

Medición y Evaluación de Calidad en Uso de Aplicaciones Web

Guillermo Juan COVELLA

*Grupo de Investigación y Desarrollo en Ingeniería de Software e Ingeniería
Web (GIDISWeb), Facultad de Ingeniería de la UNLPam,
Calle 9 y 110, (6360) General Pico, La Pampa, Argentina
covellag@ing.unlpam.edu.ar*

Presentada a la Facultad de Informática de la UNLP como
parte de los requisitos para la obtención del título de
Magíster en Ingeniería de Software

Director : *Dr.* Luis **OLSINA SANTOS**

Co-Directora : *Ing.* Marisa **de GIUSTI**

La Plata, Noviembre de 2005

Facultad de Informática
Universidad Nacional de La Plata – Argentina

Agradecimientos

A Valeria Bonessi, mi esposa. Compañera leal de todos mis esfuerzos. Por su amor incondicional, su andar incansable cubriendo mis ausencias, y también por su paciencia para corregir la ortografía y la gramática de los originales de varios capítulos de esta tesis.

A Carola y Alena, por el cariño, los besos y la eterna pregunta de respuesta elusiva “¿Papi, cuando vas a terminar ese trabajo?”, que me sirvieron de estímulo en los momentos difíciles.

A Luis Olsina, Director de esta tesis, pero también compañero de estudio y trabajo desde hace casi veinte años, por su permanente aliento y colaboración, y los aportes imprescindibles para llevar adelante todo el proceso de elaboración de la tesis.

A Marisa De Giusti, codirectora de esta tesis, por su predisposición favorable y sus oportunas contribuciones.

A mis compañeros del grupo GidisWeb, de la Facultad de Ingeniería. Especialmente a Peti Martín, Fernanda Papa y Hernán Molina, por sus aportes desinteresados y buena predisposición cuando los consulté y necesité de su colaboración.

A Luis Mendiburu y su familia , que me recibieron afectuosamente en su casa, como a uno más de sus numerosos amigos, cada vez que viajé a La Plata para asistir a las clases del master.

A Willy Bonessi, que colaboró con el diseño de arte de este trabajo

A la Facultad de Ingeniería de la Universidad Nacional de La Pampa, que financió en gran parte los gastos indispensables para el cursado de la tesis y facilitó las instalaciones para la realización de los estudios de campo y la escritura de este trabajo.

A mis compañeros de la Dirección de Computación de la Municipalidad de General Pico, La Pampa, por su colaboración, paciencia y esfuerzo solidario cuando debieron reemplazarme a causa de mis viajes por estudio.

Esta tesis se desarrolló en el contexto de las actividades de los proyectos: *“Métricas, Modelos y Herramientas para Evaluar y Controlar Calidad y Recuperar Información en Sistemas Web”*. Secretaría de Ciencia y Técnica, Facultad de Ingeniería, UNLPam - Proyecto 09-F022 del Programa de Incentivos de la Secretaría de Políticas Universitarias del Ministerio de Cultura y Educación de la Nación, Argentina entre el 01/05/2001 y 30/05/2005; y *“Modelos y Tecnologías para el Soporte de Proyectos de Medición y Evaluación en Aseguramiento de Calidad”*, desde el 01/05/2005.

Resumen

El empleo masivo de productos software, impulsado en el último lustro por el crecimiento de Internet y la WWW, ha puesto a los usuarios finales como referentes potenciales para la evaluación de calidad de productos software. La industria respondió a la demanda de un mercado global en crecimiento, pero esa respuesta no siempre tuvo en cuenta las exigencias de calidad y calidad en uso de usuarios cada vez más numerosos y mejor formados.

Referencias importantes en la discusión y la propuesta de esta tesis son varios de los estándares de la organización ISO/IEC. Específicamente en relación a calidad en uso, se consideraron el estándar 9126-1 y el borrador del estándar 9126-4.

En esta tesis se propone un enfoque ingenieril para medir y evaluar la calidad en uso, o sea la calidad percibida por los usuarios en contextos reales de uso, de productos software para la Web. Partiendo de estándares ISO/IEC, en particular el estándar 9126-1 y el borrador del estándar 9126-4, la propuesta, de carácter sistemático y disciplinado, está encuadrada por un marco de medición y evaluación, basado a su vez en una ontología de métricas e indicadores.

El marco ofrece precisiones en las definiciones necesarias para la medición y evaluación – por ejemplo: necesidad de información, modelo de concepto, concepto calculable, métrica e indicador - y líneas guía para las actividades que deben llevar a cabo usuarios y evaluadores.

La ejecución de una instancia del marco de medición se plasmó en un caso de estudio. Para llevar adelante el caso de estudio se empleó la metodología WebQEM, diseñada para evaluación de calidad de productos Web, cuyas etapas tienen relación directa con el marco de medición y evaluación. Un reto importante fue adaptar la metodología para evaluar calidad en uso.

El caso de estudio se realizó con la participación de usuarios reales de una aplicación para e-Learning. El diseño de las métricas e indicadores empleados en el caso de estudio constituye uno de los aporte significativos de esta tesis ya que, si bien los resultados de este tipo de evaluaciones no son generalizables, se generaron metadatos provenientes de definiciones exhaustivas que pueden constituir líneas base para proyectos similares.

Como líneas de trabajo futuro se visualiza la necesidad de contar con instrumentos específicos para afrontar la evaluación de los aspectos subjetivos, particularmente satisfacción, y considerar, para tener un panorama completo de la calidad en uso de la aplicación Web evaluada, el punto de vista de otros perfiles de usuario involucrados.

Tabla de Contenidos

| | |
|--|-----|
| Índice de Tablas | III |
| Índice de Figuras..... | III |
| Parte I. Introducción y Motivación | 1 |
| Capítulo 1: Motivación de la Tesis | 1 |
| 1.1. Acerca del Problema Abordado | 1 |
| 1.2. Objetivos Básicos de la Tesis | 3 |
| 1.3. Organización de la Tesis..... | 4 |
| Capítulo 2. Calidad y Calidad en Uso de Aplicaciones Software y Web: | |
| Estado del Arte | 5 |
| Introducción | 5 |
| 2.1 Calidad de Producto Software | 5 |
| 2.2. Calidad Interna y Externa..... | 7 |
| 2.3. Calidad en Uso: Estado del Arte | 9 |
| 2.3.1. Calidad en Uso en la Literatura y en la Práctica | 10 |
| 2.3.2. Calidad y Calidad en Uso | 10 |
| 2.3.3. Usabilidad y Calidad en Uso..... | 12 |
| 2.3.4. Definición y Modelo de Calidad en Uso para Producto Software | |
| | 14 |
| 2.3.5. Criterios para Medir y Evaluar Calidad en Uso..... | 16 |
| 2.4. Calidad de Aplicaciones Web | 22 |
| 2.4.1. Características Distintivas de las Aplicaciones Web | 23 |
| 2.4.2. Perspectivas de la Calidad de Aplicaciones Web | 24 |
| 2.5. Conclusiones | 26 |
| Parte II: Marco Conceptual y Método de Medición y Evaluación | 29 |
| Capítulo 3: Un Marco Conceptual de Medición y Evaluación Para Calidad | |
| en Uso..... | 31 |
| 3.1 Introducción | 31 |
| 3.2 Ontología para el Modelo de Métricas e Indicadores..... | 32 |
| 3.3 Marco de Medición y Evaluación | 33 |
| 3.3.1 Conceptos | 34 |
| 3.3.2 Ejemplos de los Principales Términos para Calidad en Uso | 39 |
| 3.4 Conclusiones. | 41 |
| Capítulo 4: El Método de Evaluación..... | 43 |
| 4.1. Introducción | 43 |
| 4.2. Panorama de la Metodología WebQEM..... | 44 |
| 4.2.1 Pasos Básicos de WebQEM | 46 |
| 4.3. Adaptando WebQEM para Calidad en Uso..... | 51 |
| 4.4. Consideraciones Finales | 52 |
| Parte III: Un Caso de Estudio de Calidad en Uso Para Aplicaciones Web..... | 55 |
| Capítulo 5: Caso de Estudio de Calidad en Uso de una Aplicación para el | |
| Dominio de e-Learning..... | 57 |
| 5.1. Introducción | 57 |
| 5.2. El Dominio de Aplicaciones Web para e-Learning | 58 |
| 5.3 Planificar y Programar la Evaluación de Calidad en Uso | 60 |
| 5.3.1 Objetivos del Caso de Estudio..... | 60 |
| 5.3.2 Perfil de la Audiencia | 60 |
| 5.3.3 Tareas para Evaluar Calidad en Uso | 62 |
| 5.4 Descripción del Contexto | 63 |

| | |
|--|-----|
| 5.4.1 Información del Producto | 63 |
| 5.4.2 Requerimientos del Producto | 63 |
| 5.4.3 Instalaciones | 64 |
| 5.4.4 Herramientas Utilizadas por los Evaluadores..... | 64 |
| 5.5 Definir y Especificar los Requerimientos de Calidad | 65 |
| 5.6 Diseñar la Evaluación Elemental | 66 |
| 5.6.1 Métricas de Calidad en Uso..... | 66 |
| 5.6.2. Métricas e Indicadores para el Caso de Estudio..... | 69 |
| 5.7 Definir la Evaluación Global..... | 75 |
| 5.8. Definición de Modelos de Agregación e Implementación de la Evaluación Global | 76 |
| 5.9 Conclusiones | 77 |
| Capítulo 6: Ejecución del Caso de Estudio | 79 |
| 6.1.Introducción | 79 |
| 6.2.1 Recolección de los Datos: Procedimiento y Herramienta | 80 |
| 6.2.2. Cómputo de las Métricas | 80 |
| 6.2.3. Cómputo de los Indicadores Elementales | 86 |
| 6.3.Implementar la Evaluación Global..... | 89 |
| 6.4. Analizar los Resultados, Elaborar las Conclusiones y Documentar | 90 |
| 6.4.1. Análisis de Resultados y Recomendaciones | 91 |
| 6.5. Conclusiones | 92 |
| Capítulo 7: Consideraciones Finales y Trabajos Futuros | 95 |
| 7.1 Consideraciones Finales | 95 |
| 7.2 Trabajos Futuros..... | 98 |
| 7.2.1. Trabajos Futuros relacionados al Dominio del Producto Evaluado | 98 |
| 7.2.2. Trabajos Futuros en relación al Enfoque para Medir y Evaluar de Calidad en Uso | 99 |
| Referencias | 101 |
| Apéndice A: Métricas del Caso de Estudio | 107 |
| A.1. Concepto Calculable (Característica): Eficacia | 107 |
| A.2. Concepto Calculable (Característica): <i>Productividad</i> | 111 |
| Tipo de Escala: intervalo. | 113 |
| A.3. Concepto Calculable (Característica): <i>Satisfacción</i> | 114 |
| Apéndice B: Ontología de Métricas e Indicadores..... | 117 |
| B.1. Introducción | 117 |
| B.2 Definición de Conceptos..... | 117 |
| B.3. Modelos UML para los conceptos de la Ontología de Métricas e Indicadores..... | 118 |
| B.4. Descripción de Atributos y Relaciones | 122 |
| Apéndice C: Diseño y definición de las tareas | 125 |
| C.1. Definición de las Tareas para el Caso de Estudio..... | 126 |
| C.2. Criterios para Evaluar Correctitud y Completitud de las Tareas .. | 127 |
| C.3. Imágenes de la Realización del Caso de Estudio | 129 |
| Apéndice D: El Caso de Estudio Presentado a los Usuarios y el Cuestionario de Satisfacción..... | 131 |
| D.1. Instrucciones para los Participantes | 131 |
| D.2. Cuestionario sobre Formación, Preferencia, Edad y Género | 133 |
| D.3. Cuestionario de Satisfacción..... | 133 |
| D.4. Utilización y Cálculo del Cuestionario | 135 |

| | |
|---|-----|
| D.4.1. Utilización del Cuestionario | 135 |
| D.4.2. Cálculo del Cuestionario | 135 |

Índice de Tablas

| | |
|---|-----|
| Tabla 2.1. Definición de las subcaracterísticas de Usabilidad prescrita en el estándar ISO 9126-1, para calidad interna y externa. | 9 |
| Tabla 2.2. <i>Definición de las cuatro características de Calidad en Uso prescritas en la ISO/IEC 9126-1.</i> | 15 |
| Tabla 6.1. Resultados obtenidos por los usuarios para la métrica <i>Número de Tareas Completadas Correctamente.</i> | 81 |
| Tabla 6.2. Resultados obtenidos por los usuarios para la métrica <i>Proporción de Tareas Completadas sobre Tareas Propuestas.</i> | 81 |
| Tabla 6.3. Resultados obtenidos por los usuarios para la métrica <i>Proporción de Tarea Completada Correctamente.</i> | 82 |
| Tabla 6.4. Resultados obtenidos por los usuarios para la métrica <i>Promedio de la Proporción de Tareas Completadas por un Usuario</i> | 82 |
| Tabla 6.5. Resultados obtenidos por los usuarios para la métrica <i>Proporción de Tareas Completadas sobre Tareas Propuestas.</i> | 83 |
| Tabla 6.6. Resultados obtenidos por los usuarios para la métrica <i>Tiempo Total de Tareas Completadas</i> | 83 |
| Tabla 6.7. Resultados obtenidos por los usuarios para la métrica <i>Proporción de Eficiencia de Tareas Completadas.</i> | 84 |
| Tabla 6.8. Resultados obtenidos por los usuarios para la métrica <i>Tiempo de Tarea.</i> | 85 |
| Tabla 6.9. Resultados obtenidos por los usuarios para la métrica <i>Proporción de Eficiencia de Tarea considerando Eficacia.</i> | 85 |
| Tabla 6.10. Resultados obtenidos por los usuarios para la métrica <i>Promedio de Eficiencia de Tareas para un Usuario considerando Eficacia</i> | 86 |
| Tabla 6.11. Resultados de los indicadores elementales, parciales y del indicador global de Calidad en Uso. | 90 |
| Tabla B.1. Definición de Conceptos..... | 117 |
| Tabla B.2. Descripción de Atributos de la Ontología de Métricas e Indicadores..... | 122 |
| Tabla B.3. Descripción de las Relaciones incluidas en la Ontología de Métricas e Indicadores. | 123 |
| Tabla C.1. Aspectos relevantes para la definición de Tareas. | 126 |

Índice de Figuras

| | |
|---|----|
| Figura 2.1. <i>Usabilidad en el estándar 9241-11.</i> | 14 |
| Figura 2.2. <i>Modelo para calidad en uso. ISO IEC 9126-1 (2001).</i> | 15 |
| Figura 2.3. <i>Marco conceptual de calidad teniendo en cuenta diferentes tipos de entidades y potenciales modelos de calidad, de interés para Ingeniería de Software e Ingeniería Web</i> | 16 |
| Figura 2.4. <i>Especificación de una instancia del modelo de Calidad en Uso.</i> | 22 |
| Figura 3.1. <i>Conceptos involucrados en la fase de Definición y Especificación de Requerimientos.</i> | 35 |
| Figura 3.2. <i>Conceptos involucrados en la fase de Diseño e Implementación de la Medición.</i> | 37 |
| Figura 3.3. <i>Conceptos involucrados en la fase de Diseño e Implementación de la Evaluación</i> | 39 |

| | |
|--|-----|
| Figura 4.1. <i>Clasificación de Visitantes de Sitios Web.</i> | 47 |
| Figura 4.2. <i>Subcaracterística Funcionalidad y Contenidos Específicos para sitios y aplicaciones con funcionalidad e-Learning.</i> | 48 |
| Figura 4.3. <i>Sección del árbol de requerimientos para ejemplificar el cálculo de indicadores parciales.</i> | 50 |
| Figura 5.1. <i>Árbol de Requerimientos para el caso de estudio de calidad en uso.</i> | 66 |
| Figura 5.2. <i>Asignación de pesos a subconceptos y atributos.</i> | 76 |
| Figura 5.3. <i>Operadores lógicos en la estructura de agregación para la evaluación global.</i> | 77 |
| Figura B.1. <i>Conceptos involucrados en la fase de Definición y Especificación de Requerimientos.</i> | 119 |
| Figura B.2. <i>Conceptos involucrados en la fase de Diseño de la Medición e Implementación.</i> | 120 |
| Figura B.3. <i>Conceptos involucrados en la fase de Diseño de la Evaluación e Implementación.</i> | 121 |
| Figura C.1. <i>Uno de los usuarios en plena sesión de trabajo, filmado con una cámara digital.</i> | 129 |
| Figura C.2. <i>Sesión de trabajo abierta, seguida desde una notebook.</i> | 129 |

Parte I. Introducción y Motivación

Capítulo 1: Motivación de la Tesis

1.1. Acerca del Problema Abordado

El empleo masivo de productos software a escala global, prácticamente en todos los ámbitos del desempeño humano, favorecido en el último lustro por la enorme tasa de crecimiento de los servicios relacionados a la red Internet, han convertido a los usuarios en potenciales referentes para evaluar la calidad de productos de software.

Además, una extraordinaria demanda de aplicaciones para la Web dio otra vuelta de tuerca al desarrollo de software. Este impulso involucró la presencia de nuevos actores, dadas ciertas particularidades que caracterizan al desarrollo para la Web y que en cierto modo también lo diferencian del desarrollo de productos software tradicionales.

La respuesta a esta demanda, por parte de la industria y la comunidad de desarrolladores fue importante en tecnología y escala de producción, pero sin embargo no es extraño encontrar hoy, a trece años de la aparición del primer browser gráfico para la WWW (NCSA-Mosaic), que las cuestiones relacionadas a la calidad –especialmente la calidad que perciben los usuarios reales respecto a las aplicaciones software que emplean- han sido postergadas o directamente ignoradas, entre otras razones por las urgencias de entregar los productos al mercado, eventuales ahorros o falta de recursos humanos capacitados, tal como reseñan [BAS01].

Justamente esta cuestión, más una cierta falta de precisión en la definición de términos clave empleados para medir y evaluar alguna de las perspectivas de la calidad de producto software (en particular de aplicaciones Web), constituyen una motivación importante para los estudios, discusiones y aportes incluidos en esta tesis.

Esta situación no es extraña en áreas y disciplinas de ámbitos del conocimiento relativamente nuevos, donde las necesidades de la industria suelen hacer que las soluciones ad-hoc, más bien prácticas, precedan a los aportes teóricos y soluciones basadas en estándares, con horizontes de mediano o largo plazo.

Esta circunstancia ha promovido la necesidad de contar con enfoques ingenieriles para abordar el problema de la creación, evaluación, mantenimiento y evolución de sitios y aplicaciones Web¹ en consideración de restric-

¹ En los últimos años se ha observado una creciente preocupación de la comunidad científica de Ingeniería de Software, Hipermedia, Interfase Persona-Computadora (entre otras comunidades) por los temas relacionados a la Web, de modo que se ha dado en llamar a la disciplina como Ingeniería Web. Por ejemplo, en 1998, se celebró el primer workshop sobre Ingeniería Web, en conjunción con la 7ma Conferencia World Wide Web (WWW), en Australia. Se celebraron cinco Conferencias Internacionales en Ingeniería Web (ICWE): España (2001), Argentina (2002), España (2003), Alemania (2004) y

ciones de calidad y presupuesto. Esto implica aplicar a este campo del desarrollo de software el empleo sistemático, disciplinado y cuantificable de principios y prácticas reconocidos de Ingeniería.

Sin embargo, también se reconocen valiosas contribuciones, tanto empíricas como teóricas, que vienen realizando, por ejemplo, expertos como Jakob Nielsen [NIE04a], últimamente en el área de sitios y aplicaciones Web, e investigadores como Nigel Bevan [BEV97a], en calidad en uso de productos software, destinadas a evaluar productos de software en contextos de uso reales, ya sea por medio de expertos y/o usuarios reales, en contextos específicos de trabajo.

Concretamente, los trabajos de Nielsen [NIE01b], de amplia difusión en relación a evaluaciones de usabilidad para la Web, tienen un enfoque basado en heurísticas, orientado a proveer criterios para cambios y mejoras a corto plazo y, en forma complementaria, rankings de calidad de aplicaciones Web basados en casos de estudio comparativos.

Para poner en contexto esta tesis también es importante repasar la evolución que han tenido las definiciones de calidad en uso a lo largo del tiempo. Como ejemplo puede presentarse la versión actual de la norma ISO/IEC 9126-1 [ISO9126-1] del año 2001, sobre calidad de producto de software, que propone definiciones taxativas y diferentes para calidad interna, calidad externa y calidad en uso, en lugar de la única definición existente en la versión anterior del año 1991 [ISO9126].

Lo significativo de esas tres definiciones es que representan tres perspectivas de calidad, sensiblemente diferentes entre sí, pero en todo caso refiriéndose a un producto de software cuando es utilizado bajo ciertas condiciones y en un contexto bien definido, para dejar en claro que calidad no es un concepto absoluto y depende de condiciones y usuarios específicos [OLS04b].

También se consideró necesario tener en cuenta una analogía de dos conceptos clave como usabilidad y calidad en uso; ya que suelen usarse frecuentemente en la literatura de modo intercambiable, o con idéntico significado, existiendo sin embargo algunos trabajos que le asignan significados sensiblemente distintos, tal como se discutirá en la sección 2.3.3 "*Usabilidad y Calidad en Uso*". Básicamente se justificará, analizando esta rela-

Australia (2005). En cuanto a publicación de libros, en 1998, se editó el libro de Powell et al denominado "*Web Site Engineering: beyond Web Page Design*", luego apareció el libro denominado "*Web Engineering: Managing Diversity and Complexity of Web Application Development*" de Murugesan et al, por citar sólo algunos. En cuanto a revistas científicas tratamientos del tema se encuentran en estos últimos tres años en el *Journal of Web Engineering* de Rinton Press (<http://www.rintonpress.com/journals/jwe/>).

ción, que el concepto de calidad en uso es más amplio y abarcativo que usabilidad.

1.2. Objetivos Básicos de la Tesis

Sin negar el aporte de los estudios heurísticos, es posible considerar que un enfoque basado en modelos, con procesos de evaluación sustentados en metodologías apropiadas, eminentemente objetivos antes que subjetivos y cuantitativos más que cualitativos, puede resultar una contribución valorable y significativa. Se propone para ello, y es el objetivos principal de esta tesis, una visión ingenieril para evaluar calidad en uso, que no descansa solamente en reglas diseñadas por expertos y eventualmente validadas por la acción de los desarrolladores, diseñadores y usuarios, sino que se sostiene en tres fundamentos:

- a) Un *modelo conceptual* de calidad en uso.
- b) Un *marco de referencia* para la adopción, diseño y cálculo de las métricas e indicadores necesarios para medir calidad en uso.
- c) Una *metodología* de referencia para llevar adelante completamente el proceso de evaluación.

La propuesta que se presenta en este trabajo, se ejemplifica en el capítulo 5 con un caso de estudio sobre evaluación de calidad en uso. El caso de estudio se diseñó y ejecutó con la participación de usuarios reales en un contexto real de uso, sobre un tipo particular de producto software: una aplicación Web para el dominio de e-Learning.

El interés sobre el dominio de e-Learning está dado por la importancia que está adquiriendo la Web como soporte a los procesos de entrenamiento y enseñanza/aprendizaje, en el ámbito empresarial y educativo respectivamente, y en este último caso con mucho énfasis a nivel universitario [MIL03].

Es posible considerar que la medición y evaluación de calidad interna, calidad externa y calidad en uso están estrechamente ligados. En relación a esta aseveración se presentan, en el capítulo 4, lineamientos de una propuesta de evaluación de calidad externa para el mismo tipo de aplicación Web que el utilizado en el caso de estudio del capítulo 5.

Además, si bien no es el objetivo de esta tesis realizar propuestas sobre calidad de procesos, se mencionarán brevemente, dentro del capítulo sobre el estado del arte, experiencias acerca de actividades de “ingeniería de usabilidad” dentro de un proceso de desarrollo, con el objetivo de remarcar el carácter interdependiente de las diferentes perspectivas de la calidad de un producto software.

En resumen, dado que medir y evaluar la calidad en uso de un producto software permite determinar el grado en que dicho producto satisface las necesidades y expectativas de un usuario final realizando tareas concretas y en un contexto de uso particular, bien definido y especificado; las discusiones, reseñas y experiencias que se presentan a continuación pretenden contribuir a la medición y evaluación sistemática y disciplinada de calidad en uso de productos Web.

1.3. Organización de la Tesis

Esta tesis está organizada en seis capítulos, además de esta introducción. Un análisis del concepto calidad en uso en la literatura y en la práctica se presenta en el capítulo 2, donde además se discute acerca de la evolución de los concepto de calidad en uso en relación a calidad y usabilidad. En el capítulo 3 se presenta un marco conceptual de medición y evaluación para calidad en uso, basado en un modelo de métricas e indicadores con soporte en una ontología específica. El capítulo 4 se refiere al método de evaluación adoptado y adaptado para la evaluación y medición de calidad en uso, se mencionan también estudios preliminares y complementarios relacionados. Luego, en el capítulo 5, se detalla el diseño de un caso de estudio sobre calidad en uso para el dominio de e-Learning; también se discuten cuestiones que justifican la elección del dominio y se detallan las métricas e indicadores propuestos. El capítulo 6 está destinado a presentar la ejecución y los resultados del caso de estudio, incluyendo conclusiones preliminares acerca de la calidad en uso del producto evaluado y justificaciones en relación a esos resultados.

En el capítulo 7, por último, están las consideraciones finales y líneas de trabajo futuro relacionadas. Luego de las referencias, un conjunto de cuatro apéndices complementan el material central de esta tesis.

Capítulo 2. Calidad y Calidad en Uso de Aplicaciones Software y Web: Estado del Arte

Introducción

Dado que el objetivo principal de esta tesis es contribuir, desde un punto de vista metodológico, a medir y evaluar calidad en uso de sitios y aplicaciones Web en forma sistemática y disciplinada, es importante revisar, inicialmente, la evolución del concepto calidad en uso. Para ello, en este capítulo se presenta y discute la literatura más relevante del área, resaltando primero definiciones y modelos acerca de calidad de producto software en general (secciones 2.1 y 2.2). Luego se presenta el estado del arte con respecto a la calidad en uso, en donde se destacan semejanzas y diferencias con el concepto de usabilidad. Más adelante se discuten modelos, métodos y técnicas que, empleados junto a ciertos criterios, permiten medir y evaluar calidad en uso de un producto software (sección 2.3) y se pueden adaptar para evaluar calidad en uso de aplicaciones Web en particular.

Antes de las conclusiones del capítulo se incluye una sección sobre calidad de aplicaciones Web (sección 2.4), con énfasis en la calidad en uso como objetivo final del proceso de diseño, implementación y evaluación de calidad de un producto software. Las conclusiones del capítulo están en la sección 2.5.

