

1.9 SITIOS WEB RECOMENDADOS

- <http://www.apa.org/ethics/code/>

Es el sitio de la *American Psychological Association* (APA) en el que se puede encontrar su código ético (*Ethical Principles of Psychologists and Code of Conduct*), es muy interesante leerlo para reflexionar sobre los diferentes aspectos éticos, especialmente su estándar número 8 relativo a investigación y publicación.

- <http://www.research.umn.edu/consent/>

Este sitio la Universidad de Minnesota ofrece un pequeño tutorial sobre el consentimiento informado.

- <http://start.aisnet.org/?CodeofResearch>

En este sitio se encuentra la última versión del Código de Conducta en Investigación de la *Association for Information Systems* (AIS) que puede servirnos de guía a la hora de abordar algunos dilemas éticos en la investigación en ingeniería del *software*.

ENCUESTAS

2.1 INTRODUCCIÓN

Las encuestas son, probablemente, el método de investigación más utilizado por todo el mundo. De hecho, es muy común que en nuestra vida diaria se nos ofrezca, muy a menudo, participar en una encuesta por nuestro papel de electores, consumidores o usuarios de servicios. Este uso tan extendido de las encuestas puede hacer que las investigaciones basadas en encuestas parezcan una forma fácil y directa de conseguir información importante sobre productos, contextos, procesos, personas, etc. Sin embargo, la realidad no es esa.

Una encuesta es un método empírico que se utiliza para recopilar información de o sobre personas para describir, comparar o explicar su conocimiento, sus actitudes o su comportamiento. También se pueden utilizar encuestas para la descripción de las características de métodos o herramientas.

En la mayoría de los casos, los datos relativos a la encuesta provendrán de cuestionarios. Pero los cuestionarios, por sí solos, no constituyen la encuesta. De hecho, una encuesta es un proceso más complejo formado por una serie de actividades bien definidas (Kitchenham y Pfleeger, 2008) que se enumeran a continuación y que se irán explorando a lo largo de este capítulo:

- Establecer los objetivos de la encuesta.
- Diseñar la encuesta.

- Desarrollar el cuestionario.
- Evaluar y validar el cuestionario.
- Obtener los datos de la encuesta.
- Analizar los datos obtenidos.
- Reportar los resultados.

2.2 PROCESO DE REALIZACIÓN DE ENCUESTAS

En este apartado se describen las tareas que es recomendable realizar en cada una de las actividades del proceso para la realización de encuestas.

2.2.1 Establecer los objetivos de la encuesta

Establecer los objetivos es siempre el primer paso de una encuesta o de cualquier otro tipo de investigación. En el caso concreto de las encuestas, cada objetivo debería establecerse como una frase relativa a los resultados esperados tras realizar la propia encuesta.

Algunos ejemplos válidos de objetivos podrían ser determinar el lenguaje de programación más atractivo para un determinado grupo de programadores de *software* o la metodología de desarrollo más utilizada por las empresas de un país.

El origen de los objetivos puede ser diverso y, por ejemplo, un objetivo puede surgir como resultado de una duda que haya aparecido tras una búsqueda en la literatura, puede que haya algún tipo de necesidad que se perciba por parte de los investigadores responsables y se pretenda explorar las distintas posibilidades asociadas a ella o puede, incluso, que los haya planteado directamente algún experto.

En cualquier caso, los objetivos de las encuestas deben ser lo más claros posibles y han de poder medirse. Dichos objetivos determinarán la mayoría del resto de actividades del proceso de realización de encuestas, de ahí que deban establecerse cautelosamente.

2.2.2 Diseñar la encuesta

Los dos tipos de diseño de encuestas más utilizados son (Kitchenham y Pfleeger, 2008):

- **Encuestas transversales:** en este tipo de estudio, se pide información a los participantes en un instante determinado. Por ejemplo, podríamos sondear a todos los miembros de la plantilla de una factoría de desarrollo de *software* a las 11:00 horas para saber qué actividades están llevando a cabo en ese preciso momento. Esta información nos daría una instantánea de qué ocurre en una organización en un momento determinado. La mayoría de encuestas que se realizan en la ingeniería del *software* son de este tipo.
- **Encuestas longitudinales:** en este tipo de estudio la meta se establece a más largo plazo y se trata de conocer la evolución a través del tiempo de una determinada población. Podemos encontrarnos con dos variantes principales de encuestas longitudinales, según si la población que elegimos en los distintos momentos en los que se realiza la encuesta es la misma o varía.

Hay algunos otros tipos más complejos de diseño de encuestas, por ejemplo, diseños que comparan poblaciones diferentes u otros que pretenden medir el impacto de un cambio. Si se está interesado en alguno de estos diseños se puede consultar *Shaddish et al. (2002)*.

Otro aspecto importante a tener en cuenta a la hora de diseñar una encuesta es cómo se va a administrar. Existen, de nuevo, diversas opciones (Kitchenham y Pfleeger, 2008):

- Cuestionarios auto-administrados (vía Internet).
- Encuestas telefónicas.
- Entrevistas personales.

Elegir una u otra de estas opciones determinará el tipo de preguntas que podrán realizarse en la encuesta. Además, algunas estrategias para obtener datos fiables, como el orden de las preguntas o el vocabulario utilizado, también vienen influidas por cómo se vaya a administrar la encuesta.

Cabe destacar que en el campo de la ingeniería del *software*, los cuestionarios auto-administrados son la opción más utilizada, por lo que será este tipo el que se estudiará en este capítulo.

2.2.3 Desarrollar el cuestionario

En muchas disciplinas, los investigadores suelen basarse en cuestionarios ya existentes o, como mucho, en ligeras variaciones de los mismos. Este enfoque es útil puesto que los cuestionarios existentes ya se han validados con anterioridad y permite, además, comparar los resultados que se obtengan en el estudio con los previamente obtenidos.

Sin embargo, en la ingeniería del *software*, la mayoría de las investigaciones parten de cero, por lo que utilizar material ya existente no suele ser una opción viable y es necesario que los investigadores desarrollen sus propios cuestionarios.

Una encuesta pretende obtener respuestas sobre una serie de preguntas por una razón determinada. Por tanto, a la hora de diseñar un cuestionario conviene partir del objetivo de la investigación, si bien traducir directamente objetivos a preguntas rara vez lleva a un cuestionario realmente útil, por lo que, si se pretende conseguir una encuesta realmente efectiva habrá que diseñar correcta y cautelosamente los cuestionarios que se utilizarán para la recolección de los datos.

En el resto de sub-apartados se comentarán las distintas tareas a llevar a cabo para desarrollar correctamente el cuestionario. En concreto, la consulta de la literatura relevante a la investigación que se esté llevando a cabo a través de la encuesta, la elección del tipo de las preguntas del cuestionario, el diseño de las mismas, así como el diseño de las respuestas a las preguntas, el formato a seguir por los cuestionarios y, por último pero no por ello menos importante, los aspectos motivacionales a tener en cuenta para conseguir una participación suficiente en la encuesta, en cuanto a cantidad y calidad.

