Esto quiere decir percepciones de las personas, opiniones y juicios. Pueden ser cuantitativas o cualitativas.

Ejemplos típicos de medidas subjetivas de un test de usabilidad

- Reflexiones sobre la
 - Facilidad de uso del producto
 - de aprender el producto
 - de hacer una determinada tarea
 - de instalar el producto
 - de encontrar información en el manual
 - de comprender la información
 - Utilidad de los ejemplos de ayuda
- Preferencias o razones de la preferencia
 - De una versión previa.
 - Sobre un producto de la competencia.
 - De la manera en la que están haciendo las tareas ahora.
- Predicciones de comportamiento
 - ¿Comprará el producto?
- Comentarios espontáneos
 - Estoy totalmente perdido.
 - Ha sido fácil.
 - No comprendo el mensaje.

3) Pensar en voz alta (*thinking aloud*)

En este método de evaluación (Nielsen, 1993) se pide a los usuarios que expresen en voz alta sus pensamientos, sentimientos y opiniones mientras interactúan con el sistema. Es muy útil en la captura de un amplio rango de actividades cognitivas.

Procedimiento

Se proporciona a los usuarios el prototipo que hay que probar y un conjunto de tareas para realizar. Se les pide que realicen las tareas y que expliquen en voz alta qué es lo que piensan al respecto mientras están trabajando con la interfaz, y que describan qué es lo que creen que está pasando, por qué toman una u otra acción o qué es lo que están intentando realizar. En definitiva, cuantas más impresiones, mejor.

Pensar en voz alta permite a los evaluadores comprender cómo el usuario se aproxima a la interfaz y qué consideraciones tiene en la mente cuando la usa.

El usuario puede expresar que la secuencia de etapas que le dicta el producto para realizar el objetivo de su tarea es diferente de la que esperaba.

Aunque el principal beneficio del método es una mejor comprensión del modelo mental del usuario y la interacción con el producto, hay asimismo otros beneficios, como por ejemplo conocer la terminología que el usuario utiliza para expresar una idea o función que debería ir incorporada en el diseño del producto o, al menos, a su documentación.

Este método tiene como ventaja más importante la simplicidad; requiere realmente poca experiencia para poder llevarlo a cabo y puede proporcionar ideas muy útiles acerca de los problemas con una interfaz (Dix y otros, 1993, pág. 386).

Otras ventajas importantes del método son:¹⁶

- a) Puede ser utilizado para observar cómo el sistema se utiliza actualmente (como complemento para la observación etnográfica).
- b) Puede realizarse en todas las fases del ciclo de vida (incluso en las más iniciales), y con cualquier tipo de prototipo.
- c) Se trata de un método muy económico.

Y como principales inconvenientes tenemos:

- a) La información propuesta es subjetiva y selectiva (dependerá de las tareas elegidas).
- b) El proceso de observación puede alterar la manera en la que los usuarios realizan sus tareas y, por lo tanto, es posible obtener vistas parciales (describir lo que uno hace con frecuencia cambia la forma de hacerlo).

Una variante de pensar en voz alta es el método conocido como *evaluación cooperativa*, que estimula al usuario a verse a sí mismo como un colaborador de la evaluación más que un simple sujeto experimental.

4) Interacción constructiva o aprendizaje por codescubrimiento

Este método es una derivación del de pensar en voz alta, e implica tener, en vez de a uno, a dos usuarios que hagan el test del sistema conjuntamente (O'Malley y otros, 1984).

16. Puede ver el epígrafe "Observación de campo" en el subapartado 4.2.2 de este capítulo.

La principal ventaja de este método es que la situación del test es mucho más natural que pensar en voz alta con usuarios individuales, ya que las personas normalmente verbalizan cuando tratan de resolver un problema conjuntamente y, además, hacen muchos más comentarios.

Un aspecto que hay que tener en cuenta es que la interacción constructiva requiere el doble de usuarios que el método de pensar en voz alta, aspecto importante si tenemos restricciones económicas.

Este método tiene la desventaja de que los usuarios pueden tener diferentes estrategias de aprendizaje.

5) Test retrospectivo

Si se ha realizado una grabación en vídeo de la sesión de test, es posible recoger más información si se hace que el usuario revise la grabación (Hewett, 1987).

Los comentarios del usuario mientras revisa el vídeo son más extensos que mientras ha estado trabajando en la tarea de test y, por tanto, es posible que el evaluador pare el vídeo y pregunte al usuario con más detalle, sin tener miedo de interferir con el test que esencialmente ha sido completado.

El aspecto negativo más evidente es que con cada usuario se tarda como mínimo el doble del tiempo necesario con cualquier otro método.