2.1 Calidad de Producto Software

Un hito en la definición de estándares de calidad de producto software, destinado a evaluación, se dio a finales de 1991, cuando ISO/IEC publicó el modelo de calidad y el proceso de evaluación [ISO9126]. Como antecedente es necesario remarcar que desde 1976 se habían publicado trabajos que fueron reconocidos y tomados en cuenta en el estándar, entre los que corresponde mencionar los realizados por Bohem [BOE76] [BOE78], Mc Call [MCC77] y la fuerza aérea de los Estados Unidos. En estos trabajos ya se definían modelos y marcos de calidad.

La virtud de la organización ISO/IEC fue lograr el consenso necesario para un emprendimiento que tuviera alcance y reconocimiento internacional. Esta especificación fue evolucionando a lo largo del tiempo, manteniéndose en buena medida la estructura y espíritu originales, tal como se verá más adelante, en la sección 2.2.

El estándar ISO/IEC 9126 [ISO9126] prescribe seis características que describen, con mínimo solapamiento, a la calidad de software. Además, informa acerca de un conjunto de subcaracterísticas de calidad para cada característica en particular. También especifica un modelo de proceso de evaluación, en donde las entradas de información para la definición de re-

querimientos de calidad son el modelo de calidad ISO y las necesidades explícitas e implícitas de los usuarios.

La definición de calidad en este estándar es: *“La totalidad de características de un ente teniendo en cuenta su capacidad de satisfacer necesidades explícitas e implícitas”*² (Téngase en cuenta que esta definición fue adoptada de la versión previa del estándar ISO 8402 titulado “Quality Vocabulary”, publicado en 1986). Las seis características prescritas son: *Usabilidad, Funcionalidad, Confiabilidad, Eficiencia, Portabilidad y Mantenibilidad*. Por ejemplo *Usabilidad*, a la que se prestará especial atención en la sección 2.3.3, está definida como *“Un conjunto de atributos teniendo en cuenta el esfuerzo necesario para usar, y sobre la evaluación individual de tal uso, realizado por un conjunto definido o implícito de usuarios”*³. A su vez, *Usabilidad* está subdividida en tres subcaracterísticas llamadas: *Comprensibilidad, Aprendibilidad* y *Operabilidad*. *Operabilidad*, por ejemplo, está definida como *“Atributos de software que tienen que ver con el esfuerzo de los usuarios para operar y controlar la operación”*⁴.

Es importante remarcar los siguientes puntos en este estándar (algunos provenientes de las contribuciones de Bohem y McCall) [OLS04b]:

- ❖ el significado de calidad es un concepto compuesto, multidimensional, que no puede ser medido directamente;
- ❖ dada la complejidad que involucra el concepto de calidad, es necesario un modelo para especificar requerimientos de calidad de producto;
- ❖ por razones de claridad y manejo, el modelo de calidad, de propósito general, contiene una cantidad mínima de características por medio de las cuales cualquier clase de software puede ser evaluado;
- ❖ para la actividad de definición de requerimientos de calidad son consideradas las necesidades explícitas e implícitas de los usuarios; y además en ciertas definiciones de características y subcaracterísticas se reconoce la importancia del usuario, por ej. en usabilidad y subcaracterísticas asociadas;
- ❖ este enfoque difiere de los enfoques tradicionales de calidad donde el énfasis está en cumplir requerimientos especificados que son principalmente de orden funcional. Interpretado literalmente, esto significa que la calidad de un producto queda en manos del individuo que realiza la especificación de requerimientos [BEV97]. En [GAR84] este

² Quality: the totality of characteristics of an entity that bear on its ability to satisfy stated and implied needs

³ A set of attributes that bear on the effort needed for use, and on the individual assessment of such use, by a stated or implied set of users

⁴ Attributes of software that bear on the users' effort for operation and operation control

enfoque se describe como “*calidad en la manufactura*”(manufacturing quality);

Como se ha señalado anteriormente, la definición de calidad en el estándar [ISO9126] remarca que la meta de la calidad es cumplir con las necesidades de los usuarios. Pero lo que no está claramente explicitado es que el propósito de la calidad de software es que sea percibido con calidad, esto es, percibido con grados de excelencia por los usuarios finales, en contextos reales de uso. Así, el estándar [ISO9126] deja bastante claro que la calidad está determinada por la presencia o ausencia de atributos, con la implicancia de que estos son atributos específicos que deben ser diseñados en el producto. En relación a esto Bevan [BEV99] ha dicho que “*Si bien los desarrolladores quisieran conocer qué atributos incorporar en el código para reducir el ‘esfuerzo requerido para el uso’, la presencia o ausencia de atributos predefinidos no puede asegurar usabilidad, en tanto no haya una forma confiable de predecir el comportamiento de los usuarios del producto final*”⁵.

2.2. Calidad Interna y Externa

Para achicar la brecha entre calidad diseñada y calidad percibida, el estándar [ISO9126] fue revisado (desde 1994) con el objetivo de especificar un nuevo marco de calidad que distingue entre tres enfoques diferentes de calidad de software. Estos son: *calidad interna*, *calidad externa* y *calidad en uso*. El estándar [ISO9126-1], que incluye estos tres enfoques de calidad fue oficialmente publicado en 2001, mientras que el modelo de proceso de evaluación inicialmente incluido en el estándar ISO 9126 fue retirado y desarrollado completamente en la serie ISO/IEC 14598 [ISO14598-1] y principalmente en [ISO14598-5].

Los enfoques de calidad interna y externa de producto software, en el estándar [ISO9126-1], pueden resumirse de la siguiente manera, tal como se reseña en [OLS04b]:

- ❖ **Calidad Interna** está especificada por un modelo de calidad (similar al modelo 9126), y puede ser medida y evaluada por medio de atributos estáticos de documentos tales como especificación de requerimientos, arquitectura o diseño; piezas de código fuente, etc. En etapas tempranas del ciclo de vida del software es posible medir, evaluar y controlar la calidad interna de estos productos, pero asegurar la calidad interna no es generalmente suficiente para asegurar calidad externa.

⁵ Although developers would like to know what attributes to incorporate in the code to reduce the “effort required for use”, presence or absence of predefined attributes cannot assure usability, as there is no reliable way to predict the behaviour of the users of the final product

- ❖ **Calidad Externa** está especificada también por un modelo de calidad (similar al modelo 9126), y puede ser medida y evaluada por medio de propiedades dinámicas del código ejecutable en un sistema de computación, esto es, cuando un módulo o la aplicación completa es ejecutado en una computadora o en una red simulando lo más cercanamente posible un ambiente real. En fases tardías del ciclo de vida del software (principalmente en distintas etapas de testing o ya en estado operativo de un producto de software o aplicación Web), es posible medir, evaluar y controlar la calidad externa de estos productos ejecutables. Más adelante, en la sección 2.3.4, se verá la relación directa existente entre calidad externa y calidad en uso.

La definición de calidad interna dada en el estándar 9126-1 está dada por *“la totalidad de atributos de un producto que determina su capacidad de satisfacer necesidades explícitas e implícitas cuando es usadas bajo condiciones específicas”*⁶; la definición de calidad externa es *“el grado en la que un producto satisface necesidades explícitas e implícitas cuando se utiliza bajo condiciones especificadas”*⁷; (nótese que estas definiciones están en el estándar [ISO14598-1]).

Estas dos definiciones diferentes de calidad, junto a la definición de calidad en uso (en lugar de la única definición existente en el estándar previo [ISO9126]), se refieren particularmente a un producto cuando es usado bajo condiciones y contextos de uso especificados, para dejar claro que calidad no es un concepto absoluto, y depende de condiciones y contextos de uso específicos, para usuarios específicos.

Para ambos modelos, el de calidad interna y el de calidad externa, se han mantenido en la revisión las seis características principales de calidad. Aun más, a nivel de subcaracterísticas se han transformado en prescriptivas en vez de informativas. Además, se han añadido nuevas subcaracterísticas y otras redefinidas en términos de “capacidad del software” para facilitar la interpretación de las mismas desde una perspectiva de calidad interna o de calidad externa. Por ejemplo, la característica usabilidad está definida en el estándar [ISO9126-1] como *“la capacidad del producto software para ser comprendido, aprendido, utilizado y atractivo para el usuario, cuando se usa bajo condiciones especificadas”*⁸. A su vez, usabilidad está subdividido en cinco subcaracterísticas, denominadas: Comprensibilidad, Facilidad para ser Aprendido (o Aprendibilidad) y Operabilidad, además de Grado de Atractivo

⁶ The totality of attributes of a product that determine its ability to satisfy stated and implied needs when used under specified conditions

⁷ The extent to which a product satisfies stated and implied needs when used under specified conditions

⁸ The capability of the software product to be understood, learned, used and attractive to the user, when used under specified conditions

y Conformidad con Estándares (ver tabla 2.1 para el detalle de la definición de estas subcaracterísticas).

| Subcaracterísticas | Definición |
|---|---|
| Comprensibilidad (Understandability) | <i>La capacidad de un producto de software para permitir al usuario comprender cuando el software es adecuado, y como puede ser usado para una tarea particular bajo determinadas condiciones de uso.</i> |
| Facilidad para ser Aprendido (Learnability) | <i>La capacidad de un producto de software para permitir al usuario aprender su aplicación.</i> |
| Operabilidad (Operability) | <i>La capacidad de un producto de software para permitir al usuario operarlo y controlarlo.</i> |
| Grado de Atractivo (Atractiveness) | <i>La capacidad de un producto de software de ser atractivo para el usuario.</i> |
| Conformidad con Estándares (Compliance) | <i>La capacidad de un producto de software para adherir a estándares, convenciones, guías de estilo o regulaciones relacionadas a usabilidad.</i> |

Tabla 2.1. Definición de las subcaracterísticas de Usabilidad prescrita en el estándar ISO 9126-1, para calidad interna y externa.

Calidad externa es, finalmente, el resultado del comportamiento compartido de un componente o aplicación de software y un sistema de computación. A su vez, midiendo y evaluando la calidad externa se puede ayudar a validar la calidad interna.

Análogamente, tomar en cuenta atributos apropiados del software para calidad interna es un prerequisite para alcanzar el comportamiento externo requerido, y considerar los atributos apropiados del software o aplicación Web para el comportamiento externo es un prerequisite para alcanzar calidad en uso, tal como se explicitará en el apartado 2.3.

2.3. Calidad en Uso: Estado del Arte

El concepto de calidad en uso (bajo este nombre o con sinónimos, como veremos) ha estado revisándose en la literatura y en la práctica a lo largo de los últimos quince años. En lo que concierne a software, su significado ha evolucionado acompañando a la notable evolución de la industria del software en ese período. En la opinión de algunos expertos, como Bevan [BEV95a], el cambio fundamental es la atención cada vez mayor que merecen tanto los usuarios como el contexto considerados, cuando se evalúa el desempeño de un producto software en uso.

Junto a la madurez que han adquirido los usuarios en el uso de herramientas de software y sistemas de información informatizados, se han in-

crementado también sus requerimientos de calidad con respecto al software en general y a las aplicaciones Web en particular, en contextos reales de uso.

A continuación se discute acerca de qué significa y cómo se usa el término Calidad en Uso en diferentes contextos, y qué influencia tienen los usuarios finales en el enfoque de algunos autores.

2.3.1. Calidad en Uso en la Literatura y en la Práctica

Las normas recientes ISO/IEC reflejan la importancia de tener en cuenta el punto de vista de los usuarios y, como ya se ha dicho, la última versión de el estándar [ISO9126-1] confirma una apertura conceptual en las definiciones, propiciando un marco de evaluación más complejo pero también más abarcativo y adaptable a distintos productos, procesos y modelos de software, ya que se especifica a la calidad en uso como un modelo de calidad diferente y complementario al modelo de calidad de producto de software.

De aquí en adelante, en este capítulo, se propondrá discutir e indagar sobre definiciones de calidad en uso y sus relaciones con otros conceptos vinculados, como una forma de encontrar criterios generales y líneas guías para una propuesta de medición y evaluación sistemática de calidad en uso para aplicaciones Web. Para lograrlo, se revisarán las relaciones existentes entre los conceptos calidad y calidad en uso, usabilidad y calidad en uso, y las definiciones, a veces complementarias y a veces contradictorias, presentes en la literatura y en las normas dadas por los organismos de estandarización.

2.3.2. Calidad y Calidad en Uso

El término calidad es empleado frecuentemente en la literatura y en la práctica de las organizaciones en general, básicamente agregando valor o realzando propiedades de un producto o servicio; no obstante, existen distintos enfoques acerca de su significado en determinados contextos, acerca de cómo puede desarrollarse, construirse, controlarse y de cómo puede afectar en particular a un producto de software.

Considérense por ejemplo las definiciones dadas en algunas normas ISO en general. En el estándar [ISO9000] del año 2000 se define calidad como el *“grado en el que un conjunto de características inherentes [al producto] cumple con los requerimientos”* aclarándose que inherente significa que existe algo especial como característica o propiedad permanente. El estándar [ISO9000] reemplazó al [ISO8402], donde se proponía calidad como la totalidad de características de un producto que posee capacidad para satisfacer necesidades explícitas e implícitas y el énfasis estaba puesto en el producto, en la existencia o no de atributos mensurables que determinaban las características deseables.

Teniendo en cuenta calidad para producto de software, la primera versión ISO/IEC 9126 de 1991 [ISO9126] adopta este punto de vista y considera seis grandes categorías de características que permitirían diseñar o evaluar la calidad de un producto de software, tal como se indicó en la sección 2.1.

Calidad vista así, y siempre y cuando fuera posible conocer las características necesarias para los potenciales usuarios, parece ser una propiedad exclusiva de un producto, pero, ¿qué ocurre si distintos grupos de usuarios tienen diferentes necesidades? Seguramente las características deseables para que el producto sea considerado de calidad serían distintas.

Ya en su trabajo sobre calidad en uso de 1997 [BEV97], Bevan y Azuma confrontan los conceptos genéricos de calidad percibida y calidad en uso, indicando que usualmente la percepción de la calidad por parte de los usuarios es juzgada como subjetiva e inexacta; particularmente discuten a Garvin [GAR84] quien afirma que esos conceptos pueden ser *"...tan subjetivos como evaluaciones estéticas"*, afirmando en cambio que hay otra razón para comprometerse con la percepción de calidad de los usuarios y es que *"un producto puede tener calidad sólo en relación al propósito para el que fue creado"*.

Como consecuencia de esta postura consideran que reconciliar los trabajos sobre calidad en uso con los enfoques tradicionales sobre calidad ha traído aparejada otra visión, más amplia y potencialmente importante, que abarca la calidad percibida por los usuarios, relacionando a la calidad directamente con las necesidades de los usuarios de productos específicos.

Se puede tomar como ejemplo la necesidad de desplazarse por una gran ciudad, con el tránsito congestionado y poco espacio de estacionamiento posiblemente. Allí un auto de gama alta, de grandes dimensiones, podría no ser considerado de calidad, dado que será difícil de maniobrar o estacionar. En cambio, uno pequeño y ágil resultará más atractivo, y será percibido como de calidad para esa circunstancia. Contrariamente, para un viaje largo, donde los requerimientos de confort son mayores, la percepción de la calidad seguramente se invertirá.

Con este nuevo enfoque, el estándar [ISO14598-1] ya no considera a la calidad disociada de las necesidades de los usuarios sino que calidad en uso puede ser evaluada de acuerdo a la eficacia, productividad y satisfacción con las que una clase de usuario específico puede alcanzar metas específicas en un ambiente particular de trabajo. En el contexto de un proceso de aseguramiento de calidad de un producto de software, los atributos de calidad (interna y externa) deberían ser el medio y la calidad en uso el fin, el objetivo a alcanzar.

También el ISO/IEC 14598 dice que las necesidades de los usuarios pueden expresarse como un conjunto de requerimientos deseables para el

comportamiento del producto en uso y que estos requerimientos, a su vez, pueden cuantificarse a partir de una serie de métricas recolectadas cuando el sistema esté en uso, en el contexto en el que será evaluado. Otra contribución importante, tratándose de una guía de evaluación, es la correspondencia que propone entre los distintos niveles de evaluación que pueden establecerse en un producto de software entre métricas internas y métricas externas. Las primeras corresponden a la medición de artefactos producidos en la fases de diseño y desarrollo (modelos, código fuente, etc.), mientras que las segundas se aplican a versiones ejecutables del producto, en las fases de integración del sistema y testing, y también sobre el producto en uso.

La versión actual de el estándar [ISO9126-1] de 2001 tiene un enfoque integrado de calidad, tal como se ha mencionado en la sección 2.1 sobre calidad de producto software, propiciando la inclusión de requerimientos de calidad en uso ya en la especificación de requerimientos del producto. Si bien considera que es adecuada la definición dada en [ISO8402], sostiene también que las metas de calidad pueden ser realmente evaluadas cuando el producto está en uso, y efectivamente puede confrontarse la percepción de los usuarios con las necesidades manifestadas inicialmente, en un contexto determinado de funcionamiento.

En otra dimensión, orientada a otro tipo de productos (equipamiento con terminales visuales), el estándar sobre ergonomía [ISO9241-11] emplea también el concepto calidad en uso con un significado amplio de calidad, semejante al que tienen las normas [ISO9126-1] y [ISO14598-1]. Concretamente sostiene que la calidad en uso de un producto es dependiente del contexto de uso. Entiende por contexto al usuario (considerando sus características específicas), las tareas a realizar, el equipamiento utilizado y el ambiente físico y social en el que se evalúa el producto, que pueden influir en la calidad en uso de un producto en un lugar de trabajo.

En este sentido es posible decir que el modelo de calidad en uso de el estándar ISO 9126-1 es una ampliación (o adaptación) del modelo ISO 9241-11, específicamente para software, considerando que la calidad en uso puede ser medida y evaluada teniendo en cuenta el grado en que el software o aplicación Web cumple con las necesidades de usuarios reales, en un contexto real, específico, de uso.

2.3.3. Usabilidad y Calidad en Uso

Los términos usabilidad y calidad en uso han estado empleándose como sinónimos en la comunidad de Ingeniería de Software e Ingeniería Web por un largo período.

En una encuesta donde se repasa el uso del término usabilidad bajo distintos enfoques, a partir de trabajos previos reconocidos, E. Folmer y J.

Bosch [FOL04] afirman, citando el trabajo seminal de Bevan, Kirakowsky y Maissel de 1991 "*What is Usability?*" [BEV91], que el término usabilidad tiene su origen en el concepto de "*user friendly*", pero que esta frase fue adquiriendo un sentido vago y extremadamente subjetivo, sugiriéndose el término usabilidad para reemplazarla. Luego usabilidad fue definida como una característica principal de la calidad de un producto de software, en el estándar [ISO9126] y fue ampliada también en el [ISO9126-1], esta vez con calidad en uso, que ofrece una idea más abarcativa y completa de calidad que usabilidad.

Además, Folmer y Bosch sostienen que se ha invertido un enorme esfuerzo en definir el término usabilidad, estableciendo atributos que pueden ser medidos y que a través de un proceso de composición definen usabilidad. Si bien juzgan importante la discusión, no consideran necesario profundizarla, dado que el objetivo del trabajo citado es estudiar usabilidad en el contexto de un método de diseño de software, en particular a nivel arquitectural.

Hay otros autores que emplean usabilidad directamente en términos de calidad en uso, por ej. en [JUR01] dicen que "*Una definición de usabilidad es calidad en uso*", remitiéndose directamente a el estándar ISO 14598-1, en sintonía con la propuesta de [BEV95b].

Un experto reconocido en usabilidad, Jakob Nielsen, quien ha estado trabajando en evaluación de usabilidad desde principios de los '90, no elabora una definición precisa de usabilidad, sino que a través de una serie de heurísticas que han sido empleadas consistentemente a lo largo del tiempo y en diferentes dominios (especialmente en comercio electrónico), ofrece criterios cualitativos para evaluar usabilidad.

Según Nielsen, Usabilidad se constituye en base a cinco atributos: Facilidad de Aprendizaje (Learnability), Eficiencia (Efficiency), Facilidad de Memorización (Memorability), Baja Tasa de Errores (Errors –low rate-) y Satisfacción (Satisfaction).

Por otro lado, el estándar [ISO9241-11] propone una definición más concreta de usabilidad, ilustrada en la Figura 2.1, pero curiosamente en los trabajos ya citados de [BEV95b] y [FOL04], donde el estándar es reseñado se emplean los términos calidad en uso y usabilidad respectivamente. En el primer caso se enfatiza el carácter más amplio del concepto de calidad en uso respecto de usabilidad, ya que puede ser influenciado por cualquiera de las características que definen la calidad del producto. En el segundo caso, en cambio, se emplea el término usabilidad literalmente, pero se remarca que mantiene una sintonía con la definición dada para calidad en uso del estándar [ISO9126-1], es decir una idea de usabilidad más amplia, que re-

presenta el objetivo del diseño o la meta final del producto, satisfaciendo las necesidades de los usuarios.

La diferencia, más allá del nombre que se ponga al concepto –calidad en uso o usabilidad-, solamente radicaría en que en el estándar ISO 9126-1 tiene seguridad (safety) como característica adicional a eficacia, productividad (ligado a eficacia) y satisfacción. En cambio en el estándar sobre ergonomía se consideran por separado cuestiones de salud y seguridad.

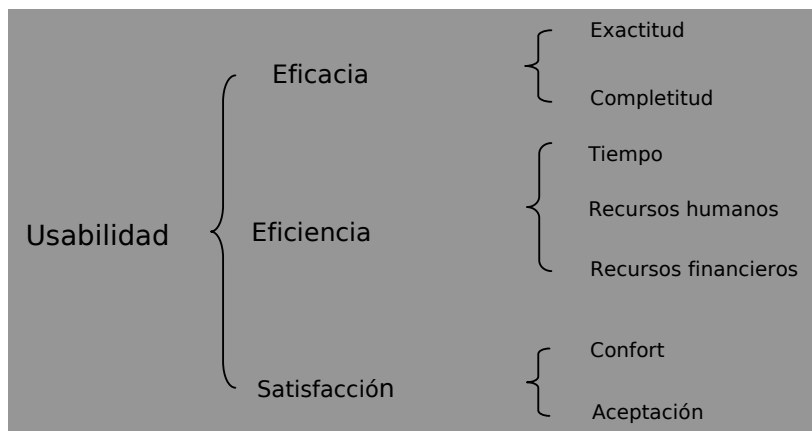


Figura 2.1. Usabilidad en el estándar 9241-11.

En la sección siguiente se presenta un modelo y una definición concreta de calidad en uso, relacionada directamente con un producto software, propuesto en el estándar [ISO9126-1] y que se adoptará como base para esta tesis.

2.3.4. Definición y Modelo de Calidad en Uso para Producto Software

El estándar [ISO9126-1] define calidad en uso como “*la capacidad de un producto de software de facilitar a usuarios específicos alcanzar metas específicas con eficacia, productividad, seguridad y satisfacción en un contexto específico de uso*”⁹.

Añade que “*calidad en uso es la visión de calidad de los usuarios de un ambiente conteniendo software, y es medida sobre los resultados de usar el software en el ambiente, antes que sobre las propiedades del software en sí mismo*”¹⁰.

⁹ The capability of the software product to enable specified users to achieve specified goals with effectiveness, productivity, safety and satisfaction in specified contexts of use.

¹⁰ Quality in use is the user's view of the quality of an environment containing software, and is measured from the results of using the software in the environment, rather than properties of the software itself.

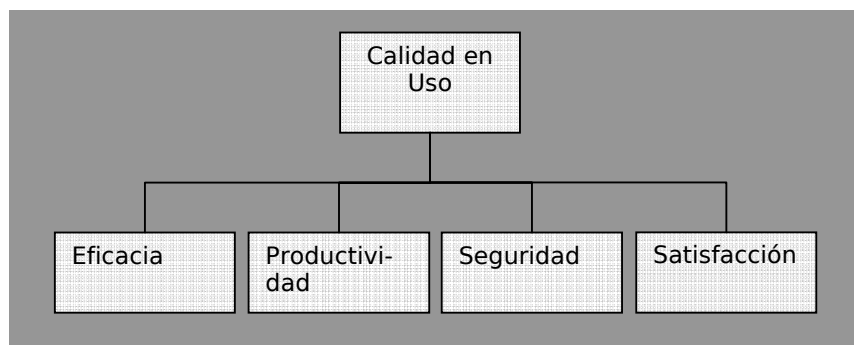


Figura 2.2. Modelo para calidad en uso. ISO IEC 9126-1 (2001).

Las características de calidad en uso son agrupadas en cuatro categorías (ilustradas en la figura 2.2): Eficacia (Effectiveness), Productividad (Productivity), Seguridad(Safety) y Satisfacción (Satisfaction), y son definidas tal como se muestran en la tabla 2.2.

| Característica | Definición |
|-----------------------|---|
| Eficacia | <i>La capacidad del producto de software para facilitar a los usuarios alcanzar metas específicas con exactitud y completitud en un contexto específico de uso</i> |
| Productividad | <i>La capacidad del producto de software para invertir la cantidad apropiada de recursos en relación a la eficacia alcanzada en un contexto específico de uso.</i> |
| Seguridad | <i>La capacidad del producto de software para alcanzar niveles aceptables de riesgo de dañar a las personas, el negocio, el software, la propiedad o el ambiente en un contexto específico de uso.</i> |
| Satisfacción | <i>La capacidad del producto de software para satisfacer a los usuarios en un contexto específico de uso.</i> Nota [del estándar ISO]: Satisfacción es la respuesta del usuario a la interacción con el producto, e incluye la actitud hacia el uso del producto. |

Tabla 2.2. Definición de las cuatro características de Calidad en Uso prescritas en la ISO/IEC 9126-1.

La relación existente entre las tres perspectivas de calidad de producto software puede sintetizarse de la siguiente manera: medir y evaluar la calidad en uso por medio de métricas e indicadores, tal como se verá en el capítulo 3, puede ayudar a validar la calidad externa del software; calidad en uso evalúa el grado de excelencia de un producto y puede ser usada para validar el grado en que el software o aplicación Web cumple con necesidades específicas de los usuarios. A su vez, medir y evaluar la calidad externa puede ayudar a validar la calidad interna.

Análogamente, tomar en cuenta atributos apropiados del software para calidad interna es un prerrequisito para alcanzar el comportamiento externo

requerido, y considerar los atributos apropiados del software para el comportamiento externo es un prerrequisito para alcanzar calidad en uso (esta dependencia se sugiere en la figura 2.3).

*¡Error! Vínculo no válido. **Figura 2.3.** Marco conceptual de calidad teniendo en cuenta diferentes tipos de entidades y potenciales modelos de calidad, de interés para Ingeniería de Software e Ingeniería Web*

Las afirmaciones anteriores se basan en notas relacionadas a la definición de calidad en uso, dentro del mismo estándar 9126-1 donde se menciona que *“Calidad en uso es la vista de calidad que tienen los usuarios. Alcanzar calidad en uso depende de alcanzar la calidad externa necesaria, la que a su vez depende de alcanzar la calidad interna necesaria”*¹¹; y agrega que *“Las mediciones se requieren normalmente a los tres niveles, ya que alcanzar criterios para mediciones internas no es suficiente para asegurar alcanzar criterios para mediciones externas, y lograr satisfacer criterios para mediciones externas de subcaracterísticas no es normalmente suficiente para asegurar alcanzar criterios de calidad en uso”*¹².