2.2.3.1 CONSULTA DE LA LITERATURA RELEVANTE

Toda buena investigación que se precie, ha de comenzar con esta fase que pretenderá conseguir un doble objetivo: por un lado se tratará de identificar qué otros estudios se han llevado a cabo con anterioridad sobre el mismo tema, por otro, será de gran ayuda conocer cómo se han conseguido los datos en estos estudios previos, si es que los hay. Más en concreto, habrá que centrarse en la posibilidad de conseguir los cuestionarios o cualquier otro método de recolección de datos que se haya realizado con anterioridad en investigaciones relacionadas con el mismo tema.

2.2.3.2 TIPOS DE PREGUNTAS

En un cuestionario pueden encontrarse dos tipos de preguntas: abiertas o cerradas. Una pregunta abierta permite al encuestado expresar la respuesta utilizando sus propias palabras, mientras que una pregunta cerrada proporciona una lista de posibles respuestas a la pregunta.

Cada tipo de pregunta tiene sus ventajas e inconvenientes. Las preguntas abiertas permiten que el encuestado se exprese libremente y utilizando el lenguaje que estime más conveniente pero, a la vez, ese mismo hecho se vuelve un inconveniente al permitir que se produzcan interpretaciones erróneas sobre las respuestas obtenidas o que se obtenga una gran cantidad de información irrelevante y no deseada. En resumen, las preguntas abiertas suelen ser difíciles de codificar y analizar, por lo que suele ser mucho más conveniente que el cuestionario esté formado por tantas preguntas cerradas como se pueda.

2.2.3.3 DISEÑO DE LAS PREGUNTAS

Una vez se tiene claro qué preguntar, es necesario pensar detenidamente cómo se quieren plantear las preguntas. Las preguntas han de ser precisas, carentes de ambigüedad y perfectamente comprensibles por parte de todas aquellas personas que vayan a responderlas. Para conseguir estas características conviene tener en cuenta los siguientes factores (Kitchenham y Pfleeger, 2008):

- El lenguaje utilizado es el apropiado para las personas que responderán la encuesta y cualquier término potencialmente ambiguo se deberá definir explícitamente.
- Se evitarán errores gramaticales, de puntuación o de deletreo.
- En cada pregunta se abordará un único concepto para conseguir preguntas concisas y concretas.
- Se evitará el uso de calificadores ambiguos.
- Nunca se utilizarán jergas ni expresiones coloquiales.
- Las preguntas podrán ser positivas o negativas, evitando dobles negaciones.
- Se evitará preguntar sobre sucesos acaecidos hace mucho tiempo.
- Se evitarán preguntas sensibles que incomoden a los participantes.
- Las preguntas se podrán responder de manera sencilla para evitar que los participantes se frustren al no conocer la respuesta a las mismas.

2.2.3.4 DISEÑO DE LAS RESPUESTAS

Las respuestas a las preguntas de un cuestionario suelen ser alguna de las contenidas en la siguiente lista:

1. Valores numéricos (p.ej. la edad).
2. Categorías (p.ej. el tipo de trabajo).
3. Respuestas SI/NO.
4. Escalas ordinales.

En el caso de los valores numéricos no se suelen encontrar dificultades, pero sí conviene hacer algunas recomendaciones sobre el resto.

En las respuestas enmarcadas en categorías será necesario que dichas categorías cumplan ciertos requisitos (Kitchenham y Pfleeger, 2008):

- Ser exhaustivas pero no demasiado largas.
- Ser mutuamente excluyentes.
- Permitir una selección múltiple si fuera necesario.
- Incluir una categoría "Otros" si no se cubren explícitamente todas las posibles respuestas.

Las respuestas SI/NO son especialmente problemáticas. Contienen un carácter restrictivo (sólo dos respuestas posibles y opuestas) y son poco fiables (la misma persona puede dar distintas opuestas en momentos diferentes). Es por ello que en respuestas que expresen actitudes o preferencias sea mucho más interesante utilizar respuestas en una escala ordinal. Hay tres tipos de escala:

1. Escalas de conformidad, p.ej. desde totalmente de acuerdo hasta totalmente en desacuerdo. Este tipo de escala también se conoce como escala de Likert (Likert, 1932).
2. Escalas de frecuencia, p.ej. desde nunca hasta siempre.
3. Escalas de evaluación, p.ej. desde terrible hasta excelente.

En la mayoría de los casos se recomienda utilizar una escala de entre cinco y siete posibles valores (Lethbridge, 1998). Además, conviene explicitar dichos valores y no utilizar números para etiquetarlos, puesto que pueden llevar a

interpretaciones por parte de los participantes que introduzcan un sesgo en los resultados (Krosnick, 1990). Como mucho, si se estima conveniente se puede incorporar una numeración a las etiquetas lingüísticas, especialmente si la diferencia entre dichas etiquetas es muy pequeña.

Un último aspecto a tener en cuenta sobre las respuestas es la posibilidad de incluir una opción NS/NC, es decir, "no sabe o no contesta". No hay un acuerdo en la comunidad científica sobre este aspecto aunque, como norma general, se recomienda no incluir dicha opción si los participantes han sido seleccionados porque deberían poder responder a todas las preguntas.

2.2.3.5 FORMATO DE LOS CUESTIONARIOS

Para conseguir que el aspecto físico de los cuestionarios que se entreguen en papel (aunque la mayoría de ellas son directamente aplicables a un cuestionario web) no introduzca complejidad innecesaria en la encuesta existen, entre otras, las siguientes recomendaciones (Kitchenham y Pfleeger, 2008):

- Dejar un espacio para que los participantes dejen sus comentarios sobre el cuestionario.
- Usar espacio entre las distintas preguntas.
- Utilizar un tamaño de fuente entre 10 y 12 puntos.
- Evitar el uso de cursivas.
- Enfatizar la información que lo requiera mediante el uso de negritas, subrayado y mayúsculas.
- No dividir entre páginas las instrucciones, una pregunta o las respuestas asociadas a una pregunta.

El orden en que se plantean las preguntas también es importante. Lo normal es seguir un orden lógico, haciendo que las preguntas más sencillas se encuentren al principio y que el nivel de dificultad avance según los participantes van rellenando el cuestionario.

En muchos casos, los cuestionarios contienen preguntas demográficas. Si bien no hay consenso sobre dónde colocar estas preguntas, deberán colocarse al principio o al final del cuestionario.

2.2.3.6 MOTIVACIÓN

Uno de los principales retos a la hora de abordar la realización de una encuesta es conseguir motivar a los participantes para que respondan de manera fiel a una encuesta en la que, en la mayoría de los casos, no han pedido participar.

En líneas generales, las personas estarán motivadas y darán respuestas completas y precisas siempre y cuando perciban que los resultados del estudio les serán útiles. Es por ello que se recomienda incluir cierta información clave en el cuestionario de la encuesta (Kitchenham y Pfleeger, 2008), como por ejemplo:

- ¿Cuál es el objetivo del estudio?
- ¿Cómo puede ser relevante para los participantes?
- ¿Por qué es importante la participación de cada individuo?
- ¿Cómo y por qué se eligió a cada participante?
- ¿Cómo se implementa la confidencialidad de las respuestas?

2.2.4 Evaluar y validar el cuestionario

Una vez se han definido las preguntas del cuestionario, los investigadores suelen pensar que ya se puede pasar a realizar la encuesta y obtener los datos de la misma. Éste es un error muy común que debe evitarse, ya que el conjunto de preguntas del cuestionario constituye únicamente el punto de partida de la construcción de los cuestionarios y, una vez creado, es esencial que se evalúe (Litwin, 1995).