6) Método del conductor (*coaching method*)

El método del conductor (Mack, 1992) es algo diferente de estos métodos de test vistos hasta ahora, porque hay una interacción explícita entre el usuario y el evaluador (o conductor). Este último, en los otros métodos, trata de interferir en el usuario tan poco como sea posible mientras está haciendo el test.

Este caso es totalmente al contrario en este aspecto: se conduce al usuario en la dirección correcta mientras se usa el sistema. Durante el test, el usuario puede preguntar al evaluador cualquier aspecto relacionado con el sistema y éste le responderá.

Este método se centra en el usuario inexperto, y su propósito es descubrir las necesidades de información de los usuarios de tal manera que se proporcione un mejor entrenamiento y documentación, al mismo tiempo que un posible rediseño de la interfaz para evitar la necesidad de preguntas.

7) Ordenación de tarjetas (*card sorting*)

Este método tiene un enfoque distinto a los presentados anteriormente y a pesar de que, como veremos, suele estar enfocado a relacionar aspectos de la arquitectura de la información en el diseño de sitios web, es un método completamente válido para relacionar la información de cualquier medio interactivo.

Cuando se empieza un nuevo ejercicio de diseño de la información, es normal encontrarse con una larga lista de ítems sin relacionar que hay que incluir. El reto es organizar esta información de manera que sea útil y comprensible para los usuarios del sistema.

Aunque es cierto que cuidadas investigaciones sobre el análisis de la información pueden revelar algunas pistas, difícilmente se podrá determinar qué temas deben agruparse entre sí.

La dificultad de organizar el contenido procede de la falta de conocimiento sobre cómo usan los usuarios reales este contenido. Y sin este conocimiento, cualquier intento de organizar esta información no deja de ser un puro ejercicio teórico (Robberston, 2001).

La técnica conocida como *ordenación de tarjetas* o *card sorting* es la utilizada para conocer cómo los usuarios visualizan la organización de la información. El diseñador utiliza las aportaciones de los usuarios para decidir cómo deberá estructurarse la información en la interfaz.

Se trata de una técnica simple –fácil de entender y aplicar–, barata, rápida y que involucra a los usuarios, especialmente indicada cuando disponemos de una serie de ítems que necesitan ser catalogados y para decidir la estructura organizativa de los sitios web.

Procedimiento

Para implementar una ordenación de tarjetas, hay que seguir los pasos siguientes:

1) *Determinar la lista de temas.* Identificar la lista de los ítems que hay que ordenar. Esta lista no debería ser muy extensa (debe ser manejable), al mismo tiempo que tiene que resultar comprensible para todos los participantes de la sesión. El evaluador no debe poner ningún tipo de indicación que pueda influenciar a los usuarios es su decisión, ni ningún tópico que induzca a la agrupación.

pación de términos (por ejemplo, archivo, edición, preguntas más frecuentes o siglas concretas que tienen significado propio en la aplicación).

2) *Crear las tarjetas.* Cada tópico deberá escribirse en una tarjeta (papel, cartón), la cual ocasionalmente puede adjuntar algún tipo de explicación (aunque no es muy recomendable). Además, deberá proporcionar tarjetas en blanco a los participantes.

3) *Seleccionar a los participantes.* Los participantes preferentemente serán usuarios finales (gerentes y otros implicados no son usuarios finales), de los que deberemos estar seguros de que representan fielmente a grupos de usuarios potenciales del sistema.

4) *Proceder con la sesión (o sesiones) de ordenación.* Cada sesión debe comenzar con una explicación del método y de los objetivos que anime a todos los participantes a organizar las tarjetas y etiquetar los grupos según sus criterios personales. El organizador de la sesión debe tomar nota de todo aquello que pueda resultar relevante para la evaluación final.

5) *Analizar las agrupaciones.* Una vez han concluido todos, el evaluador deberá analizar todas las agrupaciones en un ejercicio de "análisis democrático" para identificar aquellas agrupaciones más frecuentes y, de esta manera, poder decidir la estructura final.

4.3. Métodos de evaluación de la accesibilidad

Esta sección está dedicada a conocer las herramientas más relevantes en el campo de la validación y de la reparación de los aspectos relacionados con la accesibilidad de los sistemas interactivos. La mayoría se rigen por las pautas marcadas por la WAI, comentadas en el capítulo III.

4.3.1. Herramientas de validación

Las herramientas de evaluación o de validación permiten realizar un análisis de páginas y sitios web y proporcionan un informe o un valor de una escala pre-fijada respecto a esta accesibilidad.¹⁷

17. Es muy recomendable la nueva especificación técnica de la ISO 16071 (del 2003) como pauta para la accesibilidad de las interfaces de sistemas basados en ordenadores en general.