2.3.5. Criterios para Medir y Evaluar Calidad en Uso

Si se considera que la evaluación de calidad en uso se realiza sobre un producto en funcionamiento, es necesario emplear un contexto real de trabajo en el que el software será utilizado, en cuanto al perfil de usuario, el equipamiento y las tareas a realizar. Se trata de una evaluación que se orienta eminentemente a tareas, ya que es necesario evaluar cuan eficaces, productivos, seguros y satisfechos resultan los usuarios empleando un producto, en un contexto específico.

Teniendo en cuenta estas particularidades, y en relación al interés concreto de esta tesis, esto es evaluar sistemáticamente calidad en uso de aplicaciones Web, se revisan a continuación procesos y algunos métodos y técnicas pertinentes a la evaluación de calidad en uso de un producto software. Específicamente aquellos que han sido adaptados, personalizados y en ciertos casos automatizados por expertos e investigadores para facilitar la evaluación de calidad en uso de aplicaciones Web.

2.3.5.1. Procesos para Evaluación de Calidad en Uso

¹¹ Quality in use is the user's view of quality. Achieving quality in use is dependent on achieving the necessary external quality, which in turn is dependent on achieving the necessary internal quality.

¹² Measures are normally required at all three levels, as meeting criteria for internal measures is not usually sufficient to ensure achievement of criteria for external measures, and meeting criteria for external measures of subcharacteristics is not usually sufficient to ensure achieving criteria for quality in use.

De entre los modelos de proceso de evaluación es importante referirse al [ISO14598-5], que como se ha dicho en la sección 2.2 estaba incluido originalmente en [ISO9126], y que tiene como objetivo principal proveer un marco de evaluación genérico, abstracto, que permita a los evaluadores, junto con desarrolladores, compradores, vendedores y usuarios en general expresar requerimientos de calidad siguiendo el modelo de calidad definido en el estándar [ISO9126-1]. La evaluación debe tener en cuenta una variedad de documentos que pueden ser considerados parte del producto de software, tales como documentación de diseño, código fuente, tests o documentación para el usuario final.

Por lo tanto, este proceso de soporte es aplicable dentro de los procesos principales del ciclo de vida del software prescritos en el estándar ISO/IEC 12207, esto es: aseguramiento de calidad, verificación, validación, revisión conjunta y auditoría.

Dentro de las características deseables para cualquier proceso sistemático, el estándar menciona cuatro de carácter obligatorio: repetitividad, reproducibilidad, imparcialidad y objetividad. Si bien no indica una técnica o procedimiento en particular a emplear, no obstante especifica claramente las cinco actividades que deben llevar a cabo los evaluadores:

1. Establecer los requerimientos de evaluación
2. Especificar la evaluación en base a los requerimientos de evaluación y sobre la descripción del producto provista por quien solicita la evaluación
3. Diseñar un plan de evaluación basado en la especificación de la evaluación; aquí se deben tener en cuenta los componentes del software a ser evaluado y los métodos y herramientas propuestos por el evaluador
4. Ejecutar el plan de evaluación, que consiste en inspeccionar, modelar, medir y testear el producto y sus componentes conforme al plan. Todo lo realizado por los evaluadores deber ser registrado y los resultados obtenidos colocados en un borrador del reporte de evaluación
5. Concluir la evaluación, que consiste en la entrega del reporte de evaluación y la desinstalación o destrucción, por parte del evaluador, tanto del producto como de sus componentes, cuando éstos han sido enviados en forma independiente sólo para su evaluación.

También en relación al proceso de aseguramiento de la calidad de un producto software, otro estándar que puede ser tomado como referencia es el [ISO13407], una guía para el proceso de aseguramiento de calidad en uso de sistemas interactivos basados en computadoras, que promueve el desa-

rollo de actividades centradas en el usuario a lo largo del ciclo de vida del producto.

Este estándar describe al diseño centrado en el usuario como una actividad multidisciplinaria, que permite incorporar factores humanos y conocimiento sobre ergonomía y prescribe técnicas con el objetivo de obtener mejores tasas de efectividad y productividad, y también de mejorar el ambiente de trabajo, contrarrestando eventuales perjuicios sobre los usuarios en lo concerniente a salud y seguridad.

Se consideran cuatro actividades de diseño centradas en el usuario que debieran llevarse a cabo a lo largo de todas las etapas de un proyecto: comprender y especificar el contexto de uso, especificar los requerimientos del usuario y la organización, producir soluciones de diseño y evaluar alternativas de diseño frente a requerimientos.

Además, es importante remarcar que si bien calidad en uso sólo puede ser evaluada con el producto ejecutable en el contexto real, también se pueden incorporar actividades para asegurar calidad en uso de un producto software a lo largo de todo el ciclo de vida.

Según [FER01] las prácticas de Ingeniería de Usabilidad, tales como las que propende el estándar [ISO13407], pretenden en todo caso definir el nivel de usabilidad anticipadamente e intentan asegurar que el software desarrollado alcance ese nivel. En [HIX93] se afirma que se trata de un proceso en el que las características de usabilidad son especificadas, cuantitativa y tempranamente en el proceso de desarrollo, y medidas a lo largo de todo el proceso.

Finalmente, en el exhaustivo trabajo sobre automatización de la evaluación de usabilidad de interfaces Web, [IVO01] reporta una serie de actividades concretas que se desarrollan normalmente como parte del proceso de evaluación de usabilidad. Dichas actividades son:

1. Especificar las metas de evaluación
2. Determinar los aspectos a evaluar de la Interfaz de Usuario
3. Identificar el perfil de los usuarios de usabilidad
4. Seleccionar métricas de usabilidad
5. Seleccionar el/los método/s de evaluación
6. Seleccionar Tareas
7. Diseñar experimentos
8. Capturar datos de usabilidad
9. Analizar e interpretar los datos
10. Realizar críticas y sugerir mejoras a la Interfaz de Usuario

11. Iterar el proceso si fuera necesario

12. Presentar los resultados

2.3.5.2. Métodos y Técnicas para Evaluar Calidad en Uso

El borrador del estándar [ISO9126-4] indica genéricamente cómo medir calidad en uso de un producto software. Textualmente dice que *“Las métricas de calidad en uso miden el grado en que un producto satisface las necesidades de un usuario específico, para lograr metas específicas con eficacia, productividad, seguridad y satisfacción”*¹³ y que *“Las necesidades de calidad de los usuarios pueden ser especificadas como requerimientos de calidad a través de métricas de calidad en uso, métricas externas y, algunas veces por métricas internas. Estos requerimientos especificados como métricas deberían ser usados como criterios de evaluación cuando sea evaluado el producto”*¹⁴.

En el trabajo [FOL04] se analizan definiciones de usabilidad –en varias ocasiones como sinónimo de calidad en uso- y se reseña que entre ciertos autores tales como Shackel [SHA91], Nielsen [NIE93] y algunos estándares ([ISO9241-11], [ISO9126]) existen, como denominador común, un conjunto de “atributos mensurables” que permiten especificar y evaluar usabilidad: Eficacia, Eficiencia (como productividad), y Satisfacción. También aparece mencionado, aunque con menos frecuencia, Facilidad de Aprendizaje (Learnability).

Consideran para la evaluación tres grandes grupos o categorías de métodos de evaluación: Testing, Inspección y Consultas, cada uno de los cuales incluye la posibilidad de emplear un conjunto de métodos o técnicas concretas, como por ejemplo Protocolo de Pensar en Voz Alta (Thinking Aloud Protocol), Evaluación Heurística y Registro del Uso Real (Logging Actual Use) o Encuestas, Inspección de Características Físicas, entre otras.

También en la línea de análisis de Folmer y Bosch, en [MAT02] se asegura que los enfoques más utilizados para evaluar usabilidad de sitios Web, que pueden ser extrapolados para evaluar calidad en uso tal como se verá en el caso de estudio del capítulo 5, son métodos orientados al usuario (testing por usuarios) y métodos de inspección, basados en la revisión por parte de expertos.

En el testing realizado por usuarios se tiene en cuenta la existencia de ciertas propiedades o atributos en el producto software o aplicación Web

¹³ The quality in use metrics measure the extent to which a product meets the needs of specified users to achieve specified goals with effectiveness, productivity, safety and satisfaction in aspecified context of use. This can be only achieved in a realistic system environment.

¹⁴ User quality needs can be specified as quality requirements by quality in use metrics, by external metrics, and sometimes by internal metrics. These requirements specified by metrics should be used as criteria when a product is evaluated.

(casi siempre sobre la interfaz de usuario), mientras se observa cómo es utilizado por usuarios reales realizando tareas representativas.

La categoría de métodos de inspección agrupa un conjunto de métodos basados en la labor de expertos, que observan exhaustivamente cuestiones relacionadas a usabilidad (u otras características) y pueden evaluar y calificar un producto software o aplicación Web en base a reglas preestablecidas, basadas en la experiencia y si fuera posible reiterando los tests con cierta frecuencia. Además, los métodos de inspección han sido utilizados con frecuencia para inspeccionar modelos de calidad. Cada uno de los métodos tiene sus pros y sus contras.

En el testing con usuarios, es interesante observar la reacción de los usuarios llevando a cabo tareas reales con el software y también es factible una evaluación rápida y eficaz a partir del seguimiento sobre la interfaz de usuario. Pero, por otro lado, no es siempre sencillo replicar un medio ambiente verdaderamente representativo o encontrar los usuarios con las características semejantes a las de los usuarios reales y llevar a cabo el estudio a un costo razonable. Además, algunos usuarios pueden sentirse incómodos si saben que serán observados y modificar su actitud en el momento del testing. Así, el método puede resultar intrusivo.

Los métodos de inspección, en cambio, pueden abaratar costos en los estudios, simplificándolos y prescindiendo del equipamiento e instalaciones para el testing con los usuarios, ya que es suficiente con la labor de los expertos. Si bien se pueden obtener resultados rápidamente, la confiabilidad de los mismos está directamente relacionada al nivel de conocimiento y experiencia de los evaluadores. Particularmente para la Web la primera metodología en emplear este método fue Olsina [OLS99], que, adaptada para evaluar calidad en uso, se empleó en el caso de estudio del capítulo 5.

Ambas categorías de métodos, testing con intervención de usuarios e inspección por parte de expertos, se basan en dos técnicas que pueden utilizarse en forma alternativa: evaluación guiada por heurísticas y evaluación basada en tareas o escenarios. Triacca et al. en [TRI04] proponen, en cambio, emplearlas simultáneamente, como una forma de sumar los beneficios que ofrecen las dos técnicas.

La evaluación guiada por heurísticas se realiza a partir de un checklist y principios de usabilidad ya establecidos. Por ejemplo, se pueden consultar los aportes de Nielsen [NIE00], donde el checklist puede ser utilizado también como base de un cuestionario que deben responder los usuarios y no sólo como una lista de control de cumplimiento de objetivos o principios de usabilidad, tal como se emplea comúnmente. Una contra de esta técnica es que distintos grupos de expertos, realizando una evaluación sobre el mismo producto software o aplicación Web, suelen partir de heurísticas diferentes

y, en algunos casos, en base a principios o criterios contrapuestos tal como bien señala Ivory en [IVO01].

La evaluación basada en tareas tiene en cuenta la realización de acciones (pasos de una o más tareas) sobre el producto software por parte de un usuario. Las tareas deben estar previamente especificadas y el usuario es observado in situ, mientras se desempeña usando el producto, y eventualmente puede ser registrado por algún medio audiovisual.

Normalmente, estas técnicas se emplean por separado y aisladamente, pero esto no necesariamente tiene que ser así. Triacca et al. [TRI04], proponen adaptar un método de evaluación de usabilidad existente para aplicaciones Web para e-Learning.

Según los autores, una combinación de las dos técnicas podría reducir las limitaciones y potenciar los beneficios de ambos métodos.

En cuanto al método de consulta, es ampliamente conocido un cuestionario con una escala debidamente validada para la Web, denominado WAMMI (Website Analysis and Measurement Inventory) desarrollado por Kirakowski, Claridge y Whitehand [KIR98], que, entre otros aspectos de evaluación, incluye usabilidad. Las respuestas a los ítems del cuestionario son luego agrupadas en seis categorías: atractividad, controlabilidad, eficiencia, utilidad, aprendibilidad y usabilidad global.

2.3.5.3. Prerrequisitos para Diseñar la Medición y Evaluación de Calidad en Uso

Tal como se ha indicado en las secciones 2.2.3 y 2.2.4, las distintas perspectivas de la calidad de producto software se relacionan e influyen recíprocamente, pero es necesario remarcar que eficacia, productividad, seguridad y satisfacción son influenciadas no sólo por usabilidad, funcionalidad, confiabilidad y eficiencia del producto de software, sino también por dos componentes de tipo recurso del contexto de uso: la infraestructura y las metas del usuario.

Para el primer tipo de recurso, esto es infraestructura, se puede considerar por ej.: el equipo en que se desarrollan las tareas, la red y el medio físico de trabajo; para el segundo, metas de usuario, las tareas soportadas por la aplicación y las características del tipo de usuario tales como la experiencia, el nivel de formación, la predisposición a usar el producto o aplicación Web, etc.

Se debe tener en cuenta, además, que no es posible generalizar los resultados de cualquier tipo de evaluación de calidad en uso a otro contexto, o al mismo producto pero con diferentes tipos de usuarios o medioambiente, tal como oportunamente ha resaltado Bevan [BEV99].

Como consecuencia, cuando se diseñan y documentan procesos de medición y evaluación de calidad en uso, según [OLS04b] al menos es necesaria la siguiente información:

- ❖ Descripción de los componentes del contexto de uso incluyendo el tipo de usuario, equipamiento, medio ambiente, y tareas a realizar con el producto (entendiendo tarea a los pasos o submetas llevadas a cabo para alcanzar una meta deseada)
- ❖ Métricas e indicadores de calidad en uso de acuerdo al propósito deseado y la necesidad de información

En el marco de esta tesis, como eje del aporte principal, se utilizará el modelo de calidad en uso reseñado en la sección 2.3.4, basado en un conjunto de métricas e indicadores, soportados e integrados todos por un marco de medición y evaluación con base ontológica [OLS04a], adaptado para calidad en uso. Para llevar adelante el proceso de medición y evaluación se emplearán métodos de inspección –adaptando la metodología Web-QEM [OLS02]- y testing que se discutirán en los capítulos 3 y 4.

Para facilitar el diseño y selección de métricas e indicadores para evaluar calidad en uso es necesario asociar, como primer paso, atributos a las cuatro características: efectividad, productividad, seguridad y satisfacción. La figura 2.4 muestra, a modo de ejemplo, algunos atributos para dos características de calidad en uso: Eficacia y Productividad.

Calidad en Uso

1. Eficacia

1.1 Eficacia de Tarea (ET)

1.2 Completitud de Tareas (CT)

1.3 Frecuencia de Error (FE)

2. Productividad

2.1 Eficiencia en relación a Eficacia (EFt)

2.2 Eficiencia en relación completitud de tareas (EFtRc)

Figura 2.4. Especificación de una instancia del modelo de Calidad en Uso.

2.4. Calidad de Aplicaciones Web

Powell [POW98] afirma que los sitios y aplicaciones Web “*involucran una mezcla entre publicación de contenidos impresos y desarrollo de software, entre marketing y computación, entre comunicaciones internas relaciones externas, y entre arte y tecnología*”¹⁵. Actualmente hay una mayor conciencia y conocimiento en el mundo científico y profesional acerca de la naturaleza multidimensional de sitios y aplicaciones Web, que comprende entre otras disciplinas: computación, arquitectura de información, autoría de

¹⁵ Involve a mixture between print publishing and software development, between marketing and computing, between internal communications and external relations, and between art and technology.

contenidos, navegación, cuestiones estéticas y de presentación, multiplicidad de audiencias, cuestiones éticas y legales, seguridad, rendimiento y ambientes operativos heterogéneos.

2.4.1. Características Distintivas de las Aplicaciones Web

Los sitios y aplicaciones Web como producto o producto en uso (sin hablar acerca de características distintivas del desarrollo Web) tienen sus propias características que son sensiblemente diferentes del software tradicional. Según [POW98], [OLS00b], [MUR01] y [OLS04b] estas características son:

- ❖ Los sitios y aplicaciones Web continuarán siendo guiados por contenido, orientados a documentos. La mayoría de los sitios y aplicaciones Web, aún considerando el creciente soporte hacia funcionalidad y servicios, seguirán destinados a mostrar y entregar información -esto es una característica básica proveniente de los primeros tiempos de la Web que están siendo potenciados actualmente por iniciativas en el área de Web Semántica [DAV03];
- ❖ Los sitios Web son aplicaciones interactivas, centradas en los usuarios, donde las interfaces de usuario juegan un rol central; y continuarán muy enfocadas en cuestiones de aspecto y presentación (look and feel). Las interfaces Web deberían ser fáciles de usar, comprender y operar, ya que potencialmente miles de usuarios con diferentes perfiles y capacidades interactuarán diariamente con ellas. Más aún, en las interfaces de software tradicional podría emplearse un cierto período de tiempo para entrenar a los usuarios en la operación del producto; sin embargo, esto no ocurre del mismo modo con los sitios Web y la dispersión geográfica de los potenciales usuarios;
- ❖ La Web implica una relación mayor entre arte y ciencia que la que se encuentra generalmente en aplicaciones de software. Características estéticas y de presentación de los desarrollos Web no responden únicamente a habilidades técnicas sino también a habilidades creativas y artísticas;
- ❖ Internacionalización y accesibilidad de contenidos para usuarios con capacidades diferentes son desafíos vigentes en el desarrollo de sitios y aplicaciones Web;
- ❖ Búsqueda y navegación son dos tipos de funcionalidad básica para encontrar y explorar documentos y contenidos en un sitio (estas capacidades son heredadas de las aplicaciones hipermediales);
- ❖ La seguridad es una cuestión central en aplicaciones Web orientadas a transacciones; asimismo, el rendimiento es una cuestión crítica

para muchos sitios y aplicaciones Web (aunque también son características críticas para aplicaciones tradicionales);

- ❖ El sitio Web como un todo, o partes de él, son frecuentemente piezas evolutivas de información;
- ❖ El medio donde los sitios y aplicaciones Web son alojados y funcionan, son generalmente más impredecibles que los medios donde corren las aplicaciones tradicionales. Por ejemplo, la impredecibilidad del mantenimiento del ancho de banda, o la disponibilidad del servidor podrían afectar la calidad percibida por los usuarios en relación al sitio Web;
- ❖ La privacidad de contenidos y derechos de propiedad intelectual de los materiales distribuidos en sitios y aplicaciones Web también son cuestiones de interés específico para los involucrados en el desarrollo. La mayoría de las veces es muy difícil establecer límites legales debido a la heterogeneidad de la legislación en diversos países o, aún peor, directamente la ausencia de la misma.

La mayoría de las características mencionadas hacen de los sitios y aplicaciones Web artefactos particulares, sin embargo, tal como cualquier aplicación de software, ellos también involucran código fuente y ejecutable, estructuras de datos persistentes y requerimientos, arquitectura, diseño y especificaciones de testing. Por lo tanto, se puede argumentar que el marco de calidad ISO introducido en capítulos anteriores, que distingue entre los tres enfoques diferentes para especificar requerimientos de calidad de software, esto es, calidad interna, calidad externa y calidad en uso, es aplicable también en gran medida a los productos intermedios y al ciclo de vida de productos Web. A continuación se discute brevemente la validez y alcance de dichos modelos de calidad para sitios y aplicaciones Web.

2.4.2. Perspectivas de la Calidad de Aplicaciones Web

Al igual que cualquier línea de producción de software, el ciclo de vida Web involucra diferentes etapas de sus productos, tanto en fases tempranas como inepción y desarrollo, o en fases tardías, como implantación, operación y evolución. Para asegurar la calidad de los productos, se puede planificar la evaluación y el control de la calidad desde los productos intermedios hasta los productos finales. Así, a la pregunta general acerca de si se pueden aplicar los mismos modelos ISO de calidad interna, calidad externa y calidad en uso, la respuesta inicial es sí. Sin embargo, considerando una pregunta más específica respecto a utilizar las mismas seis características prescriptas para calidad interna y externa y las cuatro características para calidad en uso, la respuesta es sí para la última, pero es necesario hacer algunas otras consideraciones respecto a las primeras.

Particularmente, y tal como se ha señalado al principio de esta sección, la naturaleza de sitios y aplicaciones Web está dada por una mezcla de contenidos (información), funcionalidad y servicios. Se puede argumentar que las seis características prescritas por el estándar ISO 9126-1 (esto es Usabilidad, Funcionalidad, Confiabilidad, Eficiencia, Portabilidad y Mantenibilidad) no están destinadas a especificar requerimientos para calidad de información [OLS04b]. En su columna on-line, Nielsen [NIE04] ha dicho con respecto a contenidos Web para sitios informativos que “*en última instancia, los usuarios visitan el sitio Web por sus contenidos. Todo lo demás es sólo el telón de fondo*”¹⁶.

En sintonía con las consideraciones anteriores y teniendo en cuenta algunos aportes realizados en el área de calidad de información por [ALE99] y [HUA99] entre otros, en [OLS04b] se identifican cuatro subconceptos principales para la característica *Contenido*, la que se propone sea añadida a las seis ya existentes en el estándar ISO/IEC 9126-1, para poder abarcar también la evaluación de requerimientos de calidad de información de una aplicación Web:

- ❖ *Precisión de la Información (Information Accuracy)*. Esta subcaracterística enfoca la naturaleza intrínseca de la calidad de la información. Asume que la información tiene *per se* su propia calidad. La precisión de la información es el grado en el cual la información es correcta, no ambigua, confiable, objetiva y verificable. Si una parte de la información de un sitio Web adquiere fama de inexacta, el sitio Web probablemente se percibirá como de poco valor agregado y se verán reducidas sus visitas;
- ❖ *Adecuación de la Información (Information Suitability)*. Esta característica enfoca la naturaleza contextual de la calidad de la información. Enfatiza la importancia de transmitir la información apropiada de acuerdo a tareas y metas orientadas a los usuarios. Consecuentemente es el grado en el que la información es apropiada (cobertura apropiada para la audiencia destinataria), completa (cantidad suficiente), concisa (si es breve, mejor) y actualizada;
- ❖ *Accesibilidad (Accessibility)*: esta característica resalta la importancia de los aspectos técnicos de los sitios y aplicaciones Web en términos de hacer los contenidos más accesibles para usuarios con diferentes discapacidades;
- ❖ *Conformidad con Normas Legales (Legal Compliance)*: la capacidad de una pieza de información para adherir a estándares, convenciones

¹⁶ Ultimately, users visit your website for its contents. Everything else is just the backdrop

y normas relacionadas a contenidos y derechos de propiedad intelectual;

Además de las categorías anteriores, en [OLS04b] se propone tener en cuenta otros subconceptos relativos a la estructura y organización de la información. Muchas de estas subcaracterísticas tales como comprensibilidad global (implementada por mecanismos que ayudan a comprender rápidamente la estructura y contenidos del espacio de información del sitio Web tales como tablas de contenido, índices y mapa del sitio), aprendibilidad, y aún internacionalización, pueden relacionarse a la característica *Usabilidad*.

Finalmente, desde la perspectiva de calidad en uso, para la característica satisfacción, que puede ser implementada para propósitos de medición como un cuestionario, pueden ser incluidos ciertos ítems relacionados a contenido como también para navegación, estética, funciones, etc. Además, para otras características de calidad en uso como *Efectividad* y *Productividad*, pueden diseñarse e implementarse ciertas tareas específicas, obviamente orientadas a los usuarios, que involucren la realización de acciones en base a contenidos y funcionalidad.

2.5. Conclusiones

Las discusiones presentadas en las secciones iniciales (2.1, 2.2 y 2.3) acerca del significado y las relaciones entre términos como calidad y calidad en uso permiten concluir, inicialmente, en que calidad no es un concepto absoluto y que existen diferentes perspectivas sobre calidad, teniendo en cuenta un producto de software o bien un producto de software en uso. Tanto calidad interna como calidad externa y calidad en uso, como se enfoca en el estándar ISO/IEC 9126-1, pueden especificarse, medirse y finalmente evaluarse considerando que cada perspectiva tiene su valor agregado en una estrategia de aseguramiento de calidad y a lo largo de todo el ciclo de vida de un producto software. Sin embargo, el objetivo final es alcanzar la calidad en uso.

También se ha podido considerar la evolución en la teoría y la práctica del concepto calidad en uso, que partió de una idea sencilla pero discutible como *facilidad de uso* y se arribó, aunque no podría decirse que en forma definitiva, a un modelo –y a una definición también- de calidad en uso más abarcativa. En esta línea se ha discutido también una analogía entre los significados de calidad en uso y usabilidad, dado que éste último sigue usándose frecuentemente en la literatura y en la industria como sinónimo de calidad en uso.

La importancia de esta discusión radica en que si sólo se toman en cuenta las definiciones respectivas en el estándar ISO/IEC 9126-1 el significado y la dimensión de cada uno dista bastante de ser el mismo. En esta norma, calidad en uso es una de las tres perspectivas de calidad, o un *concepto calculable* al más alto nivel teniendo en cuenta el modelo de Métricas

e Indicadores, tal como se verá en el capítulo 3. *Usabilidad*, en cambio, es una de las características de calidad interna y calidad externa de un producto software, es decir que se ubica en un nivel jerárquico inferior.

Como corolario de este capítulo se presentaron algunos prerrequisitos para diseñar y evaluar calidad en uso (tópico 2.3.5.3) y precisiones acerca de qué se entiende por calidad de aplicaciones Web (2.4), donde se señala el hecho de que el diseño y la evaluación de calidad en uso de un producto software es una cuestión que puede ser abordada desde múltiples puntos de vista, con la contribución de diferentes disciplinas. Desde este punto de vista guarda una estrecha relación con los enfoques que apoyan definiciones de Ingeniería Web, como las presentadas por Murugesan y otros en [MUR01], y puede considerarse que los requerimientos obligatorios de calidad en uso de las aplicaciones Web sólo serán satisfechos si se especifican y “construyen” desde etapas tempranas del proceso de desarrollo; preferentemente en base a un modelo de calidad predefinido y consensuado de antemano entre los interesados del proyecto.

Los conceptos de las reseñas de las dos últimas secciones, junto al modelo de métricas e indicadores detallado en el capítulo 3, serán pilares para el caso de estudio presentado en el capítulos 5 y 6, donde se pone en práctica la propuesta central de esta tesis, acerca de cómo abordar, en forma sistemática y disciplinada, la evaluación de calidad en uso de sitios y aplicaciones Web con base en un marco ontológico.