En esta actividad de evaluación, o pre-test como también se conoce, se pretenden diversos objetivos (Kitchenham y Pfleeger, 2008):

- Comprobar que las preguntas se entienden correctamente.
- Evaluar el índice probable de respuestas.
- Evaluar la fiabilidad y validez del cuestionario, a través de grupos de discusión y/o de estudios pilotos.
- Comprobar que el método de análisis de datos que se utilizará será compatible con las respuestas que se van a obtener.

En los grupos de discusión se cuenta con un conjunto de personas representativo de aquellos que utilizarán los resultados de la encuesta o de aquellos que la realizarán (o incluso una mezcla de ambos). Se trata de que estos individuos rellenen el cuestionario e identifiquen cualquier problema asociado a él. Así, los grupos de discusión deberían identificar preguntas que faltan o que sobran, así como preguntas o instrucciones ambiguas.

En los estudios piloto, las encuestas se realizan tal cual se tienen diseñadas, pero utilizando una muestra mucho más pequeña de participantes. Estos estudios pretenden detectar problemas en el propio cuestionario, el tiempo permitido para responder, etc.

Una vez se ha construido y validado el cuestionario asociado a la encuesta conviene comenzar a documentar la misma (Bourque y Fielder, 1995). Normalmente, se suele comenzar con un documento descriptivo inicial, denominado especificación del cuestionario, que debería contener:

- Los objetivos del estudio.
- Una descripción del porqué de cada pregunta.
- Motivos para haber adoptado o adaptado preguntas de otras fuentes, incluyendo las citas apropiadas.
- Una descripción del proceso de evaluación.

Más tarde, cuando el cuestionario se haya publicado, se debe ir actualizando la información relativa a los datos de los participantes, cómo se administró el cuestionario, cómo se procesaron las preguntas, etc.

Un buen motivo para ir preparando esta documentación en paralelo con el desarrollo de la encuesta es que, en la mayoría de los casos, la recolección de datos puede extenderse mucho en el tiempo y puede que pasen meses entre que se distribuyen los cuestionarios y se comienzan a analizar los resultados. Esta dilación puede hacer que se olviden fácilmente detalles sobre la creación de los cuestionarios o sobre la gestión de los mismos.

2.2.5 Obtener los datos

A la hora de conseguir los datos, normalmente es imposible contar con las respuestas de toda la población implicada en el estudio, de ahí que haya que recurrir a una *muestra* de la misma, con la esperanza de que sus respuestas representen a las respuestas que hubiera dado el conjunto completo.

Dentro de los métodos de selección de la muestra, destacan tres tipos principales (Kitchenham y Pfleeger, 2008):

- **Métodos probabilísticos.** En estos métodos, cada elemento de la población objetivo tiene una probabilidad conocida y distinta de cero de ser incluido en la muestra.
- **Métodos basados en grupos (*clusters*).** En estos casos, cada individuo de la población objetivo pertenece a un grupo bien definido.
- **Métodos no probabilísticos.** Estos métodos se utilizan cuando los sujetos que realizarán la encuesta se eligen porque son fácilmente accesibles por parte de los investigadores responsables o porque se tiene alguna justificación para creer que son representativos de la población objetivo.

Con el fin de obtener una muestra representativa para una encuesta hay tres factores esenciales a tener en cuenta: ausencia de sesgos y grado de adecuación de los participantes y la relación coste-efectividad que presentan.

No hay que perder de vista varios aspectos a la hora de seleccionar quién va a participar en una encuesta. El primero de ellos es el tamaño de la muestra. Si el tamaño no es suficientemente grande, las conclusiones no serán razonables y será difícil generalizarlas. Eso sí, no hay una ecuación que marque exactamente cómo de grande debería ser una muestra (Fowler, 2002) ni hay una fórmula que permita comprobar cuándo un tamaño de la muestra pasa de ser no representativo a representativo. Lo que se recomienda es que se tenga una muestra lo más representativa (es decir, de un tamaño mayor) posible de todos los subgrupos de individuos que pudiera haber en la población.

En cualquier caso y, a pesar de lo imposible de la tarea, determinar el número de sujetos que participarán en la encuesta no es lo único importante. Cualquier encuesta que se precie ha de indicar su tasa de respuesta, es decir, el porcentaje de sujetos contactados para realizar la encuesta que finalmente completaron los cuestionarios. Aun así, una tasa de respuesta alta no implica que se vaya a conseguir resultados más precisos (Krosnick, 1990).

Encontrar un compromiso entre el tamaño de la muestra y la tasa de respuesta de una encuesta no es un proceso sencillo y en muchos casos dependerá de factores ajenos a los investigadores. Conviene, por tanto, utilizar los recursos con los que se cuente, como enviar recordatorios en listas de correo o, llegado el caso, individualmente a aquellos participantes potenciales que aún no hayan completado los cuestionarios.

2.2.6 Analizar los datos

Una vez diseñada y llevada a cabo la encuesta es el momento de pasar a analizar los datos obtenidos. En este apartado se tratarán los puntos más importantes relativos al análisis de los mismos. En concreto, se estudiará la validación de los datos obtenidos, la división en las respuestas en grupos homogéneos y, finalmente, el análisis de los datos obtenidos, tanto ordinales como nominales.

2.2.6.1 VALIDACIÓN DE LOS DATOS

Antes de embarcarse en ningún análisis, es necesario comprobar que los datos con los que se cuenta son completos y consistentes. Para ello, conviene contar con una política de gestión de cuestionarios inconsistentes. Si se observa que la mayoría de los participantes han respondido a todas las preguntas, probablemente convenga rechazar aquellos que estén incompletos. Si, por el contrario, hay una o más preguntas cuya respuesta se ha omitido sistemáticamente por parte de los participantes, convendrá rechazar esa(s) pregunta(s) en concreto.

Podría darse el caso de que, aun estando incompletos, convenga analizar todos los cuestionarios, lo que nos llevará a tener un tamaño de la muestra diferente para las preguntas. Si ocurre esta situación, habrá que documentarlo convenientemente y explicitar el tamaño de la muestra para cada pregunta. Conviene tener en cuenta que este enfoque se podrá utilizar para calcular estadísticos simples o comparar medias, pero no para estudios de correlación o regresión.

Para evitar estas situaciones es muy importante que los cuestionarios se validen con anterioridad a la realización de la encuesta.

2.2.6.2 DIVISIÓN DE LAS RESPUESTAS

En ocasiones, antes de comenzar a analizar será conveniente dividir y agrupar las preguntas para contar con grupos homogéneos. Estos agrupamientos se suelen hacer en función de la información demográfica que se ha obtenido de los participantes, por ejemplo, basándose en la edad, la localización geográfica, la experiencia en algún campo concreto, etc., según convenga en cada caso.

2.2.6.3 ANÁLISIS DE DATOS ORDINALES Y NOMINALES

Los datos numéricos generados por una encuesta se podrán analizar utilizando las técnicas habituales, sin embargo, puede que los datos ordinales y nominales generen algún problema adicional, por lo que serán éstos los que trataremos en este sub-apartado.

Si se ha utilizado un cuestionario que se responde en una escala ordinal, se suele convertir dicha escala en sus valores numéricos correspondientes (p.ej. desde 1 hasta 7) y analizar dichos datos como si fueran datos numéricos simples. Es una solución razonable pero que viola las reglas matemáticas para analizar datos ordinales, lo que conlleva un riesgo que puede dar lugar a unos resultados imprecisos.