Parte II: Marco Conceptual y Método de Medición y Evaluación

Capítulo 3: Un Marco Conceptual de Medición y Evaluación Para Calidad en Uso

3.1 Introducción

En sección 2.3.2 se discutió acerca de que el concepto calidad, y en particular respecto a un producto software, no tiene un carácter absoluto ni puede ser abarcado desde una única perspectiva.

Además, la idea subyacente en la discusión es que un producto software no sólo debe cumplir con requerimientos explícitos e implícitos de “fabricación”, sino fundamentalmente con las expectativas de usuarios específicos, en contextos reales y condiciones de uso bien definidos. Esto implica que, cualquiera sea la perspectiva desde la que se desee medirla: interna, externa, en uso, la calidad no puede medirse directamente, es decir de un modo sencillo y directo [OLS04b], excepto que sea de un modo subjetivo y trivial.

Este problema ha sido advertido y discutido en [OLS04a] y por eso, en el contexto del diseño de un sistema de catalogación de métricas e indicadores de software, proponen una ontología que sirva como cimiento para la conceptualización del dominio, donde los conceptos claves, los atributos y sus relaciones son modelados y definidos explícitamente.

Como motivación para la propuesta de la ontología, los autores consideran que si bien han sido publicados a la fecha un conjunto de modelos de medición y evaluación de calidad de producto software, tales como el ya citado [ISO9126-1] y el estándar [ISO15939], y que por otro lado son reconocidos los trabajos de Zuse [ZUS98] -formalización de un marco para la medición de calidad- y Kitchenham [KIT01] -modelado de datos de mediciones de software-, no existe en la literatura ni en la práctica, un consenso general en el uso de los términos clave utilizados.

Por lo tanto, llegado el momento de implementar un sistema de medición y evaluación de calidad, en particular de calidad en uso tal cual es el objetivo de esta tesis, es imprescindible definir con rigor el significado de los conceptos involucrados y sus relaciones. De no ser así, sería difícil establecer claramente cuáles son las metas y objetivos de la evaluación, y tampoco se sabrá qué medir y cómo llevar adelante los pasos del proceso de medición y evaluación. Por ejemplo, es necesario establecer claramente la entidad que será objeto de la medición y evaluación –producto, producto en uso, etc.-, la necesidad de información que le da origen y también el punto de vista del usuario desde el que se enfoca el estudio.

La literatura relevante en relación a la ontología y al modelo de métricas e indicadores proviene de los trabajos recientes [MAR03], [MAR04] y [OLS05a], además del citado [OLS04a], a partir de los que se realizó una sinopsis que se incluye en las secciones 3.2 y 3.3.

El capítulo 3 está organizado de la siguiente manera: luego de esta introducción se encuentra una sinopsis de una ontología de métricas e indicadores para la medición y evaluación de software y aplicaciones Web (sección 3.2); más adelante, en la sección 3.3, se presenta el marco de medición y evaluación que sustenta la propuesta de medición y evaluación sistemática de calidad en uso. Finalmente, la sección 3.4 corresponde a las conclusiones.

3.2 Ontología para el Modelo de Métricas e Indicadores

La importancia de una ontología en el ámbito de la medición y evaluación de producto software radica en que permite un acuerdo básico entre los principales usuarios del sistema de calidad –expertos, desarrolladores, gerentes, vendedores, compradores- sobre la terminología y significado de conceptos clave tales como: medida, medición, métrica, atributo, método de medición, escala, indicador, etc.

La percepción de los investigadores, al momento de estudiar el problema, era que existía una cierta profusión de información que no facilitaba precisamente el consenso sobre términos clave relacionados a procesos de evaluación, estimación y aseguramiento de calidad en general. Esa falta de acuerdo significaba también una dificultad adicional para la aplicación y reuso de los estándares.

Como ejemplo ilustrativo en [MAR04] se menciona que el término “métrica” es empleado en los estándares [ISO9126-1] e [ISO14598-1], pero no en el [ISO15939]. Aún más, en los estándares [ISO9126-1] e [ISO14598-1] se emplean los términos “medida directa” y “medida indirecta”, mientras que en el [ISO15939] se emplea el término “medida base” y “medida indirecta”. En algunos casos se percibe que ciertos términos son empleados como sinónimos pero en otros claramente no. Tal es el caso del concepto “métrica”, que en el [ISO14598-1] está definida como “*el método de medición definido y la escala de medición*”¹⁷, y no existe en el [ISO15939] un término con un significado exactamente igual.

También es ostensible la ausencia de ciertos términos en la mayoría de los estándares referidos, tales como “indicador elemental” o “indicador global”, como también de “método de cálculo”. El término indicador se emplea en [ISO15939], pero con un significado menos preciso al asignado en el contexto de la definición de la ontología de métricas e indicadores, tal como se verá en el tópico 3.3.1.3. Por otro lado, para los autores el objetivo cuando se usa un método de medición es diferente al de usar un método de cálculo. El primero está definido con el objetivo de llevar adelante una actividad de medición de una métrica directa y el segundo para una actividad de cálculo de métricas indirectas e indicadores.

¹⁷ “The defined measurement method and the measurement scale”

En el contexto de la especificación de la ontología de métricas e indicadores se establecen claramente el *dominio* a modelar, el *propósito*, el *alcance* y las fuentes de conocimiento utilizadas. A continuación, a partir de una síntesis del trabajo [MAR04], se detallan las especificaciones para los tres primeros tópicos.

Dominio modelado: Elementos para la medición y/o estimación de artefactos de software (recursos, productos, procesos) para satisfacer distintas necesidades de información.

Propósito de la ontología:

- ❖ Facilitar el intercambio de información relacionada a métricas e indicadores de software entre los ingenieros, gerentes, usuarios y demás participantes en proyectos de medición y evaluación de artefactos software.
- ❖ Conformar la base de conocimiento para el diseño y construcción de un sistema de catalogación de métricas e indicadores [MAR03] que facilite la documentación, recuperación y búsqueda semántica de toda la información relacionada, sirviendo de apoyo a los procesos de aseguramiento de calidad, definición y especificación de requerimientos no funcionales, definición de test de calidad, entre otros.
- ❖ Conformar la base de conocimiento para el diseño y construcción de un marco de medición y evaluación.

Alcance: El alcance de la ontología comprende todos los conceptos, atributos y relaciones necesarios para la catalogación y explotación de la información relacionada a métricas e indicadores en el ámbito de Ingeniería de Software e Ingeniería Web. La ontología no incluye la especificación de instancias de los conceptos definidos.

(En el apéndice B se presentan detalladamente los conceptos que intervienen en la ontología de métricas e indicadores).

3.3 Marco de Medición y Evaluación

Un marco de medición y evaluación consistente debe estar construído sobre una base conceptual claramente especificada. Para el marco de medición y evaluación adoptado, la base ontológica permite explicitar, sólida y formalmente, los principales conceptos del dominio de métricas e indicadores, como así también sus propiedades, relaciones y axiomas [OLS05a] [OLS05b].

También se basa en la suposición de que para que una organización pueda medir y evaluar con un propósito definido, debe primero especificar los requerimientos no funcionales, comenzando por las necesidades de

información, luego debe definir un conjunto de métricas útiles para el propósito de la medición y, finalmente, interpretar las métricas por medio de indicadores contextuales, con el objetivo de evaluar o estimar el grado de cumplimiento de los requerimientos especificados.

Los componentes principales del marco de medición y evaluación son:

- a) La definición y especificación de los requerimientos no funcionales
- b) El diseño y ejecución de la medición
- b) El diseño y ejecución de la evaluación
- c) La definición del proyecto en sí mismo

Cada componente tiene su soporte en los conceptos de la ontología de métricas e indicadores.

A continuación se ilustran las definiciones de los conceptos clave con un ejemplo concreto para cada uno de ellos, orientado a medir y evaluar calidad en uso de una aplicación Web.

3.3.1 Conceptos

La ontología de métricas e indicadores constituye para esta tesis el eje principal para la evaluación sistemática y disciplinada de calidad en uso. Incluye la definición de conceptos clave tales como concepto calculable, entidad, atributo, indicador, etc.; considerando inicialmente que para un proyecto de medición y evaluación de calidad es necesario conocer cual es la *necesidad de información* que le da origen.

3.3.1.1 Necesidad de Información, Modelo de Concepto y Atributos

Necesidad de Información se define como la visión y el conocimiento necesario para administrar objetivos, metas, riesgos y problemas. Básicamente remite a dos puntos de partida: metas que la organización desea alcanzar o bien riesgos y/o problemas que impiden alcanzar las metas deseadas. Una relación que se visualiza en la figura 3.1 para la clase *Necesidad de Información* es con la clase *Concepto Calculable*.

Un *Concepto Calculable* es una relación abstracta entre *atributos* de entidades y una *Necesidad de Información*; por ej. productividad, calidad y calidad en uso son instancias de concepto calculable. De acuerdo al nivel de abstracción al que esté definido, un *Concepto Calculable* puede, a su vez, estar compuesto por otros subconceptos, y podría así ser representado por un *Modelo de Concepto*. Por ejemplo, en el estándar [ISO9126-1] se especifica un modelo de calidad en uso basado en cuatro características (subconceptos).

Una *Entidad* es un objeto, tangible o intangible que puede ser caracterizado por un conjunto de *atributos* mensurables. En el contexto de esta tesis la entidad evaluada es una aplicación Web del dominio de e-Learning, o sea un producto software particular.

Un *Atributo* es una propiedad mensurable de una entidad, y si bien una entidad puede tener muchos atributos, sólo se emplea un subconjunto que es útil para el concepto calculable especificado, conforme a la necesidad de información. Cada atributo puede ser cuantificado por una o más métricas, pero en el contexto de un proyecto concreto debe seleccionarse una única métrica para cada atributo.

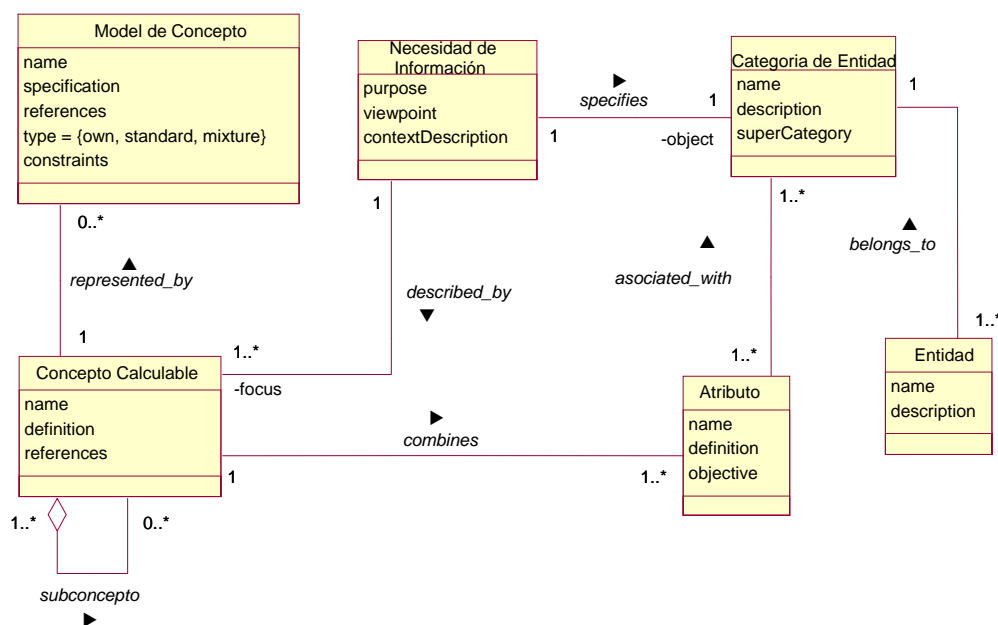


Figura 3.1. Conceptos involucrados en la fase de Definición y Especificación de Requerimientos.

3.3.1.2 Métricas y Medición

Por su parte, una métrica contiene la definición de un *método de medición* o un *método de cálculo* y la *escala* asociada, tal como se aprecia en la figura 3.2. El *método de medición* es la secuencia lógica particular de operaciones y posibles heurísticas, especificada para permitir la realización de la descripción de una métrica por una actividad de medición. Por otro lado, la *escala* se define como un conjunto de valores con propiedades definidas .

La propiedad más importante de una escala es su tipo, considerando que puede ser *Catagórica* o *Numérica*. A su vez, dependiendo de la naturaleza de la relación entre los componentes de la escala, pueden

clasificarse en: *nominal, ordinal, intervalo, proporción o absoluta*. El tipo de escala de los valores medidos define las transformaciones admisibles y afecta las operaciones matemáticas y estadísticas que pueden ser aplicadas.

Las métricas pueden ser directas, sobre las que puede aplicarse un método de medición (objetivo o subjetivo); o indirectas, que son aquellas definidas en función de otras métricas y se calculan en base al método de cálculo asociado, es decir en base a una fórmula. Formalmente una métrica m representa un mapeo $m:A \rightarrow X$, entre una variable A perteneciente al mundo empírico -un atributo de una categoría de entidad- y una variable X que pertenece al mundo real -a la que pueden asignarse valores numéricos o categóricos.

Luego, para cada métrica definida es necesario llevar a cabo el *proceso de medición*, o sea la actividad que, empleando la definición de la métrica, permite obtener valores de medidas.

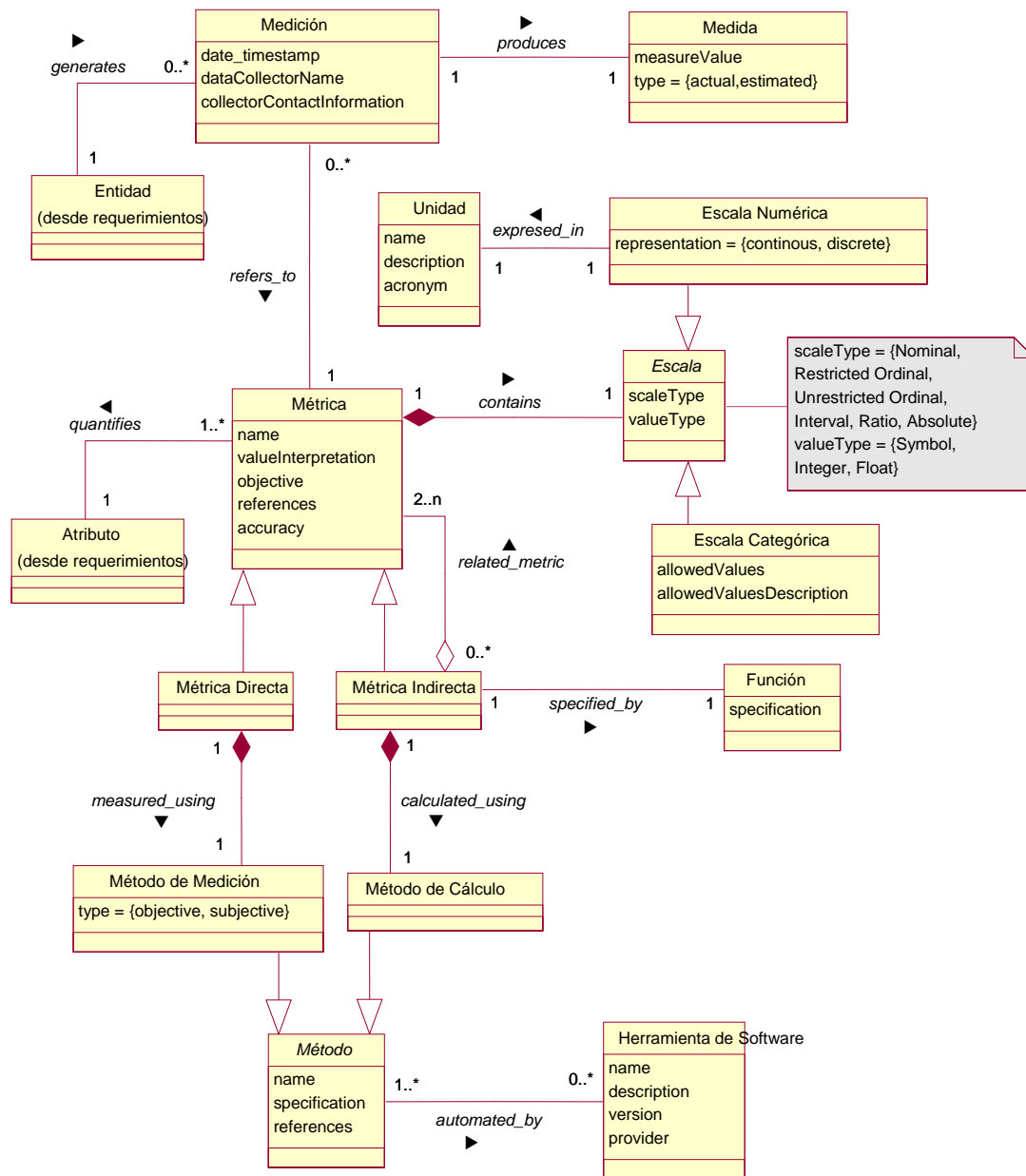


Figura 3.2. Conceptos involucrados en la fase de Diseño e Implementación de la Medición.

Sin embargo el valor de una métrica no representará por sí mismo el nivel de satisfacción de un requerimiento elemental, por ello es necesario definir un nuevo mapeo, ahora entre dos variables del mundo formal $X \rightarrow I_e$, una de ellas (X) representando a la métrica y la otra (I_e), a la que pueden asignarse valores numéricos o categóricos, representando la interpretación que se hace de la métrica, e implicando criterios de decisión

respecto a los requerimientos previamente establecidos. Esta nueva variable se denomina *Indicador Elemental* en el modelo de métricas e indicadores.

3.3.1.3 Indicadores y Evaluación

El término indicador está definido como el método de cálculo y la escala que, junto al modelo y criterios de decisión, permiten obtener una estimación o evaluación de un concepto calculable con respecto a una necesidad de información definida. Un indicador elemental es aquel que no depende de otros indicadores para evaluar o estimar un concepto a un nivel de abstracción bajo. Adicionalmente, un *indicador parcial o global* está definido como aquel que, derivado de otros indicadores, permite evaluar o estimar un concepto a un nivel de abstracción más alto. Tal como ocurre con las métricas indirectas, los indicadores poseen una escala y un método de cálculo. Estas relaciones se pueden apreciar en la figura 3.3.

Para poder establecer el mapeo o correspondencia entre una métrica y un indicador es necesario considerar, teniendo en cuenta la necesidad de información de un usuario, un *Modelo Elemental* más ciertos *Criterios de Decisión*.

Después de seleccionar un modelo de puntaje para indicadores globales, es necesario realizar el proceso de agregación, tal como se define en la estructura jerárquica del modelo de concepto. Por lo tanto, aplicando un mecanismo de agregación paso a paso entre bloques del mismo nivel se obtiene un esquema global; así, este modelo permite obtener valores de *indicadores parciales y globales* en cada paso de su ejecución. El *indicador global* representa finalmente el grado de satisfacción general en el cumplimiento de los requerimientos especificados, desde el punto de vista de un usuario específico.

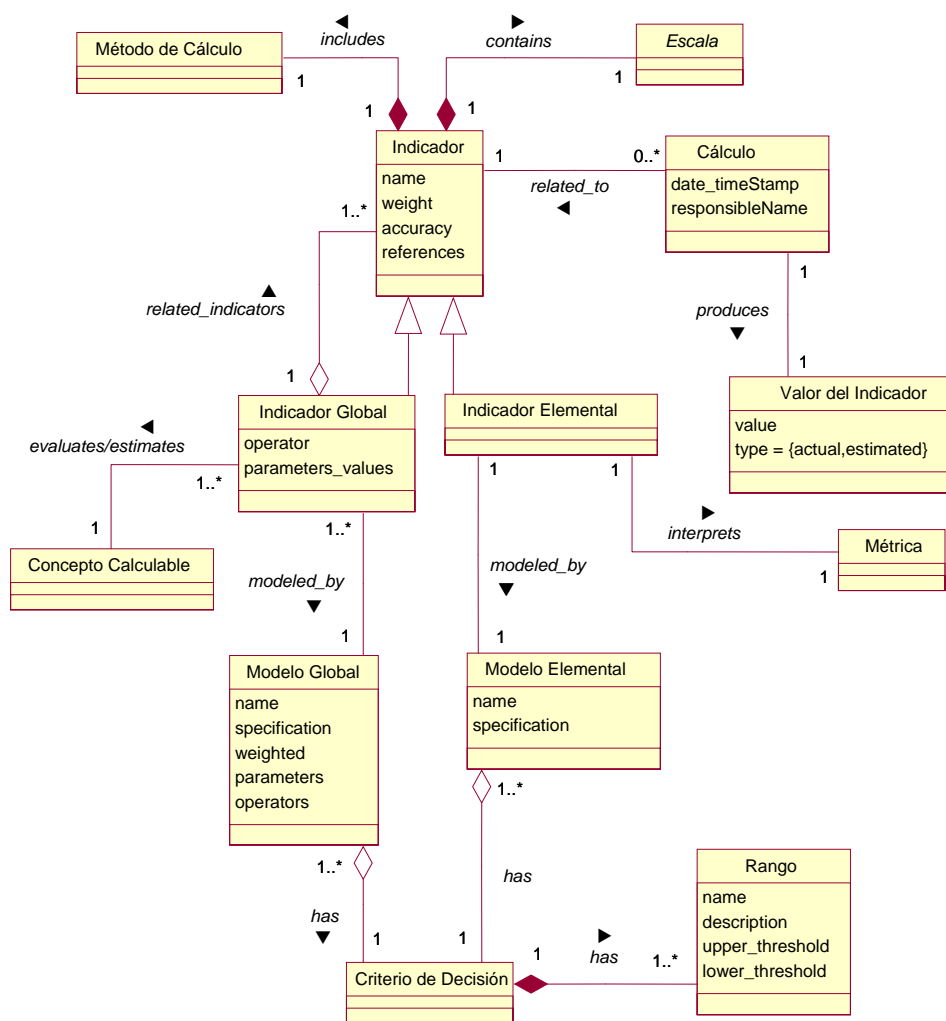


Figura 3.3. Conceptos involucrados en la fase de Diseño e Implementación de la Evaluación

3.3.2 Ejemplos de los Principales Términos para Calidad en Uso

Los ejemplos que se presentan a continuación tienen como objetivo ilustrar los conceptos más importantes del modelo de métricas e indicadores. Estos ejemplos provienen de un caso de estudio para calidad en uso de una aplicación Web, en el dominio de e-learning, que se presenta en detalle en el capítulo 5.

Para este ejemplo, la *necesidad de información* es: conocer la percepción de calidad en uso que tienen los alumnos ingresantes a la Facultad de Ingeniería de la UNLPam (año 2004) al usar una aplicación e-learning que da soporte al curso nivelatorio denominado “Introducción a la Matemática”. Para el término *concepto calculable*, la instancia en este caso

es “calidad en uso”, cuya definición se precisó en el capítulo 2. A su vez, calidad en uso posee *subconceptos* (“características” en [ISO9126-1]), tales como “Eficacia”, “Productividad”, “Seguridad” y “Satisfacción”. Además, el concepto calculable puede ser representado por un *modelo de concepto*. El modelo de concepto de referencia aquí es un modelo de calidad en uso basado en la propuesta de calidad en uso del estándar [ISO9126-1], tal como se ilustró en la figura 2.2 y se instanció en la figura 2.4.

Cada concepto posee uno o más *atributos*, los que denotan propiedades relevantes para la necesidad de información definida inicialmente. Ejemplos de atributos son “Eficacia de Tarea” y “Complejidad de Tarea”, relacionados ambos al subconcepto “Eficacia”. Por su parte, cada atributo es mensurable si está cuantificado por medio de una *métrica* apropiada y adecuadamente definida. Para el atributo “Complejidad de Tarea” se diseñó una métrica indirecta denominada “Proporción de tarea completada correctamente”, que permite conocer qué proporción de una tarea fue completada correctamente por un usuario. La *Fórmula* es: $ET = 1 - |\sum A_i|$.

Donde A_i es la parte proporcional de la tarea no realizada o incorrecta y es una métrica directa. Como se trata sólo de la proporción de una tarea realizada por un usuario, la métrica final que debe utilizarse es el “Promedio de la proporción de tareas completadas por todos los usuarios”.

Anteriormente (tópico 3.3.1.2) se dijo que una métrica tenía una escala y, si era indirecta, incluía un método de cálculo. En el ejemplo anterior, la métrica ET es numérica, con un valor “real”, y el tipo de escala es proporción. Como es una escala numérica está expresada en una determinada unidad, cuya descripción es: “Proporción de una Tarea Completada Correctamente por un Usuario”.

Para “Promedio de la Proporción de Todas las Tareas Completadas Correctamente por Todos los Usuarios” la fórmula es: $P_{ETT_u} = \frac{\sum_{j=1}^{j=n} P_{ET1u}(j)}{n}$

Donde n es el número total de usuarios intervinientes en el test y P_{ET1u} la métrica indirecta “Promedio de la proporción de tareas completadas por un usuario”.

Obsérvese que A_i es una métrica directa, y puede ser computada en forma “objetiva”, empleando un *método de medición* sencillo como una regla de conteo, facilitado por la definición explícita de las metas y submetas de cada una de las tareas que se propongan a los usuarios. Si bien las métricas son imprescindibles para el proceso de medición, es necesario contar con los indicadores correspondientes para llevar adelante el proceso de evaluación.

Para la métrica P_{ETTu} , el indicador elemental tiene por nombre “Grado de Cumplimiento respecto a proporción de tareas completadas por todos los usuarios” ($\%P_{ETTu_P}$). La especificación del *modelo elemental*, o sea el algoritmo o función que tiene asociado a su vez un criterio de decisión para este indicador es: $\%P_{ETTu_P} = P_{ETTu} \times 100$.

El criterio de decisión que puede tener el modelo de un indicador está constituido por los niveles de aceptación establecidos dentro de rangos dados por la escala. Para este caso podría ser “No satisfactorio” si $0 \leq \%P_{ETTu_P} \leq 45$, “Marginal (Regular)” si $45 < \%P_{ETTu_P} \leq 70$, o “Satisfactorio” si $70 < \%P_{ETTu_P} \leq 100$. La interpretación para “No satisfactorio” es que deben realizarse cambios en forma urgente, “Marginal” en cambio indica que deben realizarse mejoras a corto plazo.

3.4 Conclusiones.

En este capítulo se ha presentado sintéticamente un marco de medición y evaluación que, basado en una ontología sobre los términos clave implicados, permite un enfoque ingenieril de la evaluación de calidad en uso. O sea que facilita y promueve, definiendo y modelando los conceptos y sus relaciones, el empleo sistemático, disciplinado y cuantificable de principios y prácticas reconocidas de ingeniería a la medición y evaluación de calidad.

La ontología de métricas e indicadores tiene como propósitos principales: facilitar la comunicación entre los interesados de un proyecto de medición y evaluación, sentar la base de conocimiento para un sistema de catalogación de métricas e indicadores con potencia semántica y, también, conformar la base de conocimiento para el diseño y construcción de un marco de medición y evaluación más amplio.

En la sección sobre el marco de medición y evaluación se encuentran las definiciones de los conceptos más importantes involucrados en un proceso de medición y evaluación, como también las relaciones que los vinculan. Son particularmente relevantes las definiciones precisas de aquellos conceptos involucrados en las discusiones que dieron origen al modelo tales como *métrica* e *indicador*. Sin embargo, es necesario destacar que el punto de partida de cualquier proyecto de medición y evaluación es una *necesidad de información* dada, concepto que tiene a su vez una relación directa con *concepto calculable*, el que articula la relación entre atributos de entidades a medir y necesidad de información.

Finalmente, en la sección 3.3.2, el modelo de métricas e indicadores se ejemplifica considerando el modelo de calidad en uso definido en [ISO9126-1], pero visto como *modelo de concepto*, o sea compuesto por subconceptos o conceptos calculables a un nivel de abstracción más bajo. El ejemplo se relaciona con el caso de estudio cuyo diseño se detalla en el capítulo 5 y la

ejecución en el capítulo6. El caso de estudio puede ser considerado como una instancia particular del modelo de métricas e indicadores.

Capítulo 4: El Método de Evaluación

4.1. Introducción

En este capítulo se hace referencia a la metodología de evaluación de calidad que se ha empleado en el marco de esta tesis. Para contextualizar esta referencia se retoman brevemente algunos conceptos presentados en el capítulo 2, en relación a posibles estrategias de evaluación. Específicamente es necesario hacer notar que se suelen utilizar dos conceptos relacionados, tales como evaluación cualitativa o evaluación cuantitativa.

Los métodos y técnicas centradas en la evaluación cualitativa se basan generalmente en una lista de características (atributos o guías), a ser analizadas para un producto o productos competitivos. La lista puede contener características de distinto tipo (relacionadas al ente a evaluar como producto, proceso o recurso), tales como características técnicas, de costo, etc. Luego del proceso de evaluación, en la etapa de análisis y conclusiones, quienes toman las decisiones frecuentemente crean para cada sistema (a comparar) una lista de ventajas y desventajas. Finalmente, mediante un mecanismo intuitivo, comparando las ventajas y desventajas para cada característica de cada sistema, arriban a un puntaje final. Este enfoque es obviamente conveniente y atractivo cuando el objeto de la evaluación y el proceso de decisión es suficientemente simple.

Un ejemplo de esta técnica es la evaluación heurística que se encuentra en los trabajos [NIE93] [NIE04]. Es una técnica efectiva en costos pero no muy precisa, ya que en un proceso de evaluación (y eventualmente de comparación y selección) en donde intervienen e inspeccionan, por ejemplo, más de cuarenta características y atributos para cada ente seleccionado, y en donde se identifican distintas relaciones entre los mismos, carece de las propiedades de precisión y justificación objetiva necesarias.

A su vez, se podría hacer una clasificación de tipos de métodos y técnicas, conforme a si corresponden a la categoría (o clase) de testing, inspección, consulta, simulación o modelado analítico. Por ejemplo, la metodología WebQEM (Web Quality Evaluation Methodology) [OLS02] corresponde a un método de inspección de características y atributos de calidad (a partir de un modelo de calidad), en tanto que evaluación heurística corresponde a la misma categoría de inspección, pero en este caso usando guías heurísticas como técnica.

Concretamente, WebQEM es una metodología cuantitativa centrada en expertos que fue desarrollada inicialmente para evaluar la calidad de producto (particularmente ha sido empleada para evaluar sitios o aplicaciones Web) y utiliza un modelo de calidad jerárquico basado en [ISO9126-1] y

adaptado a necesidades (subcaracterísticas) propias de la Web. Esta metodología ha sido empleada en varios casos de estudio [OLS99], [OLS00a], [COV02] y en proyectos de evaluación de sitios de empresas. Específicamente en relación a un proyecto de evaluación, se podría complementar el uso de WebQEM con otras técnicas como evaluación heurística y estrategias de prueba con usuarios para evaluar calidad en uso, por citar sólo algunas.

El capítulo 4 está organizado de la siguiente manera: en la sección siguiente (4.2) se ofrece una perspectiva de la metodología WebQEM, sus principales características, objetivos y una síntesis de las fases y pasos que comprende, con ejemplos tomados de casos de estudio en la que fue empleada. La sección 4.3 se refiere a las posibilidades de adaptar WebQEM para estudios de calidad en uso y en la sección 4.4 se encuentran las consideraciones finales del capítulo.

4.2. Panorama de la Metodología WebQEM

La metodología WebQEM ha sido desarrollada a partir de mediados del 98 por Olsina [OLS00b], con el propósito de aportar una estrategia eficaz, centrada en expertos, para evaluar y analizar la calidad de sitios y aplicaciones Web en general. Está basada en un modelo jerárquico de requerimientos de calidad, partiendo de las características (subconceptos según el término empleado en 3.3.1.1) de más alto nivel prescriptas en la norma [ISO9126-1], a saber: *Usabilidad, Funcionalidad, Confiabilidad, Eficiencia, Mantenibilidad y Portabilidad* para calidad interna y externa; y *Eficiencia, Productividad, Seguridad y Satisfacción* para calidad en uso. De este modo, a partir de esas características se derivan subcaracterísticas y, a partir de éstas, siguiendo un proceso de descomposición jerárquico, se especifican (asocian) atributos. A este modelo de calidad se le ha dado en llamar mixto [OLS00b], en el sentido que parte de lo prescripto en la norma ISO, pero es personalizado a nivel de subcaracterísticas y atributos.

El principal objetivo de WebQEM es evaluar y determinar el nivel de cumplimiento de las características requeridas para un problema dado, para lo cual se analizan los indicadores (también llamados criterios de preferencia o de performance) globales, parciales y elementales obtenidos. El resultado del proceso de evaluación (y eventualmente de comparación y selección) puede ser interpretado como el grado de satisfacción de los requerimientos de calidad. Es posible sintetizar a WebQEM en los siguientes enunciados:

- ❖ Permite evaluar (y eventualmente comparar y seleccionar) la calidad o la calidad en uso de productos Web.
- ❖ Basa los requerimientos no funcionales en modelos de calidad (modelo ISO 9126, modelo mixto o modelo propio) a los que se le

asocian atributos cuantificables por medio de métricas directas e indirectas [MAR03] [OLS04a].

- ❖ Especifica criterios (modelo elemental o funciones de preferencia) para pasar de valores de métricas a indicadores.
- ❖ Especifica modelos de agregación de características, subcaracterísticas y atributos. Así, permite calcular indicadores parciales y globales, a partir de indicadores elementales.
- ❖ Se centra en juicio de evaluadores expertos antes que en juicio de usuarios finales.
- ❖ Permite trazabilidad en el proceso de evaluación.
- ❖ Permite realizar informes de recomendaciones (fortalezas y debilidades) a partir de los datos de métricas e indicadores elementales, parciales y globales.

La metodología comprende una serie de fases y actividades, y una serie de métodos, modelos y herramientas para llevarlas a cabo. Estas fases son:

- ❖ *Planear y programar* la evaluación de calidad.
- ❖ *Definir y especificar* los requerimientos de calidad.
- ❖ Definir (*diseñar*) e implementar (*ejecutar*) la evaluación elemental.
- ❖ Definir (*diseñar*) e implementar (*ejecutar*) la evaluación global.
- ❖ *Analizar* los resultados, *concluir* y *documentar* (generar el informe de recomendación).

A continuación se realiza una breve explicación de cada una de las fases:

Planear y programar la evaluación de calidad: Se refiere a actividades y procedimientos de soporte, con el fin de determinar objetivos estratégicos, tácticos y operativos. Permite establecer las principales estrategias y metas del proceso en un contexto organizacional, seleccionar un proceso de evaluación, programar y asignar métodos, agentes y recursos a las actividades, y realizar nuevas planificaciones una vez en marcha el proceso de evaluación.

Definir y especificar los requerimientos de calidad: Trata de actividades y modelos para la identificación, determinación, análisis y especificación de los requerimientos. A partir de un proceso de medición y evaluación orientado a metas, con el fin de analizar, evaluar, comparar, y mejorar características y atributos de aplicaciones Web, se establecen los requerimientos que deben responder a necesidades explícitas e implícitas de un perfil de usuario para un dominio dado. El producto de esta fase es un documento que jerárquicamente especifica todas las características y atribu-

tos cuantificables que modelan a la calidad según las necesidades del usuario.

Definir e implementar la evaluación elemental: Tiene relación con actividades, modelos, técnicas y herramientas para determinar métricas y criterios de evaluación para cada atributo cuantificable. Se consideran funciones para determinar indicadores elementales a partir de métricas, rangos de aceptabilidad, entre otros asuntos. Una vez definidos y relacionados los criterios para medir cada atributo, se debe ejecutar el proceso de recolección de datos, computar las métricas e indicadores elementales y, por último, documentar los resultados encontrados.

Definir e implementar la evaluación global: Comprende las actividades, modelos, y herramientas para determinar los criterios de agregación de los indicadores (preferencias) de calidad elemental para producir la preferencia global, para cada sistema seleccionado. Se consideran tipos de modelos de agregación (por ej. modelo aditivo, Logic Scoring of Preference (LSP) [DUJ96], entre otros) y según el modelo, se pueden representar diferentes relaciones entre atributos y características. El modelo LSP, por ejemplo, permite modelar relaciones de reemplazabilidad, simultaneidad, neutralidad y diferentes niveles de polarización. Una vez definidos y consensuados los criterios, se debe llevar a cabo el proceso de cálculo.

Analizar los resultados, concluir y documentar: Trata de actividades de análisis y comparación de los indicadores de calidad elementales, parciales y globales, y de la justificación de los resultados. Se utilizan herramientas y mecanismos de documentación para facilitar la interpretación de los datos y su seguimiento, tal es el caso de WebQEM_Tool [OLS01] y la herramienta en construcción INCAMI_Tool [MOL05], de soporte al marco de medición y evaluación.

4.2.1 Pasos Básicos de WebQEM

Los pasos mínimos definidos para aplicar la metodología WebQEM se podrían resumir de la siguiente manera:

- ❖ Definición del Alcance de la Evaluación, en relación a:
 - *Metas de la evaluación y*
 - *Perfil de la audiencia* (o usuario)
- ❖ Definición y Especificación de los Requerimientos de Calidad
- ❖ Definición de Criterios de Evaluación Elemental
 - *Proceso de medición*
 - *Obtención de Indicadores (preferencias) Elementales*
- ❖ Definición de Criterios de la Evaluación Global
 - *Modelos de Agregación de Indicadores*
 - *Cálculos*

❖ Análisis y Recomendaciones

Paso 1. *Definición de las metas de evaluación y selección del perfil de usuario.* Los evaluadores deben definir las metas y establecer el alcance del proyecto de evaluación Web. La evaluación puede llevarse a cabo tanto en la fase de desarrollo como en la fase operativa del proyecto, y se puede valorar la calidad de un producto completo o bien se puede valorar la calidad de un conjunto de características y atributos de un componente. Los resultados podrán ser utilizados para comprender, mejorar, o controlar la calidad de los productos.

Por otra parte, la importancia relativa (los pesos) de las características y atributos dependen de la meta de evaluación, del perfil de usuario seleccionado y del dominio de la aplicación. Para propósitos de evaluación en dominios Web, se han considerado tres perfiles de usuario a un alto nivel de abstracción, a saber: visitantes, desarrolladores y gerenciadorees. Siguiendo un mecanismo de descomposición es posible, por ejemplo, dividir a la categoría visitante en clases más específicas, tal como se aprecia en la figura 4.1.



Figura 4.1. *Clasificación de Visitantes de Sitios Web.*

Un visitante intencional se define como a la audiencia que tiene al menos algún conocimiento o manifiesta algún interés en un dominio y sitio/s específico/s, y desea probablemente informarse o aprender más acerca de sus contenidos y servicios. Su permanencia en el sitio es generalmente mayor que la de una audiencia casual. A su vez, se podría realizar una clasificación entre visitantes anónimos o registrados.

Paso 2. *Definición y Especificación de los requerimientos de calidad.* Los evaluadores deben elicitar, acordar y especificar los atributos y características de calidad que van a estar presentes en el proceso, agrupándolos en un árbol de requerimientos (modelo de calidad). De las características ISO antes mencionadas se derivan las subcaracterísticas y de éstas se pueden especificar atributos con un mínimo solapamiento. A cada atributo cuantificable del dominio empírico se lo cuantifica por medio de una métrica, que dará como resultado una medida en el mundo fomal.

Medición y Evaluación de Calidad en Uso de Aplicaciones Web

- 1. Usability
 -
- 2. Functionality
 -
- 2.3. Specific Functionality and Contents for E-learning
 - 2.3.1. Course Features
 - 2.3.1.1 General Information
 - 2.3.1.1.1 Descriptive Information
 - 2.3.1.1.1.1 Course Code and Name
 - 2.3.1.1.1.2 Intended Audience
 - 2.3.1.1.1.3 Contents Summary
 - 2.3.1.1.1.4 Glossary
 - 2.3.1.1.1.5 Syllabus (Table of Contents, etc.)
 - 2.3.1.1.2 Scheduling Information
 - 2.3.1.1.2.1 Status (open, closed)
 - 2.3.1.1.2.2 Category (Complementary, Indispensable, etc.)
 - 2.3.1.1.2.3 Class Scheduling
 - 2.3.1.1.2.4 Instructors
 - 2.3.1.1.3 Assessment Modality Information
 - 2.3.1.1.4 Instructional Material Information
 - 2.3.1.1.4.1 Availability
 - 2.3.1.1.4.2 Media Type
 - 2.3.1.1.4.3 Demos
 - 2.3.1.2 Requirements Information
 - 2.3.1.2.1 Academic Requirements for Admission
 - 2.3.1.2.2 Technical Requirements
 - 2.3.1.3 Application Information
 - 2.3.1.3.1 Fee Payment Information
 - 2.3.2. Student Features
 - 2.3.2.1 On-line Student Account/File
 - 2.3.2.1.1 Registration
 - 2.3.2.1.2 Security
 - 2.3.2.1.3 Configuration
 - 2.3.2.2 Online Paperwork
 - 2.3.2.2.1 Career Paperwork
 - 2.3.2.2.1.1 Enrollment
 - 2.3.2.2.1.2 Certification
 - 2.3.2.2.2 Course Paperwork
 - 2.3.2.2.2.1 Enrollment
 - 2.3.2.2.2.2 Status Tracking
 - 2.3.2.2.2.3 Mailbox
 - 2.3.3. Virtual Learning Environment Features (for a WbIS)
 - 2.3.3.1 Synchronous Communication and Collaboration Resources
 - 2.3.3.1.1 Class Session Palette
 - 2.3.3.1.2 Participation/Collaboration Mechanisms
 - 2.3.3.1.2.1 Hand raising
 - 2.3.3.1.2.2 Class Member List
 - 2.3.3.1.2.3 Synchronized Web Navigation
 - 2.3.3.1.2.4 Assistant Instructor Participation
 - 2.3.3.1.2.5 Workgroup Facility
 - 2.3.3.1.3 Synchronous Resources.
 - 2.3.3.1.3.1 Streaming Slides
 - 2.3.3.1.3.2 Streaming Video
 - 2.3.3.1.3.3 Shared Applications (e.g., Whiteboard)
 - 2.3.3.1.3.4 Audio/Video Conference
 - 2.3.3.1.3.5 Chat
 - 2.3.3.1.4 Virtual Lab
 - 2.3.3.2 Asynchronous Communication and Collaboration Resources
 - 2.3.3.2.1 Participation Mechanisms.
 - 2.3.3.2.1.1 Messaging (e-mail, discussion lists, Message Board)
 - 2.3.3.2.1.2 Agenda Facility
 - 2.3.3.2.2 Asynchronous Resources
 - 2.3.3.2.2.1 Class Playing (previously recorded)
 - 2.3.3.2.2.2 Case Studies
 - 2.3.3.2.2.3 Class Related Links
 - 2.3.3.2.2.4 Simulators
 - 2.3.3.2.2.5 Personalized Resources
 - 2.3.3.2.2.5.1 Documents Repository
 - 2.3.3.2.2.5.2 Recommendations (e.g., links, people, related courses, etc.)
 - 2.3.3.3 Assessment Mechanisms
 - 2.3.3.3.1 Self-assessment
 - 2.3.3.3.2 Proctored/Supervised Assessment
 - 2.3.3.3.3 Delivery/Return Mechanisms
 - 2.3.3.3.4 Scoring Mechanisms
 - 2.3.3.3.5 Grading Mechanisms
 - 2.3.4. Student Services
 - 2.3.4.1 Virtual Community
 - 2.3.4.1.1 Bulletin Boards
 - 2.3.4.1.2 Discussion Forums
 - 2.3.4.1.3 others
 - 2.3.4.2 Help and Support
 - 2.3.4.2.1 General Help of the Site
 - 2.3.4.2.1.1 Interpretation Help
 - 2.3.4.2.1.2 Configuration Help
 - 2.3.4.2.1.3 FAQ
 - 2.3.4.2.2 On-line Support
 - 2.3.4.2.2.1 Chat
 - 2.3.4.2.2.2 VoIP
 - 2.3.4.2.3 Off-line Support
 - 2.3.4.2.3.1 E-mail
 - 2.3.4.2.3.2 Over the Phone
 - 2.3.4.2.3.3 Fax on Demand
 - 2.3.4.3 Other Services.
 - 2.3.4.3.1 Virtual Library
 - 2.3.4.3.2 Web Mail
 - 2.3.4.3.3 Downloads
 - 2.3.5. Privacy, Certification and Guarantee Policies
 - 2.3.5.1 Information on Privacy Policy
 - 2.3.5.2 Information on Certification Policy
 - 2.3.5.2.1 Validity
 - 2.3.5.2.2 Shipment Mode
 - 2.3.5.3 Guarantee Policy Information
 - 2.3.5.3.1 Terms of Use
 - 2.3.5.3.2 Return Policy
 - 2.3.5.3.3 Changes Policy

Figura 4.2. Subcaracterística Funcionalidad y Contenidos Específicos para sitios y aplicaciones con funcionalidad e-Learning

En la figura 4.2 se ejemplifica la subcaracterística Funcionalidad y Contenidos Específicos para sitios y aplicaciones con funcionalidad e-Learning además de subcaracterísticas y atributos para un visitante intencional. Este árbol de requerimientos corresponde a un estudio realizado en 2002 [COV02], donde se empleó WebQEM para evaluación de producto.

Paso 3. *Definición de criterios de los indicadores elementales y procedimientos de medición.* Los evaluadores deben definir una base de criterios para la evaluación elemental, y realizar el ulterior proceso de medición y obtención de los valores de indicadores elementales. Un criterio de evaluación elemental declara y especifica cómo obtener un indicador a partir de una medida de una métrica (para un atributo dado). El resultado final es una preferencia o indicador elemental, y si la escala seleccionada fuera porcentual, puede ser interpretado como el grado o porcentaje del requerimiento de calidad elemental satisfecho.

Por lo tanto, para cada métrica de un atributo es necesario establecer un rango de valores aceptables y definir la función de criterio elemental, que producirá una correspondencia entre el valor de la métrica con el nuevo valor que representa la preferencia elemental. Así, luego de computar la función de preferencia elemental que modela el requerimiento del atributo A_i (a partir de la métrica m : $A_i \rightarrow X_i$), el valor del indicador (I : $X_i \rightarrow I_i$) caerá en uno de los tres niveles de aceptabilidad o barras de calidad. Esto es, si se ha consensuado una escala porcentual para el indicador, los niveles de aceptabilidad podrían ser: insatisfactorio (de 0 a 40%), marginal (desde 40,01 a 60%), y satisfactorio (desde 60,01 a 100%).

Paso 4. *Definición del Modelos de agregación e implementación de la evaluación global.* En el paso previo se producen n indicadores de calidad elemental para los n atributos considerados en el árbol de requerimientos. Por lo tanto, aplicando un mecanismo de agregación paso a paso, los indicadores elementales se pueden agrupar convenientemente para producir al final un esquema de agregación completo. Los indicadores de calidad parciales y global se pueden obtener mediante cálculo conforme al modelo de agregación y puntaje empleado. Posibles estrategias implementadas en WebQEM_Tool, son el modelo de agregación meramente aditivo [GIL69], y el modelo de agregación lógica denominado LSP [DUJ96]. El nivel de preferencia global representa el grado de satisfacción de todos los requerimientos especificados para un sitio o aplicación Web dado.

Para obtener el indicador de calidad global e indicadores parciales (IG/P), se puede emplear la siguiente estructura de agregación, tomando en cuenta el modelo aditivo:

$$IG/P = (P_1 IE_1 + \dots + P_n IE_n);$$

Donde I_{Ei} son los indicadores elementales (o parciales, conforme al nivel del árbol que se calcule) y P_i son los pesos que modelan la importancia relativa de cada indicador elemental (o parcial) dentro de un grupo. Cada P_i debe ser mayor que cero y la sumatoria en un grupo/subgrupo debe dar uno (ver [OLS00b] para más detalles).

A modo de ejemplo simple, si para el trozo siguiente de árbol de requerimientos, tomado de la figura 4.2, e ilustrado en la figura 4.3 se consideran ya acordados los pesos (encerrados entre los símbolos $< >$), y obtenidos los indicadores elementales (encerrados entre paréntesis), el valor calculado para $I_{2.3.3.1}$ (*Synchronous Communication and Collaboration Resources*) aplicando el modelo aditivo es: $(0.20 * 60.00) + (0.25 * 17.50) + (0.25 * 68.00) + (0.20 * 00.00) = 33.38$; teniendo en cuenta para el cálculo los valores de los indicadores de nivel inferior: $I_{2.3.3.1.3}$ e $I_{2.3.3.1.2}$.

2.3 Specific Functionality and Contents for E-learning

...

- 2.3.2.4 Virtual Learning Environment Features (for a WbLS) $<N.NN>$
- 2.3.3.5 Synchronous Communication and Collaboration Resources
 - 2.3.3.1.1 Class Session Palette $<0.20>$ **(60.00)**
 - 2.3.3.1.2 Participation/Collaboration Mechanisms $<0.25>$ **(17.50)**
 - 2.3.3.1.2.1 Hand Raising $<0.25>$ (0)
 - 2.3.3.1.2.2 Class Member List $<0.20>$ (100)
 - 2.3.3.1.2.3 Synchronized Web Navigation $<0.15>$ (100)
 - 2.3.3.1.2.4 Assistant Instructor Participation $<0.20>$ (50)
 - 2.3.3.1.2.5 Workgroup Facility $<0.25>$ (70)
 - 2.3.3.1.3 Synchronous Resources. $<0.25>$ **(68.00)**
 - 2.3.3.1.3.1 Streaming Slides $<0.20>$ (0)
 - 2.3.3.1.3.2 Streaming Video $<0.20>$ (100)
 - 2.3.3.1.3.3 Shared Applications (e.g., Whiteboard) $<0.20>$ (50)
 - 2.3.3.1.3.4 Audio/Video Conference $<0.20>$ (90)
 - 2.3.3.1.3.5 Chat $<0.20>$ (100)
 - 2.3.3.1.4 Virtual Lab $<0.20>$ **(0.00)**

Figura 4.3. Sección del árbol de requerimientos para ejemplificar el cálculo de indicadores parciales.

Los indicadores de más bajo nivel se calcularon de la siguiente manera:

$I_{2.3.3.1.3}$ (*Synchronous Resources*) : $(0.20 * 00.00) + (0.20 * 100.00) + (0.20 * 50.00) + (0.20 * 90.00) + (0.20 * 100.00) = 68$

$I_{2.3.3.1.2}$ (*Participation/Collaboration Mechanisms*) : $(0.25 * 00.00) + (0.20 * 100.00) + (0.15 * 100.00) + (0.20 * 50.00) + (0.25 * 70.00) = 17,50$

Este proceso recursivo debe ser completado para todos los atributos y subcaracterísticas en el árbol de requerimientos para obtener la calificación global.

Paso 5. Análisis de resultados y recomendaciones. Una vez diseñado e implementado el proyecto de evaluación, el proceso culmina con la documentación de las conclusiones y recomendaciones. Los evaluadores analizan los resultados considerando las metas y el perfil de usuario establecidos. El proceso de evaluación produce información elemental, parcial y global que puede ser fácilmente analizada por medio de un modelo de seguimiento, de generación dinámica, y que se emplea en actividades de toma de decisión. Por ejemplo, se puede definir la siguiente hipótesis a corroborar: *“la calidad de la (o las) aplicación Web satisface en general los requerimientos de calidad en consideración de un perfil de usuario. Particularmente, que el indicador global de calidad en uso satisface al menos el punto crítico de aceptabilidad del 70% de la preferencia global, conforme a los requerimientos de calidad especificados y acordados para una audiencia previamente determinada”*.

4.3. Adaptando WebQEM para Calidad en Uso

Si bien WebQEM estaba orientada inicialmente a evaluación de calidad de producto, el hecho de que base los requerimientos no funcionales en modelos de calidad (modelo ISO 9126, modelo mixto, etc.) le otorga gran flexibilidad, ya que al cambiar el modelo subyacente se pueden evaluar otras perspectivas de calidad de un ente, sin perder por ello los beneficios originales. Particularmente, para esta tesis, se experimentó con el modelo de calidad en uso del estándar ISO 9126-1.

Para el modelo de calidad en uso, a diferencia del modelo jerárquico para calidad interna y externa, el estándar ISO no prescribe subcaracterísticas.

También resultó un desafío articular la interacción entre evaluadores expertos (para lo cual WebQEM se diseñó originalmente) y usuarios finales. Por ejemplo en el caso de estudio que se presenta en el capítulo 5, los usuarios finales, de un determinado perfil, llevaron adelante una serie de tareas diseñadas específicamente para permitir la evaluación de la aplicación Web seleccionada.

Luego, a partir de los valores de métricas e indicadores obtenidos, los expertos hicieron el análisis correspondiente para arribar a las conclusiones y realizar las recomendaciones pertinentes.

A las actividades propuestas en el paso 1 *“Definición de las metas de evaluación y selección del perfil de usuario”* se agregó en este caso, como labor para los expertos, el *diseño de las tareas* que debían llevar a cabo los

usuarios para poder evaluar calidad en uso y *la especificación del contexto* en el que se llevarían a cabo las tareas. Para ello, los expertos inspeccionaron las características de la aplicación Web a evaluar, y consideraron, dentro del contexto del dominio estudiado, un conjunto de tareas típicas para el perfil de usuario seleccionado, indicando también las características del ambiente, las instalaciones y el equipamiento necesario para llevar adelante el estudio .

Otra cuestión sobre la que se hizo hincapié fue en integrar aspectos de evaluación subjetiva. Este hecho resultaba de vital importancia ya que WebQEM está orientada más a una evaluación cuantitativa que cualitativa y objetiva antes que subjetiva, pero era imprescindible evaluar satisfacción de los usuarios –característica eminentemente subjetiva- con respecto a la aplicación Web evaluada.