Para evitar esta situación existen tres posibilidades principales:

1. Utilizar las propiedades de una distribución polinomial, estimar la proporción de la población en cada categoría y determinar el error estándar de la estimación (Moses, 2000).
2. Convertir una escala ordinal en una variable dicotómica. Por ejemplo, si se pretende calcular qué proporción de la población está de acuerdo con una determinada afirmación, pueden reagruparse las respuestas afirmativas (totalmente de acuerdo, de acuerdo, etc.) en una única variable (1) y el resto en otra (0) y utilizar las propiedades de una distribución binomial.
3. Utilizar el coeficiente de correlación de *Spearman* o la *tau de Kendall* para comprobar asociaciones entre variables de escala ordinal (Siegel y Castellan, 1998).

2.2.7 Reportar los resultados

En la mayoría de las ocasiones, se tratará de publicar la información generada durante la realización de la encuesta y el análisis de sus resultados de la misma en alguna revista o congreso o incluso como un informe técnico asociado a algún proyecto de investigación.

Hay una serie de elementos que deben incluirse en el documento que servirá como documentación de la encuesta. Se trata de los siguientes (Fink, 2003):

- Título, autores, patrocinadores, localización y fecha.
- Introducción, en la que se debe dejar bien claro cuál es el problema o la necesidad a resolver e incluso las preguntas de investigación y/o las hipótesis a ser validadas.
- Características de la encuesta:
 - Tipo de cuestionarios y justificación de la elección.
 - Contenido del cuestionario: número de preguntas, descripción del contenido de las preguntas, tipos de respuestas, etc.
 - Escalas utilizadas, valoración de las preguntas y si se han agrupado preguntas.
 - Fiabilidad y validez del cuestionario, incluyendo los estudios piloto realizados, el tiempo de respuesta y cualquier otro indicador de calidad que se haya utilizado.
- Características de los administradores de la encuesta.
- Literatura relevante sobre el tema de investigación.
- Diseño de la encuesta, muestra escogida y análisis de resultados.
- Relación de los resultados con los objetivos de la encuesta.
- Conclusiones alcanzadas, implicaciones derivadas de la encuesta y trabajo futuro.

2.3 FIABILIDAD Y VALIDEZ DE LAS ENCUESTAS

Una encuesta es fiable cuando, por más veces que se repita, se obtienen siempre unos resultados similares. La fiabilidad de una encuesta puede comprobarse repitiendo exactamente la misma encuesta (o con ligeras variaciones en cuanto a la redacción o el orden de presentación de las preguntas, por ejemplo) a los mismos sujetos y verificando si las respuestas obtenidas son similares.

A nivel estadístico se suele utilizar el coeficiente alfa de *Cronbach* (Cronbach, 1951) que, en función de su valor, indica la correlación de las distintas preguntas. Se suele recomendar un valor mínimo del coeficiente de 0,7 para poder afirmar que los distintos elementos están correlacionados.

En cuanto a la validez de las encuestas, se suelen estudiar distintos tipos (Kitchenham y Pfleeger, 2008):

- **Validez de contenido:** establece una visión subjetiva sobre lo apropiado que es el contenido de los cuestionarios en relación a los sujetos que los realizan.
- **Validez de criterio:** establece la capacidad del cuestionario de distinguir a qué grupo pertenece cada uno de los sujetos que lo responden.
- **Validez del constructo:** se refiere a cómo de bien se consigue medir a través del cuestionario aquello que se pretende medir.

Desafortunadamente, la importancia que en el campo de la ingeniería del *software* se le da al estudio de la fiabilidad y la validez de las encuestas es aún bastante escasa. Aun así, existen algunos ejemplos donde sí que se llevan a cabo, como el que se presenta en Dybå (2000). En este trabajo se presenta un estudio de la fiabilidad a través del estudio de diversas correlaciones entre los elementos de la encuesta y un estudio de la validez desde los tres puntos de vista comentados con anterioridad.

Algunos otros ejemplos de encuestas que incluyen estudios sobre la validez y la fiabilidad son: Humphrey y Curtis (1991) y Ropponen y Lyytinen (2000).

2.4 EJEMPLO DE ENCUESTA

A continuación, y a modo de ejemplo, se presenta una encuesta publicada en Carrillo de Gea *et al.* (2012) y que se realizó con el objetivo de conseguir información acerca de en qué modo y hasta qué punto las herramientas *software* destinadas a la Ingeniería de Requisitos (IR) dan soporte al propio proceso de la IR a través de sus distintas capacidades. Con el fin de facilitar la lectura, en los siguientes apartados se adaptará el contenido del artículo a las distintas actividades del proceso que se ha presentado en este capítulo.

2.4.1 Establecer los objetivos de la encuesta

El principal objetivo del trabajo es arrojar luz en el estado del arte relativo a las capacidades de las herramientas de la IR. Más formalmente se define el objetivo usando la plantilla GQM (*Goal-Question-Metric*) (Basili y Rombach, 1988), que se muestra en la Tabla 2.1.

Objeto de estudio	Herramientas de IR
Propósito	Caracterizar las principales características de las herramientas de IR y evaluar el escenario actual de las herramientas de IR.
Enfoque de calidad	Efectividad, coste, y presencia de las herramientas de IR.
Perspectiva	Investigadores y compradores de herramientas de IR.
Contexto	Este estudio se realiza con fabricantes de herramientas de IR y herramientas que aparecieron en alguna de las base de datos consideradas.

Tabla 2.1. Objetivo de la encuesta

Para conseguir el objetivo, en primer lugar se creó un marco de clasificación consistente en 146 *items* basándose en el informe técnico ISO/IEC TR24766 (ISO, 2009). Dicho informe presenta las características o capacidades recomendables que debería tener cualquier herramienta de soporte a la IR. El marco de clasificación propuesto consta de 8 categorías, cada una varios *items* o grupos (ver ejemplo en la Tabla 2.2).

Categoría	Grupo de capacidades
Elicitación de requisitos	Captura; Listas de comprobación, plantillas; Importación y exportación desde y hacia otras fuentes; Documentación de la elicitación.
Análisis de requisitos	Análisis de requisitos de calidad; Análisis de viabilidad; Análisis de atributos; Análisis y gestión de riesgos.
Especificación de requisitos	Documentación de la especificación de requisitos.
Modelado de requisitos	Análisis del modelado; Lenguajes de modelado y de especificación.
Verificación y validación (V&V) de requisitos	V&V.
Gestión de requisitos	Línea base de requisitos; Gestión de cambios en los requisitos; Gestión de proyectos; Modelo de datos abierto o cerrado.
Trazabilidad de requisitos	Trazabilidad; Flexibilidad en las trazas; Trazabilidad bidireccional; Análisis de la trazabilidad.
Otras capacidades	Información de gestión de las herramientas de requisitos; Interfaz de usuario gráfica; Integración de datos.

Tabla 2.2. Ejemplo del marco de clasificación

A partir de estas categorías, los autores establecen como hipótesis de trabajo que las herramientas de la IR actuales dan un soporte adecuado a:

- H₁: Elicitación de requisitos.
- H₂: Análisis de requisitos.
- H₃: Especificación de requisitos.
- H₄: Modelado de requisitos.
- H₅: Verificación y validación (V&V) de requisitos.
- H₆: Gestión de requisitos.
- H₇: Trazabilidad de requisitos.
- H₈: Otras capacidades.
- H₉: Todas las características anteriores.

Se establece además, que cada una de las características será tratada como una variable en la encuesta. Para calcular el valor de cada variable se tomará el porcentaje de características de cada categoría que presenta cada herramienta estudiada, en función de las respuestas afirmativas que los participantes en la encuesta den sobre las distintas funcionalidades que incorporan las herramientas. Se calculará también un valor global para cada herramienta que englobará todas las categorías establecidas.

Además, se introducen otras 2 variables para cada herramienta: el coste individual por licencia y las licencias en uso.

2.4.2 Diseñar la encuesta

En cuanto al diseño de la encuesta, se trata de una encuesta transversal, realizada on-line en un período de 2 meses.

Se establecen una serie de valores numéricos que se utilizan para la evaluación de los datos recogidos a través de la encuesta.

En primer lugar, para cada categoría (c) y cada herramienta (t) determinada se establece si se tiene en cuenta o no (si es un participante) siempre y cuando se cumpla el criterio que se muestra en la siguiente ecuación:

$$participante(t:c) = \begin{cases} \text{verdadero: } NA(t:c) \geq NQ(c)/2 \\ \text{falso: en caso contrario} \end{cases}$$

Es decir, que el número de respuestas dadas superaba el 50% de preguntas relativas a una determinada categoría.

Si se consideraba que una herramienta (t) participaba en los cálculos relativos a una determinada categoría (c), se calculaba su puntuación según la fórmula:

$$puntuación(t:c) = \frac{\sum_{q=1}^{NQ(c)} puntuación(t:q)}{NQ(c)}$$

Esta puntuación obtenía un valor en porcentaje sobre el grado de cumplimiento de cada herramienta en cada categoría establecida. Los valores de las puntuaciones obtenidas para cada herramienta en cada categoría se discretizaron según los siguientes intervalos:

$$discretizar(s) = \begin{cases} \text{muy alto,} & s \in (0,875, 1] \\ \text{alto,} & s \in (0,625, 0,875] \\ \text{medio,} & s \in (0,375, 0,625] \\ \text{bajo,} & s \in (0,125, 0,375] \\ \text{muy bajo,} & s \in (0, 0,125] \end{cases}$$

A su vez, para calcular la puntuación global de una herramienta en una determinada categoría se estableció la siguiente fórmula:

$$puntuación(c) = \frac{\sum_{t=1}^{NP(c)} puntuación(t:c)}{NP(c)}$$

El valor $NP(c)$ corresponde al número de participantes en la categoría c.

Con el fin de validar las hipótesis del estudio, se estableció la veracidad de las mismas siempre que la puntuación obtenida superara al 70% de respuestas positivas en esa categoría.

Para cuantificar el precio en dólares de cada licencia (p) y el número de licencias (l) de cada herramienta se utilizaron las siguientes fórmulas:

$$\text{cuantificar}(p) = \begin{cases} 0:875: & p > 1000 \\ 0:625: & p \in (501:1000] \\ 0:375: & p \in [100:500] \\ 0:125: & p < 100 \\ \text{sin respuesta:} & p \text{ sin respuesta} \end{cases}$$

$$\text{cuantificar}(l) = \begin{cases} 0:875: & l > 10000 \\ 0:625: & l \in (1001:10000] \\ 0:375: & l \in (101:1000] \\ 0:125: & l < 100 \\ \text{sin respuesta:} & l \text{ sin respuesta} \end{cases}$$

2.4.3 Desarrollar el cuestionario

Ya se comentó la importancia de contextualizar la investigación que se está desarrollando a través de un estudio bibliográfico de aquellos temas relacionados con ella. En este artículo, este estudio se centra en tres puntos principales, estudiando los trabajos relevantes ya publicados sobre cada uno de ellos: herramientas existentes de la IR, marcos de trabajo para la comparación de herramientas de la IR y, finalmente, encuestas sobre herramientas de la IR.

El cuestionario de la encuesta contenía un total de 146 preguntas, 126 recogían información técnica de las funcionalidades de cada herramienta y las 20 restantes información básica de gestión como el nombre de la herramienta, el fabricante, la versión actual, etc. La gran mayoría de las preguntas era de respuesta cerrada, aunque se incorporaron algunas de respuesta abierta para que los participantes pudieran, por ejemplo, añadir alguna funcionalidad de la herramienta que no se hubiera recogido en la lista de preguntas. A modo de ejemplo, se presentan varias preguntas en la Tabla 2.3, agrupadas por categoría. Todas estas preguntas eran cerradas y admitían como posibles respuestas "Sí", "No", "No lo sé" o "Sin respuesta". El cuestionario está disponible en <http://www.um.es/giisw/EN/re-tools-survey/part1.pdf>.

Categoría	Grupo de capacidades	Pregunta
Elicitación	Captura de requisitos	¿Da la herramienta soporte a la captura de requisitos permitiendo que el usuario almacene y gestione la documentación de las entrevistas, talleres y otras observaciones?
Análisis	Análisis de atributos	¿Da la herramienta soporte al rastreo de cualquier atributo definido por el usuario o proporcionado por la herramienta a través de la detección y la marca de atributos ausentes o perdidos?
Especificación	Documentación de la especificación de requisitos	¿Da la herramienta soporte a la documentación de la especificación de requisitos manteniendo una relación de requisitos a través de los riesgos que hayan surgido y los que se hayan mitigado?
Modelado	Análisis del modelado	¿Da la herramienta soporte al análisis del modelado evaluando los requisitos basados en objetivos de negocio?
Verificación y validación	Verificación y validación	¿Da la herramienta soporte a la verificación y validación generando informes de excepción en los requisitos que no tengan casos de plan de verificación y casos de plan de verificación que no estén vinculados a los requisitos?
Gestión	Gestión de proyectos	¿Da la herramienta soporte a la gestión de proyectos registrando, monitorizando e informando sobre el estado del proceso general de gestión de requisitos?
Trazabilidad	Trazabilidad	¿Da la herramienta soporte a la trazabilidad a través del mantenimiento automático de trazas?
Otras capacidades	Interfaz gráfica de usuario	¿Da la herramienta soporte a interfaces gráficas de usuario permitiendo el uso de navegadores web?

Tabla 2.3. Ejemplos de preguntas

En el artículo se explicita el problema detectado al haber, en algunos casos, respuestas del estilo "no lo sé" o "sin respuesta". La solución que se adoptó consistía en tener en cuenta únicamente una categoría, si se habían respondido (bien positiva o negativamente) al menos al 50% de las preguntas relativas a esa categoría.

2.4.4 Evaluar y validar el cuestionario

Los autores elaboraron el cuestionario en un período de 3 meses, incluyendo 5 ciclos de revisión del mismo. Durante estas revisiones, se discutieron y acordaron la formulación de todas las preguntas hasta que se consiguió que todas fueran claras y simples.

Además, una persona del equipo de investigación se dedicó a comprobar que todos los aspectos técnicos relativos a la realización on-line de la encuesta funcionaban correctamente.