De entre las posibilidades existentes se optó por adaptar para satisfacción una encuesta basada en una escala validada estadísticamente, de uso difundido, normalizada en sintonía con la de los indicadores diseñados para las métricas de atributos de las otras dos características (*Eficacia y Productividad*). Esta decisión permitió incluir la media de las encuestas como el valor del indicador de satisfacción, directamente en el paso de cálculo del indicador de preferencia global (paso 4). Esto se discutirá en detalle en secciones 5.6.2.1.3 y 5.6.2.2.5.

4.4. Consideraciones Finales

La *Metodología de Evaluación de Calidad de Sitios y Aplicaciones Web* (WebQEM) realiza un aporte ingenieril al sentar las bases para el uso de un enfoque sistemático, disciplinado y cuantitativo que se adecue a la evaluación, comparación y análisis de la calidad de sitios y aplicaciones Web más o menos complejas. En la sección 4.2 se discutieron las fases y actividades principales y los métodos y modelos para realizarlas.

Durante la fase de Definición y Especificación de los Requerimientos de Calidad se realizan actividades y procedimientos para la elicitación, modelado y especificación de los requerimientos. Este enfoque se emplea para representar y acordar el árbol de requerimientos de calidad de artefactos Web, para una meta, perfil de usuario y dominio dados. Si se considera la naturaleza de los sitios y aplicaciones Web, se encuentran varios atributos y requerimientos que los hacen distintivos respecto de los artefactos de software tradicional. No obstante, las características de calidad de alto nivel de abstracción prescritas en el estándar ISO son, en muchos casos, suficientes para modelar y contener las particularidades de las aplicaciones Web, aunque según se ha discutido en la sección 2.4, para tener un modelo realmente completo será necesario tener en cuenta una característica adicional: la de Contenido. Finalmente, el producto de esta fase es empleado como

entrada en la preparación de la evaluación elemental y global (y retroalimenta a las actividades de planificación y programación).

Durante la fase de Definición (Diseño) e Implementación de la Evaluación Elemental, se definen métricas para cada atributo cuantificable, y criterios de evaluación de calidad elemental centrados en indicadores, funciones y rangos de aceptabilidad, entre otros asuntos. Una vez definidos y consensuados los criterios (métricas) para medir cada atributo, se ejecuta el proceso de recolección de datos, el de cómputo de los valores de indicadores elementales, y la documentación de los resultados. Los valores de indicadores elementales obtenidos sirven de entrada al proceso de evaluación global (por medio de modelos de agregación).

Durante la fase de Definición (Diseño) e Implementación de la Evaluación Global se debe seleccionar un modelo que permita derivar y calcular los valores de indicadores para las subcaracterísticas y características de más alto nivel de abstracción, a partir de los valores de indicadores elementales obtenidos para cada atributo del árbol. Se indicó la existencia de dos enfoques cuantitativos de utilidad en las actividades de estructuración de atributos y características y de cálculo de puntaje para los requerimientos de calidad, a saber: el modelo meramente aditivo, y el modelo de agregación lógica. Este último modelo fue el seleccionado en los casos de estudio realizados, debido a la cantidad de atributos, subcaracterísticas y características intervinientes, y a la necesidad de modelar diferentes tipos de relaciones entre los mismos. Una vez que fueron estructurados y acordados todos los criterios, pesos y operadores de agregación, los evaluadores pueden utilizar la herramienta para calcular los resultados parciales y finales de indicadores.

En la fase de Análisis de Resultados, Documentación y Recomendaciones los evaluadores realizan actividades de análisis y recomendaciones, permitiendo la justificación de los resultados. En consideración de las metas establecidas y el perfil de usuario seleccionado para una necesidad de información dada, el proceso culmina con un documento para entregar, que contiene las conclusiones y recomendaciones del caso. El proceso de evaluación produce información elemental, parcial y global que puede ser fácilmente analizada por medio de un modelo de seguimiento y empleada en actividades de toma de decisión.

WebQEM es una metodología cuantitativa de inspección de características, subcaracterísticas y atributos. Está centrada en evaluadores expertos. Puede ser complementada con otros métodos cualitativos como la evaluación heurística, o con estrategias que permitan detectar la calidad en uso o usabilidad percibida a partir de instrumentos como cuestionarios de satisfacción (técnicas y métodos de consulta, centrados en usuarios finales). En conclusión WebQEM permite comprender el estado actual de la calidad de

sitios y aplicaciones Web operativos al facilitar, por ejemplo, detectar requerimientos pobremente implementados; subcaracterísticas y atributos ausentes; problemas de interfase, de navegación, de contenido, de performance, entre otros. Por otra parte, permite especificar y controlar requerimientos de calidad para nuevos proyectos. En definitiva permite realizar recomendaciones para mejorar la calidad de los productos Web.

En el capítulo siguiente se presentará un caso de estudio para calidad en uso de aplicaciones Web, en el dominio de e-Learning, donde se especifica el diseño de un test para ser llevado a cabo por usuarios finales, en un contexto específico de uso, en el que se empleará WebQEM como la metodología para llevar adelante el proceso de medición y evaluación -con las adaptaciones discutidas en la sección 4.3-.

Parte III: Un Caso de Estudio de Calidad en Uso Para Aplicaciones Web

Capítulo 5: Caso de Estudio de Calidad en Uso de una Aplicación para el Dominio de e-Learning

5.1. Introducción

El caso de estudio que se presenta en esta capítulo se desarrolló con dos objetivos principales: el primero, de índole general, era validar en la práctica la propuesta de evaluación sistemática de calidad en uso para aplicaciones Web; y el segundo, más específico, estaba dado por la necesidad de conocer la calidad en uso de una aplicación Web particular, considerando un determinado perfil de usuario.

Teniendo en cuenta que la metodología adoptada para llevar adelante el caso de estudio fue WebQEM, con las adaptaciones necesarias para evaluar calidad en uso (sección 4.3 del capítulo 4), la presentación del caso de estudio se realizará siguiendo las fases de dicha metodología. Pero con el objetivo de facilitar la comprensión se desdoblará la presentación considerándose en este capítulo sólo lo referente a definición y diseño de la evaluación, y en el capítulo siguiente, lo concerniente a ejecución del estudio y análisis preliminar de los resultados.

Es necesario resaltar que hay una correspondencia entre las fases de WebQEM y los componentes del marco de medición y evaluación mencionado en el capítulo 3. Esta relación proporciona, a su vez, una guía para el empleo sistemático de los conceptos provistos por la ontología de métricas e indicadores en los distintos pasos de cada fase de WebQEM. Por ejemplo, en la Definición de Requerimientos intervienen, entre otros, los siguientes conceptos: *Necesidad de Información*, *Concepto Calculable*, *Modelo de Concepto*, *Entidad* y *Atributo*.

Además, para el diseño de las métricas apropiadas para el dominio de la aplicación Web evaluada, se implementó una instancia del modelo de métricas e indicadores. Esta instanciación permitió, por un lado, adoptar y adaptar algunas métricas sugeridas en el estándar [ISO 9126-4] y, por otro lado, desarrollar y dotar de significado a un nuevo conjunto de métricas e indicadores específicos para calidad en uso.

El modelo de calidad subyacente seleccionado, funcional tanto a WebQEM como a la ontología de métricas e indicadores, fue el modelo de calidad en uso propuesto en el estándar [ISO9126-1], del que se han considerado sólo tres de los cuatro subconceptos (características) principales, esto es: Eficiencia, Productividad y Satisfacción.

Un detalle importante, ya señalado en el capítulo 2, es que la evaluación de calidad en uso se realiza a partir de un contexto de trabajo y un perfil de usuario específicos y conocidos. Esto implica que si bien las métricas (en particular los metadatos asociados) se pueden reutilizar, no ocurre lo mismo con los indicadores, que deben responder a objetivos y metas es-

pecíficos. O sea que conclusiones obtenidas a partir de los indicadores no pueden generalizarse, ni parcial ni globalmente a otros dominios, aplicaciones Web o perfiles de usuario diferentes.

El caso de estudio presentado en este capítulo se llevó a cabo en abril de 2004, para una aplicación Web denominada “QPlus Campus Virtual”, empleada desde fines de 2002 por la Facultad de Ingeniería de la Universidad Nacional de La Pampa (UNLPam). El objetivo de la iniciativa es ofrecer a los alumnos ingresantes un curso de Introducción a la Matemática. Las tareas propuestas a los usuarios fueron diseñadas en relación a un conjunto de tareas típicas para el perfil de usuario al que pertenecen los alumnos, teniéndose también en cuenta la funcionalidad y servicios de la aplicación Web a evaluar.

El capítulo 5 está organizado de la siguiente manera: en la sección 5.2 se ofrece un breve panorama del dominio las aplicaciones Web para e-Learning, dado que la aplicación Web evaluada pertenece a ese dominio y es de particular interés en el ámbito académico; en la sección 5.3 se comienza con la primera fase de la metodología de evaluación: planear y programar la evaluación de calidad, a continuación el diseño de tareas (sección 5.3) y la especificación del contexto (sección 5.4). En la sección 5.5 se presenta la definición y especificación de los requerimientos de calidad; en la sección 5.6 se encuentra el diseño de la evaluación elemental con un detalle de las métricas e indicadores desarrollados y/o adaptados específicamente para el caso de estudio. En la secciones 5.7 y 5.8 se presentan aspectos de la evaluación global. Por último, en la sección 5.9, se encuentran las conclusiones del capítulo.

5.2. El Dominio de Aplicaciones Web para e-Learning

Si bien no forma parte de los objetivo de esta tesis realizar aporte alguno sobre aspectos específicos del dominio de e-Learning, resulta oportuna una breve reseña sobre la perspectiva que ofrecen ciertas aplicaciones Web en el contexto del proceso de enseñanza/aprendizaje y del interés que despiertan, como para justificar un caso de estudio de calidad en uso.

Las aplicaciones Web en el dominio de e-Learning han tenido un crecimiento significativo en los últimos años, en el contexto de lo que es la evolución de Internet y de la WWW, han recibido especial atención de parte de compañías y organizaciones, tanto del sector industrial/comercial como del sector educativo, y en este último caso especialmente a nivel universitario.

Según [GOO00], e-Learning se define como *“el uso sistemático de tecnologías multimediales basadas en computadoras para potenciar a los estudiantes, mejorar el aprendizaje, conectar a los estudiantes con personas y recursos que ayuden a sus necesidades, y a conjugar aprendizaje con rendi-*

miento y metas individuales con metas organizacionales"¹⁸. El núcleo de una aplicación Web que da soporte a un proceso de e-Learning, brindando un conjunto de servicios integrados para la creación y distribución de información en línea, la comunicación y colaboración entre los usuarios y la administración de recursos, se denomina genéricamente LMS (Learning Management System).

En la perspectiva de los miembros de una comunidad educativa, es decir en un contexto real de enseñanza/aprendizaje, si el LMS se emplea desde una interfaz homogénea que integra el acceso a contenidos y servicios, se trata de un Sistema de Enseñanza y Aprendizaje basado en la Web, conocido por el acrónimo en inglés WbIS (Web-based Instructional System) [RET02].

De acuerdo a las definiciones anteriores, a la perspectiva de los usuarios que participaron del caso de estudio, y al contexto en que se utilizó, la aplicación Web evaluada en este caso de estudio reúne las características de un WbIS.

Numerosos trabajos, de entre los que resulta de interés mencionar el de Avgeriou et al [AVG01], resaltan la madurez alcanzada por distintos LMSs comerciales. Pero como esas soluciones son en muchos casos realmente sofisticadas y costosas, se produjo también un crecimiento proporcional de productos semejantes, pero desarrollados bajo licencia de uso libre y/o gratuito. Esta circunstancia ha permitido iniciativas importantes a organizaciones educativas con capacidad económico/financiera limitada, tal como señala [GRO04], y la realización de iniciativas experimentales de educación con soporte vía Internet, sin afrontar grandes inversiones hasta no contar con los resultados de evaluaciones preliminares. El caso de estudio presentado en este capítulo se realizó también en el marco de una de esas experiencias exploratorias [COV02a].

Además, en la opinión de algunos expertos, como Milutinovic et al [MIL03], las universidades que no posean un WbIS de calidad no estarán en condiciones de competir con una oferta educativa aceptable, en un desafío que visualizan proyectado a escala global a corto y mediano plazo.

Estos antecedentes y referencias constituyen, en el ámbito académico universitario, una motivación suficiente para avanzar en estudios y prácticas que, en forma metódica y sistemática, permitan evaluar la calidad de producto y la calidad en uso de aplicaciones Web que dan soporte al proceso de enseñanza/aprendizaje.

A continuación se encuentra la primer fase del proceso de medición y evaluación, con las actividades relativas a la definición del proyecto.

¹⁸e-Learning is the systematic use of networked multimedia computer technologies to empower learners, improve learning, connect learners to people and resources supportive of their needs, and to integrate learning with performance and individual with organizational goals.

5.3 Planificar y Programar la Evaluación de Calidad en Uso¹⁹

En el contexto de esta fase se consideraron dos pasos o actividades primordiales y comunes a toda evaluación de calidad siguiendo la metodología WebQEM: definición de las metas de evaluación y selección del perfil de usuario; más otros dos pasos propios de una evaluación de calidad en uso: el diseño de las tareas a llevar a cabo por los usuarios y la especificación del contexto en que se realizará el caso de estudio.

5.3.1 Objetivos del Caso de Estudio

Los objetivos del caso de estudio para evaluación de calidad en uso fueron: 1) Validar un modelo de evaluación de calidad en uso para aplicaciones Web y, en particular, 2) conocer la percepción de la calidad en uso que tienen, como usuarios finales, los alumnos ingresantes a la Facultad de Ingeniería de la UNLPam, respecto de la aplicación Web “Qplus Campus Virtual” [QPL04].

La hipótesis para el caso de estudio en cuanto al segundo objetivo es que, al menos a nivel de subconceptos (características), el sitio está por encima de un umbral de satisfacción aceptable.

La aplicación Web generada a partir del WbIS “Qplus Campus Virtual”, fue empleada sistemáticamente por la Facultad de Ingeniería en la provisión de contenidos vía Web para los alumnos ingresantes en los años 2003, 2004 y 2005.

5.3.2 Perfil de la Audiencia

Para el diseño e implementación de un caso de estudio de evaluación de calidad en uso, una meta típica de evaluación es determinar el cumplimiento de requerimientos elementales, parciales y global de calidad para una aplicación Web operativa, considerando el perfil de visitante intencional.

Un visitante intencional se define como la audiencia que tiene al menos algún conocimiento o manifiesta algún interés en un dominio específico, y probablemente desea informarse o aprender más acerca de sus contenidos y servicios para emplearlos frecuentemente. Su permanencia en el sitio es generalmente mayor que la de una audiencia casual.

Para este caso de estudio, con el objetivo de completar un conjunto definido de tareas, fue convocado un grupo representativo de usuarios, asistentes a un curso introductorio de primer año, cuyo contenido es afín a las carreras que se dictan en la Facultad de Ingeniería de la UNLPam. La población de usuarios sobre la que se realizó este estudio estaba constituida por

¹⁹ Corresponde a la *Definición del Proyecto* del Marco de Medición y Evaluación INCAMI [OLS05a]

estudiantes, específicamente alumnos ingresantes a las carreras de Ingeniería Electromecánica e Ingeniería en Sistemas.

Este grupo de alumnos tuvo a disposición vía Web, a través del WbIS “Qplus Campus Virtual”, un curso de introducción a la Matemática, durante los meses de diciembre de 2003 a febrero de 2004. Durante ese período de tiempo los responsables del curso y miembros del personal técnico estuvieron a disposición de los alumnos, vía correo electrónico, para responder sobre cuestiones relacionadas al curso o bien de operación y funcionamiento del sitio Web, respectivamente.

El número potencial de usuarios ascendía a ochenta aproximadamente, y se tenía interés en conocer la percepción de la calidad en uso que tenían sobre “Qplus Campus Virtual”, habiendo participado del curso inicial de “Introducción a la Matemática”.

Para llevar adelante el caso de estudio fueron seleccionados, empleando un mecanismo al azar, 6 (seis) alumnos voluntarios del curso de Introducción a la Matemática 2004, que reunían las características de *usuarios intencionales* y usaron la aplicación Web durante el período diciembre-2003/febrero-2004.

Para definir la cantidad de usuarios participantes se tuvo en cuenta la opinión de Jakob Nielsen, que en su columna on-line Alertbox [NIE04] afirma que pueden obtenerse resultados significativos en una evaluación de usabilidad con la participación de un número pequeño de usuarios –normalmente de 3 a 5- si es factible repetir el test varias veces. Un beneficio adicional de esta propuesta es la reducción de los costos del caso de estudio.

Otros usuarios que pueden considerarse para estudios semejantes, dentro del mismo perfil, tal como se propone en el trabajo ya citado [AVG03] son: *profesores y tutores* y considerando un perfil de usuario experto: los *administradores*.

Antes de llevar a cabo el caso de estudio se hicieron tests exploratorios. En un caso con un usuario experto en el dominio y en otro con un usuario avanzado en el uso de herramientas y servicios Web, pero no experto en el dominio. La meta era tener información para contrastar, de modo provisorio, los resultados obtenidos por distintos perfiles de usuario.

Otras características que podrían influir en la evaluación general, tales como edad, género y disposición a usar herramientas informáticas para estudiar, también fueron relevadas empleando un cuestionario específico que puede consultarse en el apéndice D.

Una consideración adicional, que incidió luego en la definición de las tareas, es que si bien se sabía que los usuarios habían ganado cierta experiencia en el uso de “Qplus Campus Virtual”, en realidad no habían tenido un entrenamiento sistemático en el uso de todos y cada uno de los servicios de la aplicación Web. El seguimiento de la actividad de los usuarios se hizo a

través del Sistema de Gestión de la aplicación, que permite la traza de las actividades realizadas, tomando en cuenta los archivos de registro de uso o log-files.

Además, el hecho de que muchos alumnos ingresantes provenían de ciudades distintas a donde está la sede de la Facultad de Ingeniería, y de que la aplicación Web había sido presentada brevemente en el momento de la inscripción a la carrera, justificaban la importancia de evaluar la calidad en uso del producto a utilizar. Luego de ese primer acercamiento al producto, los alumnos debieron trabajar por su cuenta en sus domicilios o lugares públicos de acceso a Internet.

5.3.3 Tareas para Evaluar Calidad en Uso

En primer término, para el diseño de las tareas que se propusieron a los usuarios en el caso de estudio se consideraron a nivel general, y teniendo en cuenta las propuestas del trabajo citado [AVG03], un conjunto de tareas básicas que es factible desarrollar utilizando un WbIS, esto es: distribuir información, proveer material de aprendizaje, ofrecer diversos mecanismos de comunicación sincrónica y asincrónica y gestionar un aula virtual.

En segundo término, se analizaron las tareas específicas para un alumno, cuyo perfil es el del usuario intencional seleccionado para el caso de estudio. Esas tareas específicas son:

- a. Buscar y recuperar información
- b. Compilar y descargar material para actividades fuera de línea
- c. Completar evaluaciones
- d. Personalizar el LMS (Organizar, de acuerdo a sus preferencias, los cursos en los que está inscripto y el material de estudio)
- e. Realizar consultas al instructor (completar los campos obligatorios y enviar el mensaje)
- f. Inscribirse en un curso
- g. Responder cuestionarios
- h. Completar ejercicios on-line
- i. Acceder a los eventos de una fecha del calendario
- j. Invitar a un colega a un evento on-line

Finalmente, se tuvieron en cuenta las particularidades del WbIS a evaluar, considerando que no todos ofrecen la misma funcionalidad y servicios. Y que, por otra parte, por razones de eficacia en el testing de usuarios la cantidad de tareas a llevar a cabo por cada usuario debe ser limitada. De este análisis surgieron los objetivos específicos para cuatro tareas, que se consideraron representativas para el perfil mencionado:

- 1) Buscar y recuperar contenidos del material de estudio
- 2) Realizar una consulta a un profesor
- 3) Participar en un foro de opinión
- 4) Organizar el material de estudio disponible

Ver el apéndice C para el detalle del diseño de las tareas.

5.4 Descripción del Contexto

El contexto en el que se desarrolla un caso de estudio es fundamental en la evaluación de calidad en uso. La importancia del contexto radica en que no es factible obtener conclusiones válidas al evaluar un mismo producto en contextos distintos o bien en un contexto que no está correctamente especificado. Esta cuestión fue planteada inicialmente en la sección 2.3.5, en las discusiones conceptuales sobre calidad en uso.

Los aspectos relevantes que se presentan en relación al contexto son: la descripción del producto, los requerimientos de hardware y software, las instalaciones y el perfil de usuario (audiencia) detallado en la sección anterior.

5.4.1 Información del Producto

El producto evaluado en el caso de estudio, denominado “Qplus Campus Virtual”, es una aplicación Web que ofrece servicios típicos de un WbIS, permitiendo a los miembros de una comunidad educativa (alumnos, profesores, tutores, etc.) llevar adelante el proceso de enseñanza/aprendizaje, empleando recursos y servicios provistos típicamente por la Web. Se puede utilizar en una modalidad a distancia o complementando un sistema presencial o semipresencial.

“Qplus Campus Virtual” es utilizada actualmente, entre otras, por los siguientes organizaciones: Universidad Nacional de Quilmes, Asociación de Trabajadores del Estado (Santa Fé), Instituto Da Vinci (Pringles, Provincia de Buenos Aires), y también la Facultad de Ingeniería de la UNLPam.

Información adicional del producto puede obtenerse en <http://www.qplus.com.ar/productos.htm>.

5.4.2 Requerimientos del Producto

Los requerimientos para utilizar “Qplus Campus Virtual” son bastante simples. Requiere, del lado del cliente, un browser de ante-última o última generación y la máquina virtual Java. El acceso a los contenidos debe ser franqueado empleando nombre de usuario y contraseña provistos por el administrador. La página principal del sitio ofrece la posibilidad de completar un formulario de inscripción en línea.

Para que un usuario pueda utilizar el producto es suficiente una computadora personal basada en procesador Pentium, con 64 Mbytes de memoria RAM, monitor color de 15”, teclado, dispositivo apuntador y un sistema operativo con interfaz gráfica que soporte un browser entre los más conocidos (Opera, Internet Explorer, Netscape, Mozilla).

“Qplus Campus Virtual” puede ser utilizado por un usuario con capacidad de manejo de una computadora personal y conociendo como se emplea

un browser para buscar y recuperar información de la Web, y acceder así a servicios tales como motores de búsqueda, Web mail, foros de opinión, Chat, etc.

5.4.3 Instalaciones

El caso de estudio se llevó a cabo en el laboratorio del grupo GIDIS_Web [<http://gidis.ing.unlpam.edu.ar>], en la Facultad de Ingeniería, que fue levemente adaptado para simular un ambiente de estudio para un estudiante, contándose con una computadora personal destinada a la realización de las tareas, sobre un escritorio con suficiente espacio como para apoyar las hojas con las instrucciones del estudio y la encuesta provista por los evaluadores.

Los participantes trabajaron individualmente, sin ser interrumpidos, y recibieron respuestas a consultas sobre cuestiones de orden general cuando lo solicitaron. También fueron videofilmados y seguidos a través de una sesión de pantalla remota. Cada sesión fue registrada y grabada en un archivo en el disco de la computadora, como complemento de la filmación.

Para el caso de estudio se empleó una computadora personal basada en un microprocesador Pentium IV con 256MB de RAM, teclado de 104 teclas, un Mouse de dos botones y rueda central, con un monitor color de 17", configurado a 800x600 píxeles en color verdadero. El sistema operativo empleado fue MS Windows 2000 edición Profesional, con Java VM 1.1.3. Los browsers empleados, según las preferencias de los usuarios, fueron Microsoft IE 6.0 y Netscape Communicator 7.0.

5.4.4 Herramientas Utilizadas por los Evaluadores

Para registrar la actividad de los usuarios se empleó una variedad de herramientas y equipamiento. Por ejemplo el tiempo de las tareas dentro de las sesiones fue calculado usando Hanks Usability Logger; las secuencias de pantalla de sesiones grabadas en disco usando Lotus Screen Cam versión NT y videofilmadas usando una cámara de video digital JVC, modelo GR-DVL300. Luego las sesiones de video fueron procesadas con Ulead Video Studio 5.0.

Los cálculos de la duración de las tareas fueron posteriormente verificados, en el proceso de comprobación del modo en que fueron abordadas y completadas total o parcialmente, mediante un seguimiento realizado a partir de la reproducción de las secuencias de pantalla y los videos correspondientes.

Para la evaluación de satisfacción (ver sección 5.6.2.1.3) se empleó una escala de diez ítems, adaptada de [SUS96] -ver Apéndice D para más detalles sobre el cuestionario-, que permite a los evaluadores tener una idea global de satisfacción percibida por los usuarios. Si bien esta escala no fue diseñada específicamente para aplicaciones Web, con leves adaptaciones se

puede emplear para cumplir inicialmente con esa meta ya que tiene a su favor: a) ser sencilla de completar y calcular, b) estar estadísticamente validada y c) haber sido empleada con anterioridad para evaluar aplicaciones software.

La sencillez para ser completado no es una cualidad menor de un cuestionario, ya que se debe confeccionar después de haber trabajado durante un cierto período de tiempo con el producto que se desea evaluar. Un cuestionario complejo puede predisponer a una actitud negativa a los usuarios, y no reflejará cabalmente el grado de satisfacción respecto al producto que se utilizó.

5.5 Definir y Especificar los Requerimientos de Calidad²⁰

En esta fase, teniendo en cuenta los aspectos definidos en la fase anterior respecto a metas de evaluación, tareas y contexto, se deben establecer atributos y subconceptos (características) de calidad cuantificables que, agrupados jerárquicamente, representen un modelo de calidad apropiado para el perfil de usuario seleccionado.

En cierta medida la discusión y ejemplos acerca de los subconceptos que intervienen en esta evaluación de calidad en uso, ya se presentaron en la sección 3.3.2. La necesidad de información definida en esa sección era: *“conocer la percepción de calidad en uso que tienen los alumnos ingresantes a la Facultad de Ingeniería de la UNLPam (año 2004) al usar un WbIS que da soporte al curso nivelatorio denominado ‘Introducción a la Matemática’”*. Para el concepto calculable “calidad en uso”, cuyo modelo de concepto es el modelo de calidad en uso del estándar [ISO9126-1], los subconceptos utilizados fueron: “Eficacia”, “Productividad”, y “Satisfacción”.

La exclusión del subconcepto “Seguridad”, propuesta en el modelo de concepto de referencia, obedece a que se consideró poco relevante para el producto y el contexto en que se realizó el estudio. Se dio por sentado que el eventual mal funcionamiento de la aplicación Web evaluada no ocasionaría daños, ni perjuicios de ninguna clase a las personas involucradas, ni el deterioro del equipamiento o de las instalaciones utilizadas.