2.4.5 Obtener los datos de la encuesta

Para la gestión de los cuestionarios se utilizó la herramienta *LimeSurvey*². Para reclutar a las personas que participarían en la encuestas contactaron mediante correo electrónico con los distintos desarrolladores y/o proveedores de 100 herramientas, para invitarles a participar en la encuesta. Tras un proceso de otros 2 meses se consiguió que 38 de ellos participaran aportando los datos sobre cada una de sus herramientas, debiendo reducir el conjunto original de herramientas candidatas de 100 a 94 porque no pudieron conseguir la dirección de correo electrónico de sus representantes.

2.4.6 Analizar los datos obtenidos

Para establecer correlaciones entre variables, los autores utilizaron el coeficiente de correlación de Pearson (Ott y Longnecker, 2010) y como herramientas de cálculo estadístico y procesamiento de datos utilizaron el paquete *SPSS 19.0* y la herramienta *MS-Office Excel 2007*.

Los distintos resultados obtenidos se desgranán en los siguientes apartados, comenzando con una descripción de los participantes a través de la información de gestión recogida en la encuestas, continuando con el análisis de correlación entre las variables definidas, el contraste de hipótesis y finalizando con el procedimiento de validación de datos.

² www.limesurvey.org

2.4.6.1 PARTICIPANTES

Finalmente se consiguieron examinar un total de 38 herramientas de las 94 candidatas, lo que supone una participación del 40,4%.

Con las preguntas que recogían información de gestión de las herramientas se realizó un análisis que se resume gráficamente en la Figura 2.1.

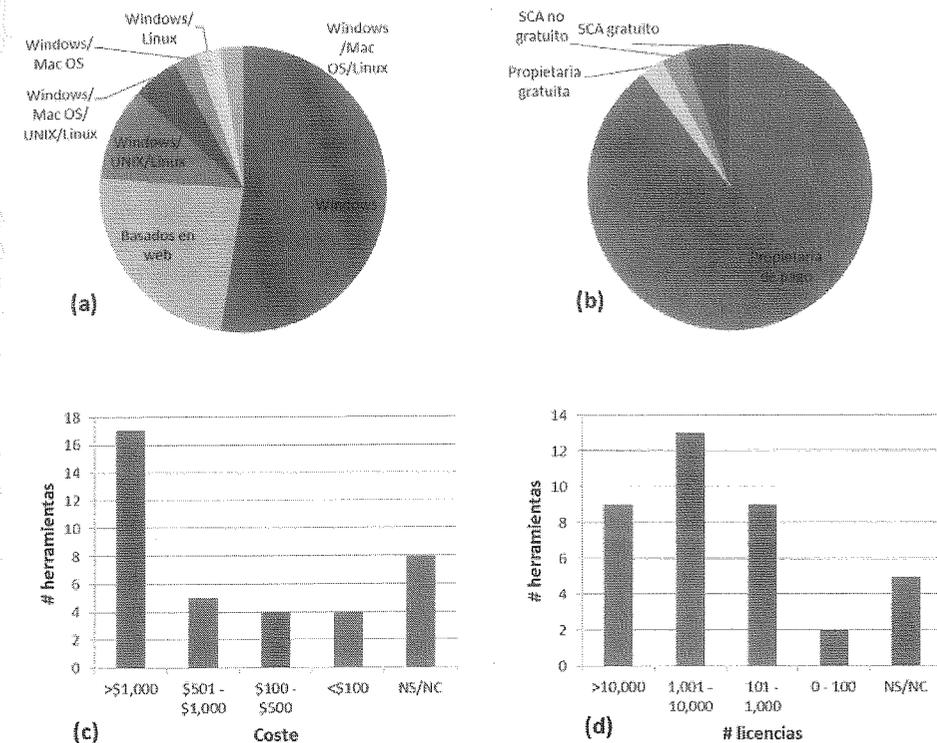


Figura 2.1. Datos relativos a las herramientas, Plataforma software requerida (a), Tipo de licencias (b), Coste individual por licencia (c) y Licencias en uso (d)

De la Figura 2.1 se puede extraer la siguiente información:

- Con respecto a la plataforma *software*, hay una gran predominancia de sistemas Windows como se puede observar en la Figura 2.1(a). Los clientes basados en la Web también son comunes, con el propósito de facilitar el acceso colaborativo a los recursos. Otros sistemas operativos como UNIX o Linux, y particularmente Mac, tienen una presencia más limitada.

- Las licencias eran mayoritariamente propietarias y no gratis (ver Figura 2.1 (b)), con poca influencia del *software* libre. Con respecto al promedio de coste por licencia (n dólares), la mayoría de las herramientas cuestan \$1.000 o más como se puede observar en la Figura 2.1(c).
- Finalmente el número de licencias en uso es aproximadamente entre 1.001 y 10.000 (ver Figura 2.1 (d), aunque hay un número considerable de herramientas más extendidas (más de 10.000 licencias) y otro grupo importante de herramientas con menos representación en el mercado (entre 101 y 1.000 licencias).

2.4.6.2 CORRELACIONES ENTRE VARIABLES

Para medir la correlación entre las variables estudiadas se utilizó un test de correlación bivariado. En la Figura 2.2 se presentan los resultados obtenidos. En esta figura cada celda contiene dos valores, el valor de *r* que indica la fortaleza y la dirección (-+) de la correlación (son mejores los valores mayores), y "*" o "***" indica que la hipótesis nula H_0 se puede rechazar, lo que significa que las variables están correlacionadas. El segundo valor es el número de pares en la muestra.

En este tipo de análisis, un valor positivo elevado representa una correlación positiva igualmente elevada entre las 2 variables involucradas. A modo de ejemplo, mencionamos la correlación de la elicitación con el análisis (0,723) o entre la elicitación y la especificación (0,800), entre otras. Esto implica que los valores de las variables correlacionadas, aumentan o disminuyen simultánea y proporcionalmente.

Si analizamos el Coste, vemos que la mayor correlación la tiene con la especificación de requisitos, aunque si observamos la correlación con la puntuación global no es significativa con es positiva (0,358).

Finalmente, comentar la correlación entre el número de licencias en uso y el resto de variables. Se puede observar en Figura 2.2 una alta correlación positiva con la categoría de otras capacidades (0,513) Y por sin embargo, no hay correlación entre el número de licencias en uso y el coste medio por licencia (0,243).

Matriz de correlación: Elicitación; Análisis; eSpecificación; Modelado; V&V; Gestión; Trazabilidad; Otras; Global; Coste por licencia; Licencias en uso; N: tamaño de la muestra

	EL	AN	SP	MO	VV	GE	TR	OT	GL	CO	LI
EL	1										
N	35										
AN	0,763***	1									
N	34	35									
SP	0,800**	0,653**	1								
N	35	35	36								
MO	0,470**	0,499**	0,629**	1							
N	33	33	34	34							
VV	0,792**	0,740**	0,641**	0,521**	1						
N	27	27	27	25	29						
GE	0,729**	0,725**	0,716**	0,592**	0,694**	1					
N	31	31	32	31	28	34					
TR	0,776**	0,618**	0,781**	0,656**	0,791**	0,737**	1				
N	31	31	32	30	29	33	34				
OT	0,777**	0,610**	0,718**	0,628**	0,678**	0,801**	0,732**	1			
N	29	29	29	28	28	31	31	31			
GL	0,903**	0,793**	0,869**	0,799**	0,830**	0,934**	0,913**	0,872**	1		
N	25	25	25	25	25	25	25	25	25		
CO	0,269	0,336*	0,545**	0,404*	0,169	0,094	0,329*	0,147	0,358	1	
N	29	29	30	28	23	27	24	21	21	30	
LI	0,185	-0,068	0,285	-0,005	-0,054	0,245	0,012	0,513**	0,183	0,243	1
N	31	31	32	30	25	30	30	27	22	29	33

* correlación significativa a nivel 0,05

** correlación significativa a nivel 0,01

Figura 2.2. Correlaciones entre variables

2.4.6.3 EVALUACIÓN DE LAS HIPÓTESIS

Para hacer el contraste de hipótesis se utilizó un análisis de los estadísticos descriptivos. La Figura 2.3 muestra las puntuaciones para cada una de las capacidades consideradas, obtenidas para las 38 herramientas cuyos fabricantes contestaron a la encuesta.