Los atributos seleccionados, funcionales a la necesidad de información definida inicialmente, fueron: *“Complejidad de Tarea”* y *“Eficacia de Tarea”*, relacionados ambos al subconcepto *“Eficacia”*; en tanto que los atributos *“Eficiencia en a relación Complejidad de Tareas”* y *“Eficiencia en relación a Eficacia”* pertenecen al subconcepto *“Productividad”*.

No se han considerado atributos para el subconcepto Satisfacción, de evaluación eminentemente subjetiva, ya que en este caso de estudio se de-

²⁰ Corresponde a la *Definición y Especificación de Requerimientos No Funcionales* del Marco de Medición y Evaluación

ció emplear directamente un cuestionario de satisfacción de índole general [SUS96]. No obstante, de ser necesario, es posible encontrar atributos para evaluar aspectos específicos de satisfacción a un nivel de granularidad más fino, tal como se verá en la sección 5.6.1.1, en la discusión sobre las métricas propuestas en el borrador del estándar [ISO9126-4].

De acuerdo a lo expuesto, los requerimientos que modelan la calidad en uso según las necesidades del perfil de usuario seleccionado, se especifican en el árbol de requerimientos según se observa en la Figura 5.1.

1. Calidad en Uso

1.1. Eficacia

1.1.1. *Compleitud de Tareas*

1.1.2. *Eficacia de Tarea*

1.2. Productividad

1.2.1. *Eficiencia en relación a Compleitud de Tareas*

1.2.2. *Eficiencia en relación a Eficacia*

1.3. Satisfacción

Figura 5.1. *Árbol de Requerimientos para el caso de estudio de calidad en uso.*

Para poder llevar a cabo el proceso de medición será necesario, en la siguiente fase, determinar las métricas y criterios elementales de evaluación para los atributos seleccionados.

5.6 Diseñar la Evaluación Elemental²¹

En esta fase es necesario diseñar y seleccionar las métricas que permitan cuantificar cada uno de los atributos del árbol de requerimientos. De otro modo, sin una métrica apropiada y adecuadamente definida para cada atributo mensurable, no sería posible llevar adelante el proceso de medición. Las métricas e indicadores de calidad en uso empleadas en este caso de estudio fueron desarrolladas teniendo como guía el marco de medición y evaluación presentado en el Capítulo 3.

5.6.1 Métricas de Calidad en Uso

En el borrador del estándar [ISO9126-4] se proponen un conjunto de métricas relacionadas a cada una de las cuatro características del modelo de calidad en uso, con el objetivo de evaluar calidad en uso de un producto software en general. La propuesta de las métricas, que no es prescriptiva, fue analizada detalladamente con el objetivo de adoptar y/o adaptar alguna de las métricas para ser empleadas en el caso de estudio. Por esta razón se incluye, en la sección 5.6.1.1, una breve reseña del objetivo que busca cada

²¹ Corresponde al *Diseño de la Medición* del Marco de Medición y Evaluación

grupo de métricas, en relación a la característica a la que pertenece el atributo que se desea medir.

En la sección 5.6.2 se presentan las métricas e indicadores que se desarrollaron y emplearon para evaluar la aplicación Web seleccionada. Estas métricas e indicadores constituyen, junto a los ejemplos expuestos en el capítulo 3, una instancia del marco de medición y evaluación.

5.6.1.1 Métricas para Calidad en Uso en el Estándar 9126-4

Las métricas propuestas en el borrador del estándar ISO 9126-4 tienen el objetivo de medir los atributos de calidad en. Según el propio reporte, las métricas no constituyen un conjunto exhaustivo y los destinatarios, sean estos los responsables de la gestión de calidad, los compradores, los desarrolladores o los evaluadores; pueden modificar estas métricas y añadir otras no detalladas en el mismo.

Teóricamente estas métricas son aplicables a cualquier producto de software, aunque no todas son aplicables a cada clase de producto. Asimismo no se asignan rangos de valores de las métricas a niveles establecidos o grados de cumplimiento, dado que se considera que esos valores deben definirse para cada producto de software o parte del producto de software de acuerdo a su naturaleza y contexto organizacional.

Métricas de Eficacia

Las métricas de eficacia permiten evaluar si las tareas realizadas por los usuarios alcanzan las metas especificadas con exactitud y completitud en un contexto de uso especificado. No tienen en cuenta cómo fueron alcanzadas esas metas sino solamente en qué medida fueron alcanzadas.

Ejemplos:

Efectividad de Tareas (Task effectiveness): Esta métrica permite evaluar qué proporción de los objetivos de una tarea se logran correctamente.

Completitud de Tareas (Task completion): Esta métrica permite evaluar la proporción de tareas que se completan con respecto a las que se intentan.

Métricas de Productividad

Las métricas de productividad permiten valorar los recursos que consumen los usuarios en relación a la efectividad alcanzada en un contexto especificado de uso. El recurso más común a consumir es tiempo para completar la tarea. Otros recursos podrían ser esfuerzo, materiales o un eventual costo financiero de uso de recursos.

Al igual que las métricas de eficacia, estas métricas sólo pueden obtenerse una vez definida las tareas que deben realizar los usuarios.

Ejemplos:

Tiempo de Tarea (Task time): mide cuánto tiempo lleva completar una tarea. Se trata de una métrica directa que se usa para calcular otras métricas derivadas o indirectas.

Eficiencia relativa de usuario (Relative user efficiency): permite comparar la eficiencia de un usuario medio respecto a un usuario experto. Se debe medir la eficiencia de tarea para un usuario del perfil que se está evaluando y para un usuario experto, luego se puede establecer una proporción o comparación.

Métricas de Seguridad

Las métricas de seguridad están orientadas a evaluar el nivel de riesgo de dañar personas, instalaciones, software, propiedades o el medioambiente en un contexto especificado de uso. Esto incluye la salud y la seguridad de los usuarios y de terceros afectados por el uso, como también las consecuencias físicas o económicas indeseadas.

Ejemplos:

Seguridad y Salud del Usuario (User health and safety): permite detectar cuál es la incidencia de los problemas de salud entre los usuarios del producto.

Seguridad de las personas afectadas por el uso del sistema (Safety of people affected by use of the system): permite evaluar cuál es la posibilidad de riesgo para las personas afectadas por el sistema.

Daño Económico (Economic damage): permite medir cuál es la incidencia del daño económico debido al uso del producto.

Métricas de Satisfacción

Para medir la satisfacción del usuario normalmente se utilizan cuestionarios, compuestos de una batería de preguntas cuyo objetivo es considerar aspectos como la complejidad de las interfaces, la calidad de la documentación, la facilidad y contenidos de la ayuda o la adecuación de la funcionalidad entre otros.

Ejemplo

Nivel de Satisfacción (Satisfaction scale): permite evaluar la satisfacción de los usuarios respecto al producto que usan. En general se usa un instrumento estadísticamente validado, con un mecanismo de cálculo que permite obtener un valor general de acuerdo a las respuestas de cada usuario. La sumatoria de la calificación de cada cuestionario válido dividido la cantidad de respuestas válidas, indicará el grado de satisfacción para el grupo de usuarios participantes.

Se puede utilizar de forma directa respecto a los usuarios que utilizan el producto, o comparar el grado de satisfacción de un perfil de usuario específico respecto al grado de satisfacción de una población media.

5.6.2. Métricas e Indicadores para el Caso de Estudio

Para poder evaluar concretamente calidad en uso es necesario primero medir, empleando las métricas apropiadas, el rendimiento obtenido por usuarios reales, realizando tareas representativas, en un contexto determinado. Por lo tanto, el diseño de las métricas debe tener en cuenta qué aspectos son importantes a evaluar para cada tarea, en relación al atributo que se desea medir. Ejemplos de esos aspectos son: el tiempo (u otro recurso) que ha insumido la realización de la tarea, considerar si la tarea ha sido completada o no, y si no lo fue, qué metas u objetivos parciales se completaron correctamente.

En la sección 4 del anexo A del estándar [ISO9126-1], se señala que el proceso de diseño de la métricas debe ser riguroso de modo que permita la realización de comparaciones confiables. Indicándose además, que las métricas deben poder ser medidas con la suficiente precisión como para permitir la definición de los criterios de comparación y la realización de las comparaciones en sí.

Las comparaciones pueden ser en base a métricas de atributos de entes del mismo tipo, pertenecientes a distintos productos, o para un único ente, con respecto a criterios específicos previamente establecidos. En este caso de estudio, las comparaciones se realizaron con respecto a criterios de decisión. Estos criterios fueron dados por los niveles de aceptabilidad asociados a cada uno de los indicadores.

Las condiciones fundamentales para que una métrica resulte confiable en relación al objetivo que persigue son tres: debe ser *objetiva*, debe ser *empírica* (empleando una escala adecuada) y debe ser *reproducible*.

Para que una métrica sea objetiva debe existir un procedimiento (método de medición) previamente especificado acerca de cómo obtener un valor numérico o categórico. Para ser empírica debe obtenerse luego de un proceso observacional o a través de un cuestionario debidamente validado, empleando una escala de tipo intervalo, proporción o absoluta. Finalmente, para ser reproducible, los resultados obtenidos por diferentes personas, en diferentes ocasiones, empleando los mismos procedimientos, deben arrojar mediciones similares, considerando una tolerancia razonable respecto posibles errores humanos o de calibración de las herramientas e instrumentos de medición.

Para las métricas del caso de estudio se han tenido en cuenta éstas condiciones. Cada métrica tiene un método de medición o de cálculo debidamente especificado, fué obtenida luego de un proceso de observación a través de distintos medios visuales (videofilmación, secuencias de pantalla) o, para la característica Satisfacción, empleando un cuestionario validado estadísticamente. El tipo de escala siempre es proporción o intervalo y las mediciones fueron verificadas dos veces.

5.6.2.1. Diseño y Selección de Métricas

Cada métrica se presentan en base a una plantilla, donde constan los atributos principales de acuerdo a la Ontología de Métricas e Indicadores: Nombre, Objetivo, Método de Medición, Fórmula, Escala, Tipo de Escala y Unidad.

Las métricas presentadas en esta sección son métricas indirectas y corresponden a cada uno de los atributos del árbol de requerimientos representado en la Figura 5.1.

Además de las métricas necesarias para cuantificar los atributos presentes en el árbol de requerimientos fue necesario considerar otras, cuyos atributos no están en el árbol pero que son necesarias y, en el contexto de un proyecto de evaluación, se deben registrar y almacenar igual que las otras métricas. Se trata en muchos casos de métricas directas, que constituyen la base para la medición, o bien de métricas indirectas que forman parte del método de cálculo de alguna de las métricas de los atributos que sí están en el árbol de requerimientos.

Un ejemplo es la métrica *Tiempo de Tarea Completada* (TtC, A.2.2.3.1.1), que permite conocer el tiempo que llevó a un usuario completar correctamente una tarea, y forma parte de la fórmula de cálculo de otra métrica indirecta, como *Tiempo Total de Tareas Completadas* (TtCc, A.2.2.2.2.1). Sin embargo, la métrica TtC no cuantifica específicamente un atributo del árbol de requerimientos.

Las métricas directas e indirectas que forman parte de los respectivos métodos de cálculo se detallan en el Apéndice A. Por ejemplo, la métrica indirecta *Proporción de Tareas Completadas sobre Tareas Propuestas* (A.1.2.2.1.1), necesita para el método de cálculo las métricas directas *Número de Tareas Propuestas* (A.1.2.3.1.1) y *Número de tareas Completadas* (A.1.2.3.2.1).

En relación a los métodos de medición, es necesario acotar que no se emplearon herramientas que los automatizaran, excepto para la medición del tiempo de tarea, que se realizó en forma semiautomática con ayuda de la herramienta indicada en la sección 5.4.5. Para conocer la cantidad de tareas completas o las metas parciales completadas correctamente se realizó el conteo manual, a partir de un seguimiento observacional, tanto de la videofilmación como de la secuencia de pantallas, donde quedó registrada la actividad de los usuarios.

Las métricas detalladas a continuación permiten cuantificar los atributos especificados en el árbol de requerimientos de la 5.1.

5.6.2.1.1. Concepto Calculable (Característica): *Eficacia*

5.6.2.1.1.1. Atributo: *Compleitud de Tareas.*

Definición: proporción de las tareas completadas por los usuarios respecto de las tareas propuestas.

Objetivo: permite conocer en qué medida las tareas propuestas han sido completadas por los usuarios. A diferencia del atributo *Eficacia de Tarea*, aquí se toman en cuenta sólo las tareas que han sido completadas correctamente en su totalidad.

5.6.2.1.1.1. Nombre de la Métrica: *Promedio de la Proporción de Tareas Completadas sobre Tareas Propuestas para Todos los Usuarios* (o sea n usuarios realizando m tareas), P_{CT} .

Interpretación: $0 \leq P_{CT} \leq 1$, cuanto más próximo a 1 mejor,

Objetivo: calcular, para todos los usuarios de un perfil, el promedio de la proporción de tareas completadas en relación a las tareas propuestas.

Método de Cálculo(Fórmula):
$$P_{CT} = \frac{\sum_{j=1}^{j=n} CT(j)}{n},$$

n : Cantidad de Usuarios, CT:Compleitud de Tareas (A.1.2.2.1.1),

Escala: numérica.

Tipo de Escala: proporción

Unidad (tipo, descripción): Tarea, proporción de tareas completadas sobre tareas propuestas.

5.6.2.1.1.2. Atributo: *Eficacia de Tarea*

Definición: correctitud en la realización total o parcial de las tareas, para todos los usuarios.

Objetivo: permite conocer en qué medida el producto puede ser utilizado eficazmente por los usuarios en la realización de tareas típicas. Se considera que una tarea completada parcialmente otorga igualmente un cierto grado de eficacia, en la medida en que las metas parciales se cumplen correctamente.

5.6.2.1.1.2.1. Nombre de la Métrica: *Promedio de la Proporción de Tareas Completadas por Todos los Usuarios*, P_{ETT_u} .

Interpretación: $0 \leq P_{ETT_u} \leq 1$, cuanto más próximo a 1 mejor

Objetivo: Calcular el promedio general de la proporción de los objetivos completados correctamente, o sea de todos los usuarios de un mismo perfil en relación a todas las tareas propuestas.

Método de Cálculo(Fórmula):
$$P_{ETT_u} = \frac{\sum_{j=1}^{j=n} P_{ETT_u}(j)}{n},$$

n : número de usuarios que intervienen en el estudio, P_{ETT_u} es la métrica definida en A.1.1.2.1.1.

Escala: Numérica

Tipo de Escala: Proporción

Unidad (tipo, descripción): tarea, promedio de la proporción de tareas realizadas correctamente por todos los usuarios

5.6.2.1.2. Concepto Calculable (Característica): *Productividad*

5.6.2.1.2.1. Atributo: *Eficiencia en Relación a Completitud de Tareas.*

Definición: productividad de los usuarios en las tareas completadas totalmente.

Objetivo: conocer el nivel de productividad obtenido por los usuarios, considerando sólo aquellas tareas que fueron totalmente completadas

5.6.2.1.2.1.1. Nombre de la Métrica: *Promedio de Proporción de Eficiencia de Tareas Completadas.*

Interpretación: $0 < P_EFtRcTu$, cuanto mayor, mejor.

Objetivo: Calcular el promedio de eficiencia general, para todos los usuarios intervinientes, en la realización de las tareas propuestas.

$$\text{Método de Cálculo(Fórmula): } P_EFtRcTu = \frac{\sum_{j=1}^{j=n} EFtRc(j)}{n}$$

n : número de usuarios que intervienen en el test., $EftRc$: *Proporción de Eficiencia de Tareas Completadas*, (A.2.2.2.1.1).

Escala: numérica.

Tipo de Escala: intervalo.

Unidad: tarea/minutos.

5.6.2.1.2.2. Atributo *Eficiencia en relación a Eficacia.*

Definición: productividad obtenida por los usuarios realizando las tareas completadas parcial o totalmente.

Objetivo: conocer el nivel de productividad obtenido por los usuarios, considerando aquellas tareas que fueron parcial o totalmente completadas.

5.6.2.1.2.2.1. Nombre de la Métrica: *Promedio de Eficiencia de Tareas para Todos los Usuarios considerando Eficacia*, P_EFtTu .

Interpretación: $0 \leq P_EFtTu$, cuanto más grande, mejor.

Objetivo: Calcular el promedio de eficiencia general, para todos los usuarios, en la realización de las tareas propuestas.

Referencia: Apéndice A, en la definición de la métrica A.6.2.1.2.1.2.

$$\text{Método de Cálculo(Fórmula): } P_EFtTu = \frac{\sum_{j=1}^{j=n} P_EFt1u(j)}{n}$$

n : número de usuarios intervinientes en el test, P_EFt1u : *Promedio de Eficiencia de Tareas para un Usuario considerando Eficacia* (A.2.1.2.1.1).

Escala: numérica.

Tipo de Escala: **proporción.**

Unidad: 1/tiempo(minutos).

5.6.2.1.3. Concepto Calculable (Característica): Satisfacción

5.6.2.1.3.1. Atributo: Satisfacción.

Definición: es el nivel de satisfacción expresado por los usuarios en relación al producto evaluado.

Objetivo: permite conocer en que medida los usuarios se sintieron satisfechos empleando el producto. Si bien representa el aspecto subjetivo de la evaluación de calidad en uso, es posible cuantificarlo. La evaluación subjetiva toma en cuenta aspectos del producto tales como: facilidad para ser aprendido y utilizado, adecuación del diseño y presentación de la interfaz, etc.

5.6.2.1.3.1.1. Nombre de la Métrica: *Promedio del grado de Satisfacción, P_GS.*

Interpretación $0 \leq P_{GS} \leq 40$, cuanto más próximo a 40, mejor.

Objetivo Calcular el promedio del grado de satisfacción para todos los usuarios.

Método de Cálculo(Fórmula):
$$P_{GS} = \frac{\left(\sum_{j=1}^{j=n} GS(j) \right)}{n}$$

n: número de usuarios que intervienen en el test, GS: *Satisfacción para un Usuario* (A.3.1.2.1.1).

Escala: numérica.

Tipo de Escala: proporción.

Unidad: grado de satisfacción.

Aspectos relativos al cuestionario pueden consultarse en el Apéndice D.

5.6.2.2. Diseño de los Indicadores

La necesidad de los indicadores proviene del hecho de que solamente con los valores de las métricas no es posible estimar o evaluar un concepto calculable. Además, facilitan la evaluación a partir de los criterios de decisión asociados.

Para este caso de estudio se diseñaron los siguientes indicadores: *Grado de Cumplimiento respecto a proporción de tareas completadas por todos los usuarios, Grado de Cumplimiento respecto a Proporción de Tareas Completadas sobre Tareas Propuestas, Grado de Cumplimiento respecto de Eficiencia de Tareas considerando Eficacia, Grado de Cumplimiento respecto a Proporción de Eficiencia de Tareas Completadas y Grado de Cumplimiento respecto a Satisfacción de Usuarios.*

En todos los casos un puntaje en el rango “no satisfactorio” indica la urgencia de poner en práctica acciones de mejora.

Los ítems de la plantilla de presentación de los indicadores corresponden a los atributos especificados en el modelo de evaluación, detallado e ilustrado en la sección 3.3.1.3.

Indicadores Relacionados al Concepto Calculable *Eficacia*

5.6.2.2.1. Nombre: *Grado de Cumplimiento Respecto a Proporción de Tareas Completadas sobre Tareas Propuestas, %P_CT_P.*

Modelo elemental: $\%P_CT_P = P_CT \times 100$, *P_CT: Promedio de la Proporción de Tareas Completadas sobre Tareas Propuestas para Todos los Usuarios, definida en 5.6.2.1.1.1.1.*

Criterio de Decisión

Rango (niveles de aceptabilidad):

$0 \leq \%P_CT_P \leq 45$: no satisfactorio.

$45 < \%P_CT_P \leq 70$: marginal (regular).

$70 < \%P_CT_P \leq 100$: satisfactorio.

Escala: numérica.

Tipo de Escala: proporción.

Unidad: porcentaje.

5.6.2.2.2. Nombre: *Grado de Cumplimiento respecto a Proporción de Tareas Completadas por Todos los Usuarios, %P_ETTu_P.*

Modelo elemental: $\%P_ETTu_P = P_ETTu \times 100$, *P_ETTu: Promedio de la Proporción de Tareas Completadas por Todos los Usuarios, definida en 5.6.2.1.1.2.1.*

Criterio de Decisión

Rango (niveles de aceptabilidad):

$0 \leq \%P_ETTu_P \leq 45$: no satisfactorio.

$45 < \%P_ETTu_P \leq 70$: marginal(regular).

$70 < \%P_ETTu_P \leq 100$: satisfactorio.

Escala: numérica.

Tipo de Escala: proporción.

Unidad: porcentaje.

Indicadores Relacionados al Concepto Calculable *Productividad*.

5.6.2.2.3. Nombre: *Grado de Cumplimiento respecto a Proporción de Eficiencia de Tareas Completadas, %P_EFtRcTu_P.*

Modelo elemental: $\%P_EFtRcTu_P = P_EFtRcTu \times 100$, *P_EftRcTu: Promedio de Proporción de Eficiencia de Tareas Completadas, definida en 5.6.2.1.2.1.1.*

Criterios de Decisión

Rango (Niveles de aceptabilidad):

$0 \leq \%P_EFtRcTu_P \leq 45$: no satisfactorio.

$45 < \%P_EFtRcTu_P \leq 70$: marginal (regular).

$70 < \%P_EFtRcTu_P \leq 100$: satisfactorio.

Escala: numérica.

Tipo de Escala: proporción.

Unidad: porcentaje.

5.6.2.2.4.Nombre: Grado de Cumplimiento Respecto de Eficiencia de Tareas considerando Eficacia, %P_EftTu_P.

$$\%P_{EftTu_P} = 100$$

$$\text{si } P_{EftTu} = P_{EftTuMax}$$

Modelo elemental: $\text{sino } \%P_{EftTu_P} = \frac{(P_{EftTu} - P_{EftTuMin})}{(P_{EftTuMax} - P_{EftTuMin})} \times 100$

$$\text{si } P_{EftTuMin} < P_{EftTu} < P_{EftTuMax}$$

$$\text{sino } \%P_{EftTu_P} = 0$$

$$\text{si } P_{EftTu} = P_{EftTuMin}$$

P_{EftTu} : Promedio de Eficiencia de Tareas para Todos los Usuarios considerando Eficacia, definida en 5.6.2.1.2.2.1

$P_{EftTuMin}$: Peor valor obtenido por un usuario para la métrica P_{EftTu} .

$P_{EftTuMax}$: Mejor valor obtenido por un usuario para la métrica P_{EftTu} .

Criterios de Decisión (rango o niveles de aceptabilidad):

$0 \leq \%P_{EftTu_P} \leq 45$: no satisfactorio.

$45 < \%P_{EftTu_P} \leq 70$: marginal (regular).

$70 < \%P_{EftTu_P} \leq 100$: satisfactorio.

Escala: numérica.

Tipo de Escala proporción.

Unidad: porcentaje.

Indicadores Relacionados al Concepto Calculable *Satisfacción*

5.6.2.2.1.5. Nombre: grado de Cumplimiento respecto a Satisfacción de usuarios, %P_GS_P.

Modelo elemental: $\%P_{GS_P} = P_{GS} \times 2.5$

Criterios de Decisión (rango o niveles de aceptabilidad)

$0 \leq \%P_{GS_P} \leq 45$: no satisfactorio.

$45 \leq \%P_{GS_P} \leq 70$: marginal (Regular).

$70 \leq \%P_{GS_P} \leq 100$: satisfactorio.

Escala: numérica.

Tipo de Escala: proporción.

Unidad: porcentaje.

5.7 Definir la Evaluación Global²²

Para poder finalizar la evaluación es necesario definir un *indicador global*, que permita evaluar o estimar globalmente el concepto calculable calidad en uso.

Para el caso de estudio sobre “QPlus Campus Virtual” la asignación de los pesos correspondientes a los atributos y subconceptos fue equivalente, tal como puede visualizarse en la 5.2, es decir que se asignó igual importancia

²² Corresponde al *Diseño de la evaluación* del Marco de Medición y Evaluación

a cada atributo dentro de cada subconcepto y, del mismo modo, a cada subconcepto que conforma el concepto calculable calidad en uso. Este criterio se empleó teniendo en cuenta que se trataba de un estudio exploratorio.

Dependiendo de la necesidad de información y del contexto de la evaluación, los evaluadores podrían acordar la asignación de diferentes pesos para distintos atributos o subconceptos de un mismo nivel, como una forma de reflejar la mayor importancia de unos respecto de otros en relación a los objetivos del caso de estudio. Además, se podrían realizar encuestas a los usuarios para determinar y ponderar los pesos de los atributos y subconceptos, tal como propone [YIP05].

- 1 Calidad en Uso
 - 1.1 Eficacia <0,33>
 - 1.1.1 *Complejidad de Tareas* <0,5>
 - 1.1.2 *Eficacia de Tarea* <0,5>
 - 1.2 Productividad <0,33>
 - 1.2.1 *Eficiencia en relación complejidad de tareas* <0,5>
 - 1.2.2 *Eficiencia en relación a Eficacia* <0,5>
 - 1.3 Satisfacción <0,33>
 - 1.3.1 *Satisfacción (cuestionario)* <1,0>

Figura 5.2. Asignación de pesos a subconceptos y atributos.

5.8. Definición de Modelos de Agregación e Implementación de la Evaluación Global

Para este caso de estudio, al momento de seleccionar un modelo de agregación y cálculo del puntaje, se consideró más apropiado un modelo de agregación lógico que uno simplemente lineal y aditivo. Esta decisión está justificada en el hecho de que un modelo de agregación lógico multicriterio permite manejar consistentemente relaciones de simultaneidad, reemplazabilidad y neutralidad, además de superar el problema de modelar requerimientos obligatorios que padece un modelo lineal y aditivo.

En esta evaluación se empleó el modelo LSP [DUJ96], que se encuentra implementado en la herramienta WebQEM_Tool. Esta herramienta se empleó en la ejecución de la evaluación, para el cálculo del indicador global de calidad en uso.

A nivel de subconceptos el operador lógico empleado fue C-, un operador para modelar una relación de cuasi-conjunción débil. Esto significa que una calificación cero en alguno de los atributos de entrada no producirá un cero en la salida, aunque castigará a la misma. La 5.3 permite visualizar gráficamente la asignación de los operadores lógicos.