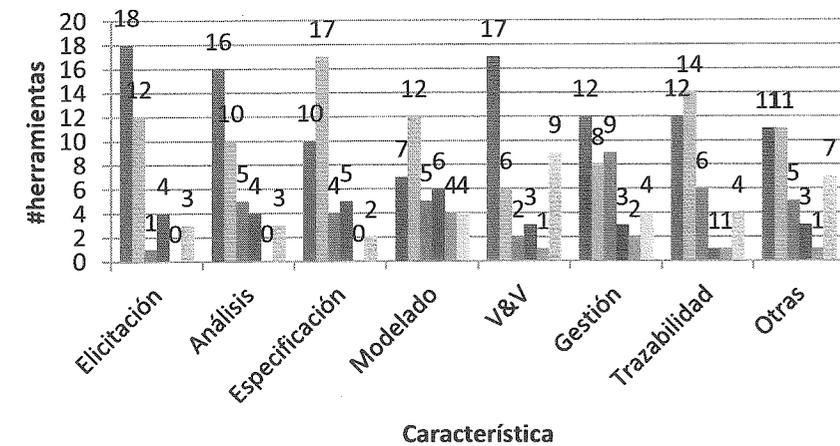


Figura 2.3. Puntuaciones obtenidas por las herramientas (valores de izquierda a derecha, muy alta, alta, media, baja, muy baja, sin participantes)

Como resumen de los resultados obtenidos en relación a las hipótesis planteadas, únicamente las hipótesis H_4 y H_6 (las relativas al modelado y la gestión de los requisitos) fueron rechazadas, por lo que el resto de hipótesis se aceptaron. Esto significa que las herramientas estudiadas sobre IR dan soporte adecuado al modelado y a la gestión de los requisitos, pero no así al resto de características estudiadas.

Por último, la Figura 2.4 muestra las puntuaciones globales obtenidas por las distintas herramientas, que muestra que las puntuaciones globales en general fueron altas (72% obtuvieron una puntuación alta o muy alta), lo que demuestra que el nivel de capacidades de las herramientas era alto, por lo que la hipótesis H_9 se pudo ser aceptadas. No obstante, los valores individuales obtenidos para cada una de las capacidades, como comentamos previamente no todos fueron favorables.

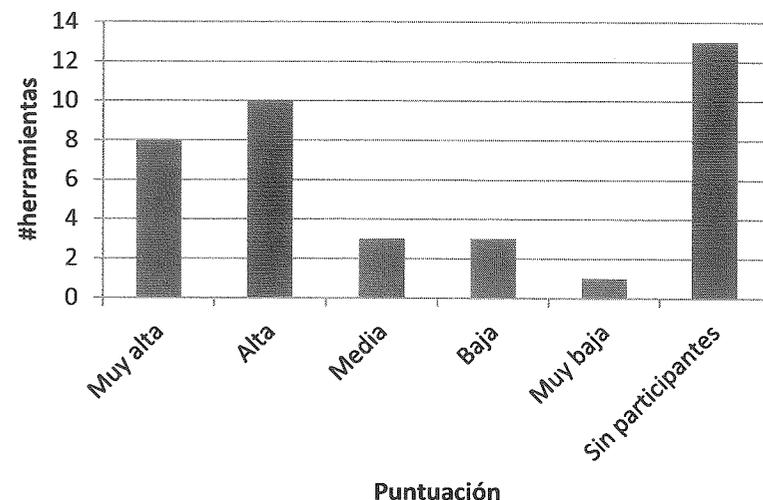


Figura 2.4. Puntuaciones globales de las herramientas

2.4.6.4 VALIDACIÓN DE LOS DATOS

Se presentan en este apartado las posibles anomalías que detectaron respecto de los datos. Principalmente exponen que la evaluación de cada herramienta se llevaba a cabo por sus fabricantes, lo que podía influir en un grado de subjetividad superior a lo normal con el fin de conseguir una puntuación positiva de la herramienta. Con el fin de mitigar este problema, los autores realizaron un estudio seleccionando de manera aleatoria cinco de las 38 herramientas para comprobar que las respuestas que habían recibido se ajustaban con la realidad. También se realizó un análisis denominado "fiabilidad entre

evaluadores" (*interrater reliability*) que permite con el propósito de determinar el grado de consistencia entre las puntuaciones y para medir la fiabilidad de los datos obtenidos a través de las repuestas de los encuestados. Tras hacer este análisis se llegó a la conclusión que las puntuaciones obtenidas eran consistentes y fiables.

2.4.7 Limitaciones del ejemplo

Como es habitual en cualquier estudio empírico se analizaron aquellos aspectos que podían amenazar la validez de los resultados obtenidos:

- **Validez interna:** Se refiere a la fiabilidad de la encuesta. La validez del material recolectado a través del cuestionario es altamente dependiente de la experiencia de los encuestados. La mayoría de los encuestados son profesionales con varios de experiencia en el uso de las herramientas de IR. El riesgo de maduración se tuvo en cuenta y por ello se diseñó un cuestionario que no requiriera más de 20 minutos para completarlo. A pesar de esto, hubo 3 encuestados que comenzaron a completar el cuestionario y lo abandonaron. Otro aspecto que se intentó mitigar es el sesgo que puede producirse por el hecho de que las herramientas fueran evaluadas por sus propios fabricantes, para ello se tuvo especial cuidado en el diseño del cuestionario y además se hizo un análisis de la fiabilidad entre evaluadores y se encontró un alto grado de acuerdo.
- **Validez externa:** Se refiere a la generalización de los resultados a entornos industriales. La selección de las herramientas se hizo en listas conocidas, evitando así considerar herramientas poco conocidas. Y los encuestados eran sus fabricantes, con lo que consideramos que tanto las herramientas como los encuestados se consideran adecuados. Sin embargo hay algunas herramientas relevantes, que no se pudieron incluir, como es el caso de Enterprise Architect, aunque se intentó en reiteradas ocasiones contactar con sus fabricantes. Además se consideraron capacidades de las herramientas escogidas de un marco específico diseñado por expertos para la evaluación de herramientas de IR y que se adecuo para considerar algunos aspectos específicos necesarios en la industria (por ejemplo, modelos de datos abiertos, integración de datos, etc.). Sin embargo, puede que nos sean las capacidades deseables por los usuarios de las herramientas. Los autores de este estudio creen que la evidencia obtenida se puede generalizar a entornos de IR industriales específicos y los resultados de este estudio pueden utilizarse por los ingenieros de requisitos, teniendo en cuenta que la tecnología evoluciona y que pueden surgir nuevas herramientas y que las herramientas existentes pueden incorporar nuevas capacidades.

- **Validez de la conclusiones:** El tamaño de la muestra (38 herramientas) es pequeño para producir un poder estadístico aceptable y es innegable que existan herramientas que no se consideraron. Por ello, no es recomendable considerar como definitivos los resultados obtenidos. Es aconsejable continuar este estudio con una mayor cantidad de herramientas. No obstante, el número de participantes representa un porcentaje relevante de la comunidad de fabricantes de herramientas de IR, incluyendo herramientas de empresas establecidas en tres continentes (Asia, América y Europa). Además el estudio realizó siguiendo un proceso y se publicó con un alto nivel de detalle para permitir que el proceso sea reproducible. Aunque puede ocurrir que en el futuro que el número de herramientas obtenido en las búsquedas varíe en el futuro.

2.4.8 Conclusiones del ejemplo

En el ejemplo presentado se realizó un estudio bibliográfico sobre las herramientas de soporte a la IR. También se creó un marco de clasificación, basado en el informe técnico ISO/IEC TR 24766 para establecer las características deseables de estas herramientas. Además se llevó a cabo una encuesta con el objetivo de analizar hasta qué punto las herramientas existentes para la IR cubren las capacidades deseables. Los hallazgos obtenidos en esta encuesta pueden ser útiles para investigadores porque les proporcionan una idea del estado del arte sobre herramientas de IR y también para los profesionales, ya que les ayudará a ser conscientes de las características más frecuentes que contienen las herramientas de IR existentes y a conocer su coste, el número de licencias en uso, etc. Esta información puede ser de gran interés al momento de decidir qué herramienta IR se debe utilizar en sus organizaciones.

Entre otros, los principales hallazgos obtenidos son:

- Respecto de los resultados obtenidos en el análisis de correlación, demuestran que hay una fuerte correlación entre muchas de las características deseables en las herramientas, lo que lleva a pensar que si una herramienta se comporta de manera deseable con respecto a una de las características, con mucha probabilidad lo hará también con el resto.
- La característica que mejor se trata, en general, es la elicitación de requisitos y en concreto la captura de requisitos, aunque existe deficiencia en el soporte de la herramientas a las plantillas y lista de comprobación utilizadas en la elicitación de requisitos. El análisis, la especificación, la trazabilidad, la V&V de los requisitos también se tratan adecuadamente.

- Por último, indicar que las herramientas estudiadas dan peor soporte a los aspectos relacionados con el modelado, la gestión de requisitos y otras capacidades. Esto indica que estas capacidades deberían ser mejoradas en las herramientas de requisitos.
- Las puntuaciones obtenidas para el modelado fueron menores que las obtenidas en la especificación y otras categorías. Esto se puede deber a que las herramientas de requisitos han estado tradicionalmente más orientadas a requisitos textuales, en lenguaje natural en comparación con notaciones de modelado como BPMN, UML, E/R.
- Con respecto a la categoría de gestión de requisitos, se detectó la falta de mecanismos para modelo de datos abierto. Al parecer, las herramientas de IR actuales no dan soporte tales características, a pesar de que le proporcionan tanto a los desarrolladores como los usuarios muchos beneficios importantes: apoyo para aumentar la comunicación y la automatización, una amplia personalización para el usuario final, secuencias de comandos y funciones de macro, agentes inteligentes externos y tutores, comandos potentes para buscar y reemplazar, buscar y reemplazar comandos, fácil suministro de correctores ortográficos conocidos, marcas semánticas, interfaces alternativas sin reimplementación, la capacidad de tener los *plug-ins* que operan en el mismo espacio, y una significativamente mayor reutilización de código común para los implementadores. Algunos de los participantes que han mencionado la posibilidad de dar soporte a modelos de datos abiertos han señalado que se puede hacer por medio de APIs específicas.
- A veces, sobre todo en las grandes organizaciones, existen diversas fuentes de datos que contienen datos críticos de la empresa. Por otra parte, la gestión de estos datos dispersos depende de sistemas diferentes. Esta diversidad de fuentes de datos es causada por muchos factores que se encuentran típicamente en proyectos de Desarrollo Global de *Software*, como la falta de coordinación entre las diferentes partes de la organización, las diferentes tasas de adopción de tecnologías nuevas, fusiones y adquisiciones, y la distancia geográfica entre los grupos de colaboradores. Por lo tanto, las herramientas de IR deben ofrecer mecanismos con los que combinar la información de estos sistemas diversos, sobre todo en ambientes de trabajo distribuidos.

2.5 OTROS EJEMPLOS DE ENCUESTAS

Entre otros ejemplos de encuestas que pueden encontrarse en la literatura, caben destacar los siguientes: Grossman *et al.* (2005), Conradi *et al.* (2005), Jørgensen y Moløkken-Østvold (2006), Rodríguez *et al.* (2012) y Vizcaíno *et al.* (2013).

2.6 LECTURAS RECOMENDADAS

- **Fink, A.** (2003). *The survey kit*. SAGE Publications. Se trata de una colección de libros que pretende ser una enciclopedia sobre la realización de encuestas en general. Resulta interesante al estar dividido en distintos libros específicos para distintos aspectos relativos a las encuestas, si bien no se enmarca específicamente en el campo de la ingeniería del *software*.
- **Ciolkowski, M., Laitenberger, O., Vegas, S. y Biffel, S.** (2003). *Practical experiences in the design and conduct of surveys in empirical software engineering*. ESERNET 2001-2003, LNCS 2765, pp. 104-128. En este trabajo se presenta un proceso para preparar, realizar y analizar una encuesta, basada en cuestionarios, específicamente en campo de la ingeniería del *software* con el fin de proporcionar a los investigadores del área una aproximación sistemática y disciplinada a la hora de realizar encuestas y detectar y evitar obstáculos a la hora de realizar encuestas y conseguir resultados interesantes.
- **Kitchenham, B. A. y Pfleeger, S. L.** (2008). *Personal opinion surveys*. Chapter 3 in: Shull, F., Singer, J., Sjøberg, D.I.K. (eds.) *Guide to Advanced Empirical Software Engineering*: Springer. Capítulo de referencia en la actualidad en la comunidad acerca del diseño y realización de encuestas. Este capítulo tiene como origen la serie de artículos que los autores publicaron en la revista *Software Engineering Notes* entre 2001 y 2003, llamados *Principles of Survey Research*.

2.7 HERRAMIENTAS Y SITIOS WEB RECOMENDADOS

Existen multitud de herramientas y páginas web que permiten realizar encuestas *online*. Entre las más utilizadas y que ofrecen sus servicios de forma gratuita, cabe destacar:

- <http://www.limesurvey.org/>
- <https://www.murvey.com/>
- <http://kwiksurveys.com/>
- <http://freeonlinesurveys.com/>

Algunas otras herramientas ofrecen funcionalidades limitadas de manera gratuita, pero es necesario adquirir una licencia para acceder a todas las posibilidades que ofrecen. Entre ellas destacan:

- <http://es.surveymonkey.com>
- <http://www.zoomerang.com/>
- <http://www.surveygizmo.com/>

Por último, indicar que cualquier paquete estadístico que se utilice normalmente (R, SPSS, etc.) será más que suficiente para realizar el análisis de los datos obtenidos durante la realización de una encuesta.