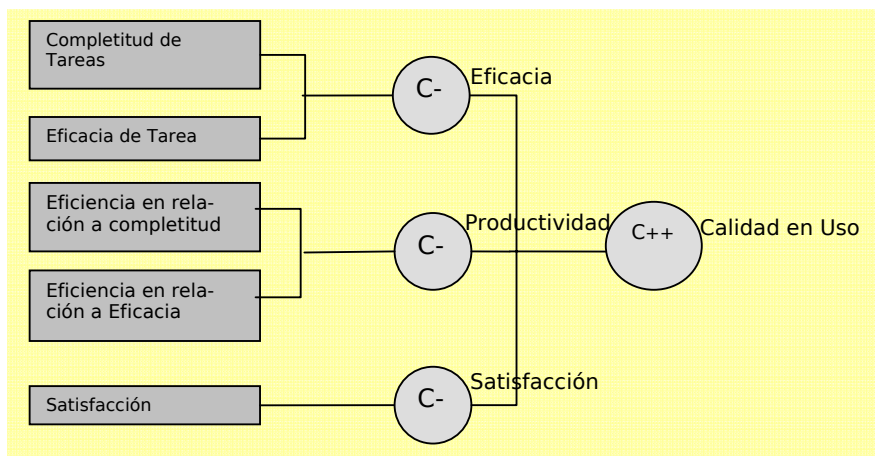


Figura 5.3. Operadores lógicos en la estructura de agregación para la evaluación global.

A nivel de concepto calculable –calidad en uso–, se empleó el operador lógico C++ para modelar una relación de cuasi-conjunción fuerte, lo que implica que para obtener valores distintos de cero de salida es obligatoria la existencia de valores distintos de cero en las subcaracterísticas de entrada. Además, cuanto mayores sean esos valores de entrada, mejor será el indicador global.

5.9 Conclusiones

En este capítulo se ha presentado el diseño del caso de estudio para evaluar calidad en uso. Los resultados se encuentran en el capítulo 6, ampliado a su vez con información acerca del dominio de la aplicación seleccionada, y también con aspectos que tienen que ver con la adaptación de la metodología utilizada para la evaluación de calidad en uso. Se ha enfatizado la necesidad de especificar adecuadamente el contexto (sección 5.4) en el que se desarrolló la evaluación y el diseño de las tareas (sección 5.3) que llevaron a cabo los usuarios, como aspectos específicos del diseño de una evaluación de calidad en uso.

También se ha presentado un conjunto de métricas e indicadores diseñados para evaluar calidad en uso, teniendo como referencia el Marco de Medición y Evaluación presentado en el Capítulo 3. La importancia de esta propuesta, implementada en el caso de estudio sobre una aplicación Web, radica en la definición precisa de cada métrica e indicador y en una instancia del modelo de Calidad en Uso. Esto permitió no sólo registrar y almacenar los valores obtenidos en el proceso de medición, sino también los metadatos asociados para ser reutilizados en proyectos semejantes.

En relación al primero de los objetivos del caso de estudio: “*validar una propuesta de evaluación sistemática de calidad en uso para aplicaciones Web*”, se puede afirmar que la metodología y el marco de medición em-

pleados resultaron suficientemente robustos y a la vez flexibles, como para evaluar calidad en uso en forma metódica y sistemática.

El marco y la metodología resultaron robustos porque habiéndose probado anteriormente en casos de estudio de calidad de producto, cada fase y las actividades dentro de cada fase, tienen también sentido y resultan apropiadas para diseñar la evaluación de calidad en uso. Se evidenciaron flexibles porque permitieron la integración de actividades propias de la evaluación de calidad en uso, tales como la definición de tareas y la especificación del contexto, y el uso de un modelo específico de calidad en uso (distinto al de calidad de producto, por ejemplo), en la fase de definición y especificación de los requerimientos de calidad.

Sobre el segundo objetivo: *“conocer la calidad en uso de una aplicación Web particular, considerando un determinado perfil de usuario”*, las conclusiones se encuentran en el siguiente capítulo, donde se presentan los resultados del caso de estudio.

Capítulo 6: Ejecución del Caso de Estudio

6.1. Introducción

En el capítulo 5 se presentó la definición y diseño de la evaluación para el caso de estudio. En este capítulo se exponen los resultados de la ejecución del estudio y también el análisis de los resultados.

Las fases de la metodología WebQEM que faltan presentar se corresponden con la ejecución de la evaluación; es decir aquellas instancias donde se recolectan los datos, se computan las métricas e indicadores elementales, y también se calcula el indicador global, que permite obtener un puntaje para el concepto calculable calidad en uso.

Una vez realizado el cálculo de los indicadores parciales y el indicador global, están dadas las condiciones para que los evaluadores puedan analizar los resultados, contrastarlos con la hipótesis planteada para el caso de estudio, elaborar conclusiones y realizar las recomendaciones pertinentes.

El capítulo 6 está organizado de la siguiente manera: la Sección 6.2 trata de la implementación de la evaluación elemental que incluye la recolección de los datos (6.2.1), el cómputo de las métricas (6.2.2). y el cómputo de los indicadores elementales (6.2.3.). En la sección 6.3 se detalla la implementación de la evaluación global, en la sección 6.4. se analizan los resultados obtenidos y se presentan las conclusiones del caso de estudio. Finalmente, en la sección 6.5. están las conclusiones del capítulo.

6.2. Implementar la Evaluación Elemental

Implementar la evaluación elemental significa básicamente ejecutar el proceso de recolección de datos, computar las métricas e indicadores elementales y documentar los resultados encontrados.

Teniendo en cuenta las consideraciones sobre el diseño de tareas para la evaluación de calidad en uso realizadas en la sección 5.3.3, se diseñaron las siguientes tareas para ser completadas por los participantes del estudio (ver, además el Apéndice C):

Tarea 1. Navegar contenidos disponibles para la asignatura, en base a una pregunta sobre material de estudio ya utilizado.

Tarea 2. Realizar una consulta a un profesor, sobre el tema desplegado en la tarea anterior, empleando el mecanismo de comunicación asincrónico apropiado.

Tarea 3. Participar en un foro de opinión, debiéndose tipear un mensaje en la sección pertinente al tema de la opinión.

Tarea 4. Organizar el material de estudio. Aquí los alumnos debían transferir contenido entre distintas carpetas, creadas en el WbIS y gestionadas por los propios usuarios, debiendo quedar establecido que el material de estudio de una clase había sido revisado completamente.

6.2.1 Recolección de los Datos: Procedimiento y Herramienta

En la sección 5.4.3 se señaló que las sesiones de trabajo de los usuarios habían sido videofilmadas y también registradas con un software que permite el registro de secuencias de pantalla de una sesión de trabajo. Estas sesiones de trabajo se llevaron a cabo entre los días 19/04/2004 y el 21/04/2004. La recolección de los datos correspondientes a tiempos de tarea, metas parciales alcanzadas y tareas completadas se realizó, por parte de los evaluadores, durante los dos días subsiguientes, a partir de los registros detallados de las sesiones de trabajo.

Además, durante las sesiones de trabajo los evaluadores tomaron nota de cuestiones que, a su criterio, podrían influir en la recolección de los datos o el cómputo de las métricas.

El procedimiento para obtención de los datos fue observacional, a partir del seguimiento de la video filmación como primera fuente de información y de la reproducción de secuencia de ventanas como respaldo. Excepto para conocer el tiempo insumido en completar las tareas, el resto de las métricas directas fue obtenido manualmente.

6.2.2. Cómputo de las Métricas

Para cada métrica directa, la recolección de datos fue guiada por la secuencia de pasos que indican los métodos de medición.

Una vez obtenidos los datos que permiten conocer los valores de las métricas directas, fue factible computar las métricas indirectas, a partir de los respectivos métodos de cálculo.

Si bien se realizaron estudios preliminares con usuarios expertos, estos resultados no fueron tenidos en cuenta para predecir los resultados de los usuarios seleccionados para el caso de estudio, correspondientes a un perfil diferente. Sin embargo, se esperaban respuestas al cuestionario sobre satisfacción con puntajes muy favorables (75% o más), considerando como antecedente el puntaje obtenido en un estudio anterior, del año 2003. En ese estudio se empleó el mismo instrumento para una población semejante, aunque sobre un número mayor de respuestas (veintitrés).

El criterio para presentar los datos de la medición y cálculo de las métricas está dado por el árbol de requerimientos presentado en la sección 5.5. Primero se encuentran las métricas que permiten mensurar los atributos del subconcepto *Eficiencia*, luego las métricas de los atributos de

Productividad y finalmente la de *Satisfacción*. En todos los casos las métricas están precedidas por los atributos relacionados.

6.2.2.1. Concepto Calculable (Característica): *Eficacia*

6.2.2.1.1 Atributo: *Complejidad de Tareas*

En primer término se abordó la medición y cálculo de las métricas necesarias para evaluar el atributo *Complejidad de Tareas*, correspondiente al subconcepto *Eficacia* del árbol de requerimientos.

Para la métrica *Número de Tareas Completadas Correctamente* (A.1.2.3.2.1) y de acuerdo a su método de medición, se contaron para cada usuario, las tareas completadas correctamente. Es decir, cada tarea en la que el usuario completó correctamente todas y cada una de las submetas definidas. En la Tabla 6.1 se presenta un resumen de los resultados obtenidos.

| Usuario | Resultados |
|-----------|------------|
| Usuario 1 | 2 |
| Usuario 2 | 1 |
| Usuario 3 | 1 |
| Usuario 4 | 1 |
| Usuario 5 | 4 |
| Usuario 6 | 4 |

Tabla 6.1. Resultados obtenidos por los usuarios para la métrica *Número de Tareas Completadas Correctamente*.

Luego fue posible calcular la métrica *Proporción de Tareas Completadas sobre Tareas Propuestas*, CT (A.1.2.2.1.1), cuyo método de cálculo es el cociente entre las *Tareas Completadas Correctamente* (A.1.2.3.2.1) y las *Tareas Propuestas* (A.1.2.3.1.1). Considerando que el número de tareas propuestas es el mismo para todos los usuarios (cuatro), los resultados obtenidos son los que se visualizan en la Tabla 6.2.

| Usuario | Resultados |
|-----------|------------|
| Usuario 1 | 0,5 |
| Usuario 2 | 0,25 |
| Usuario 3 | 0,25 |
| Usuario 4 | 0,25 |
| Usuario 5 | 1 |
| Usuario 6 | 1 |

Tabla 6.2. Resultados obtenidos por los usuarios para la métrica *Proporción de Tareas Completadas sobre Tareas Propuestas*.

Evidentemente los mejores resultados correspondieron a los usuarios que habían completado correctamente todas las tareas. La interpretación de la métrica indica que el resultado debe ser entre 0 y 1 y que es mejor cuanto más próximo a uno se encuentre.

En el paso siguiente fue posible calcular la métrica *Promedio de la Proporción de Tareas Completadas sobre Tareas Propuestas para Todos los*

Usuarios (A.1.2.1) realizando el cociente entre la sumatoria de los resultados obtenidos para la métrica *Proporción de Tareas Completadas sobre Tareas Propuestas* y el número de usuarios (seis). El resultado obtenido fue **0,542**.

6.2.2.1.2. Atributo: Eficacia de Tarea

Para poder calcular la métrica relacionada directamente con el atributo *Eficacia de Tarea* fue necesario primero coleccionar datos acerca del cumplimiento, parcial o total, de las submetas por parte de los usuarios para cada tarea. Las submetas de cada tarea están definidas en el apéndice C. Para la métrica *Proporción de Tarea Completada Correctamente* (A.1.1.2.2.1) los resultados obtenidos son los que se pueden visualizar en la Tabla 6.3.

| Usuario | Resultados | | | |
|-----------|------------|---------|---------|---------|
| | Tarea 1 | Tarea 2 | Tarea 3 | Tarea 4 |
| Usuario 1 | 1 | 0,67 | 0,6 | 1 |
| Usuario 2 | 1 | 0 | 0 | 0 |
| Usuario 3 | 1 | 0 | 0 | 0 |
| Usuario 4 | 0 | 1 | 0,8 | 0,67 |
| Usuario 5 | 1 | 1 | 1 | 1 |
| Usuario 6 | 1 | 1 | 1 | 1 |

Tabla 6.3. Resultados obtenidos por los usuarios para la métrica *Proporción de Tarea Completada Correctamente*.

Los resultados de la Tabla 6.3 se obtuvieron de sustraer a uno la fracción correspondiente a la proporción de tarea *no* completada correctamente, tal como indica el método de cálculo de la métrica.

En base a la métrica anterior fue posible calcular la métrica *Promedio de la Proporción de Tareas Completadas por un Usuario* (A.1.1.2.1.1). Los resultados, reflejados en la Tabla 6.4, se obtuvieron a partir del cociente entre la sumatoria de los valores de la *Proporción de Tarea Completada Correctamente* y la cantidad de tareas (cuatro).

| Usuario | Resultado |
|-----------|-----------|
| Usuario 1 | 0,82 |
| Usuario 2 | 0,25 |
| Usuario 3 | 0,25 |
| Usuario 4 | 0,62 |
| Usuario 5 | 1,00 |
| Usuario 6 | 1,00 |

Tabla 6.4. Resultados obtenidos por los usuarios para la métrica *Promedio de la Proporción de Tareas Completadas por un Usuario*.

A continuación se obtuvo el valor de la métrica *Promedio de la Proporción de Tareas Completadas por Todos los Usuarios* (A.1.1.1). Para ello se realizó el cociente entre la sumatoria de los resultados obtenidos para la métrica *Promedio de la Proporción de Tareas Completadas por un Usuario* y

el número de usuarios participantes (seis). El resultado obtenido fue **0,656**. La métrica calculada en este paso permitió mensurar el atributo *Eficacia de Tarea*.

6.2.2.2. Concepto Calculable (Característica): **Productividad**

6.2.2.2.1. Atributo: **Eficiencia en relación a Completitud de Tareas**

La medición del atributo *Eficiencia en Relación a Completitud de Tareas* refleja la productividad alcanzada por los usuarios, empleando el producto evaluado, pero únicamente en relación a las tareas que fueron completadas en su totalidad.

En primer lugar se calculó la métrica *Proporción de Tareas Completadas sobre Tareas Propuestas*, presentada en la Tabla 6.5, calculada como el cociente entre la métrica *Número de Tareas Completadas Correctamente* (A.1.2.3.2.1, Tabla 6.1) y el número total de tareas (cuatro).

| Usuario | Resultados |
|-----------|------------|
| Usuario 1 | 0,5 |
| Usuario 2 | 0,25 |
| Usuario 3 | 0,25 |
| Usuario 4 | 0,25 |
| Usuario 5 | 1 |
| Usuario 6 | 1 |

Tabla 6.5. Resultados obtenidos por los usuarios para la métrica *Proporción de Tareas Completadas sobre Tareas Propuestas*.

Según la interpretación de la métrica *Proporción de Tareas Completadas sobre Tareas Propuestas* cuanto más próximo a 1 es el resultado, mejor resulta el desempeño del usuario usando el producto.

Más adelante, con el objetivo de calcular las otras métricas relacionadas a Productividad, se midió el *Tiempo Total de Tareas Completadas*, esto es el tiempo insumido en llevar adelante solamente aquellas tareas que fueron totalmente completadas y en forma correcta. Las mediciones del *Tiempo Total de Tareas Completadas* (A.2.2.2.2.1), expresadas en formato *mm:ss,dd* (minutos:segundos, décimas de segundo) están en la Tabla 6.6.

| Usuario | Resultados |
|-----------|------------|
| Usuario 1 | 08:16,3 |
| Usuario 2 | 03:38,0 |
| Usuario 3 | 06:51,0 |
| Usuario 4 | 01:31,0 |
| Usuario 5 | 11:44,0 |
| Usuario 6 | 06:25,0 |

Tabla 6.6. Resultados obtenidos por los usuarios para la métrica *Tiempo Total de Tareas Completadas*

Evidentemente aquellos usuarios que hayan completado un menor número de tareas podrían computar menos tiempo que aquellos que hayan completado todas. Sin embargo, el hecho de haber completado menos tareas resultará en una penalización en la métrica *Proporción de Eficiencia de Tareas Completadas*. Esta métrica, cuyos resultados están en la tabla 6.7, se calculó para cada usuario, como el cociente entre las métricas *Proporción de Tareas Completadas sobre Tareas Propuestas* (A.1.2.2.1.1, Tabla 6.5) y el *Tiempo Total de Tareas Completadas* (A.2.2.2.1, Tabla 6.6).

| Usuario | Resultados |
|-----------|------------|
| Usuario 1 | 0,497 |
| Usuario 2 | 0,497 |
| Usuario 3 | 0,495 |
| Usuario 4 | 0,499 |
| Usuario 5 | 0,498 |
| Usuario 6 | 0,499 |

Tabla 6.7. Resultados obtenidos por los usuarios para la métrica *Proporción de Eficiencia de Tareas Completadas*.

Luego fue posible calcular la métrica para el atributo *Eficiencia en relación a Completitud de Tareas*, denominada *Promedio de Proporción de Eficiencia de Tareas Completadas*. El método de cálculo para esta métrica (A.2.2.1) se indica como el cociente entre la sumatoria de los resultados de la métrica *Proporción de Eficiencia de Tareas Completadas* (A.2.2.2.1.1, Tabla 6.7) y el número de usuarios (seis). El resultado que se obtuvo fue **0,498**.

6.2.2.2.2. Atributo: Eficiencia en relación a Eficacia

Este atributo permite conocer el nivel de productividad obtenido por los usuarios, considerando aquellas tareas que fueron parcial o totalmente completadas en función del tiempo empleado para ello.

Para poder calcular la métrica el *Promedio de Eficiencia de Tareas para todos los Usuarios Considerando Eficacia*, se midió primero el tiempo empleado por los usuarios en llevar adelante las tareas. Las mediciones están en la Tabla 6.8.

En aquellas tareas donde los usuarios no completaron bien ni siquiera una submeta, no se consideró el tiempo utilizado, etiquetándose con n/a en la entrada correspondiente. Un ejemplo se encuentra en la Tabla 6.8 para el Usuario 3 en la Tarea 3.

| | Resultados | | | |
|-----------|------------|---------|---------|---------|
| Usuario | Tarea 1 | Tarea 2 | Tarea 3 | Tarea 4 |
| Usuario 1 | 07:15,0 | 09:30,0 | 02:09,0 | 01:01,3 |
| Usuario 2 | 03:38,0 | 08:40,0 | n/a | n/a |
| Usuario 3 | 06:51,0 | n/a | n/a | n/a |
| Usuario 4 | n/a | 01:31,0 | 03:11,0 | 01:33,0 |
| Usuario 5 | 07:06,0 | 03:33,0 | 00:21,0 | 00:44,0 |
| Usuario 6 | 00:36,0 | 02:07,0 | 02:55,0 | 00:47,0 |

Tabla 6.8. Resultados obtenidos por los usuarios para la métrica *Tiempo de Tarea*.

Las tareas cuyas submetas han sido completadas en una proporción muy baja o que directamente no hayan sido completadas, impactarán negativamente en otras métricas indirectas.

Por ejemplo, en la métrica siguiente, *Proporción de Eficiencia de Tarea considerando Eficacia* (A.2.1.2.2.1, Tabla 6.9), es elocuente la calificación cero asignadas a las tareas donde los usuarios no pudieron completar ninguna submeta, en correspondencia con las entradas de la Tabla 6.8 ya que no se pudo computar el tiempo.

| | Resultados | | | |
|-----------|------------|---------|---------|---------|
| Usuario | Tarea 1 | Tarea 2 | Tarea 3 | Tarea 4 |
| Usuario 1 | 1,98 | 1,32 | 1,20 | 2,00 |
| Usuario 2 | 1,99 | 0,00 | 0,00 | 0,00 |
| Usuario 3 | 1,98 | 0,00 | 0,00 | 0,00 |
| Usuario 4 | 0,00 | 2,00 | 1,59 | 1,34 |
| Usuario 5 | 1,98 | 1,99 | 2,00 | 2,00 |
| Usuario 6 | 2,00 | 1,99 | 1,99 | 2,00 |
| | | | | |

Tabla 6.9. Resultados obtenidos por los usuarios para la métrica *Proporción de Eficiencia de Tarea considerando Eficacia*.

La métrica que se presenta en la Tabla 6.9 se calcula como el cociente entre la *Proporción de Tarea Completada Correctamente* (A.1.1.2.2.1), y el *Tiempo Empleado por un Usuario para Completar una Tarea* (métrica A.2.1.3.1.1), excepto para los casos donde el tiempo no se computó (por no haberse completado bien ninguna submeta), en cuyo caso el resultado es cero.

Antes de calcular la métrica relacionada directamente con el atributo *Eficiencia en Relación a Eficacia*, se calculó el *Promedio de Eficiencia de Tareas para un Usuario considerando Eficacia* (A.2.1.2.1.1, Tabla 6.10), como el cociente entre la sumatoria de los resultados obtenidos por todos los usuarios en la métrica anterior y el número de tareas propuestas (cuatro).

| Usuario | Resultados |
|-----------|------------|
| Usuario 1 | 1,624 |
| Usuario 2 | 0,497 |
| Usuario 3 | 0,495 |
| Usuario 4 | 1,231 |
| Usuario 5 | 1,992 |
| Usuario 6 | 1,996 |

Tabla 6.10. Resultados obtenidos por los usuarios para la métrica *Promedio de Eficiencia de Tareas para un Usuario considerando Eficacia*.

En última instancia, se calculó el *Promedio de Eficiencia de Tareas para todos los Usuarios considerando Eficacia* (A.2.1.1), obtenido a partir del cociente entre la sumatoria de los resultados de todos los usuarios y el número de usuarios participantes. El resultado fue **1,306**.

6.2.2.3. Concepto Calculable (Característica): **Satisfacción**

Para evaluar la satisfacción de los usuarios se empleó un cuestionario compuesto de diez ítems (ver apéndice D). Cada ítem recibe una puntuación de 0 a 4, de modo que el rango para la puntuación de cada cuestionario, en forma global, estará entre 0(cero) y 40 (cuarenta).

Los resultados de cada cuestionario completado, reflejados por la métrica *Satisfacción para un Usuario* (A.3.1.2.1.1), se presentan en la Tabla 6.10. Las puntuaciones más altas, próximas a 40, reflejan una disposición más favorable de los usuarios respecto al producto utilizado.

| Usuario | Resultados |
|-----------|------------|
| Usuario 1 | 39,00 |
| Usuario 2 | 28,00 |
| Usuario 3 | 39,00 |
| Usuario 4 | 28,00 |
| Usuario 5 | 38,00 |
| Usuario 6 | 37,00 |

Tabla 6.10. Resultados obtenidos por los usuarios para la métrica *Satisfacción para un usuario*.

Para obtener un valor que represente la perspectiva global de los usuarios para este subconcepto se computó la métrica *Promedio del Grado de Satisfacción* (A.3.1.1). El cálculo se realizó como el cociente entre la sumatoria de los resultados de la métrica *Satisfacción para un Usuario* y el número de usuarios participantes(seis), siendo el resultado **34,83**.

6.2.3. Cómputo de los Indicadores Elementales

En este paso, una vez especificados los criterios, tal como consta en la Sección 5.6.2.2, y un modelo elemental para pasar de valores de métricas a indicadores, se realizó el cálculo correspondiente a los indicadores elementales. De este modo, es posible realizar una interpretación inicial de las métricas, al nivel más bajo (atributos) del árbol de requerimientos.

Al igual que con las métricas, el cómputo se realizó empleando una hoja de cálculo que permite expresar y calcular las fórmulas correspondientes a los indicadores elementales, partiendo de los resultados de las métricas relacionadas.

Los resultados se presentan, nuevamente, siguiendo la jerarquía del árbol de requerimientos, primero los indicadores para el subconcepto *Eficiencia*, luego para *Productividad* y finalmente para *Satisfacción*.

Para el cálculo de los indicadores elementales es necesario estandarizar las unidades en que están expresados. Esto permite una interpretación adecuada de las métricas relacionadas y facilita el cálculo de los indicadores parciales y del indicador global. En este caso, los métodos de cálculo para los distintos indicadores permiten obtener valores como porcentaje de modo que se puedan contrastar, en forma evidente, los resultados obtenidos versus los rangos de aceptabilidad establecidos en el diseño del caso de estudio.

6.2.3.1. Indicadores Relacionados al Concepto Calculable Eficacia

Los indicadores elementales relacionados a Eficacia permiten conocer cuál es el nivel de cumplimiento que se alcanzó para los atributos *Completitud de Tareas* y *Eficacia de Tarea*, cuando se comparan los valores de los indicadores con los niveles de aceptabilidad preestablecidos. En la Tabla 6.11 los indicadores siguientes están agrupados bajo el código de jerarquía 1.1 denominado *Nivel de Eficacia*.

6.2.3.1.1. Grado de Cumplimiento Respecto a Proporción de Tareas Completadas sobre Tareas Propuestas.

El cálculo del modelo elemental para este indicador (definido en 5.6.2.2.1) se obtuvo como el producto entre el valor de la métrica *Promedio de la Proporción de Tareas Completadas sobre Tareas Propuestas para Todos los Usuarios* (0,542, calculado en 6.2.2.1.1) y 100.

Se obtuvo **54,20%** como valor del indicador.

6.2.3.1.2. Grado de Cumplimiento Respecto a Proporción de Tareas Completadas por Todos los Usuarios

Para este indicador, el modelo elemental utilizado (definido en 5.6.2.2.2) expresa que el valor se obtiene como el producto entre el valor de la métrica *Promedio de la Proporción de Tareas Completadas por Todos los Usuarios* (0,656, calculado en 6.2.2.1.2) y 100.

Se obtuvo **65,60%** como valor del indicador.

6.2.3.2. Indicadores Relacionados al Concepto Calculable Productividad

Los indicadores elementales relacionados a Productividad permiten conocer cuál es el nivel de cumplimiento que se alcanzó para los atributos *Eficiencia en relación a Completitud de Tareas* y *Eficiencia en relación a Eficacia*, cuando se comparan los valores de los indicadores con los niveles de aceptabilidad preestablecidos. En la Tabla 6.11 los indicadores siguientes están agrupados bajo el código de jerarquía 1.2 denominado *Nivel de Productividad*.

6.2.3.2.1. Grado de Cumplimiento Respecto a Proporción de Eficiencia de Tareas Completadas

En este caso, el modelo elemental (definido en 5.6.2.2.3) se calculó como el producto entre la métrica Promedio de proporción de Eficiencia de tareas completadas (0,498, calculado en 6.2.2.2.1) y 100.

Se obtuvo **49,80%** como valor del indicador.

6.2.3.2.2. Grado de Cumplimiento Respecto de Eficiencia de Tareas Considerando Eficacia

Aquí el modelo elemental (definido en 5.6.2.2.4) no es un producto entre el valor de la métrica y un factor (100 en los casos anteriores), sino una fórmula donde debe calcularse el cociente entre la diferencia de la métrica con el valor mínimo (dividendo) sobre la diferencia entre los valores máximos y mínimos obtenidos por los usuarios (divisor). Luego puede obtenerse el valor del indicador, al multiplicarse el resultado del cociente anterior por 100. Los valores de las métricas necesarios para el método de cálculo de esta métrica, *Eficiencia de Tareas para todos los Usuarios considerando Eficacia* y *Promedio de Eficiencia de Tareas para un Usuario considerando Eficacia*, han sido calculados en la sección 6.2.2.2.2.

Se obtuvo **54,04%** como valor del indicador.

6.2.3.3. Indicador Relacionado al Concepto Calculable Satisfacción

El subconcepto Satisfacción no posee ningún atributo en el modelo de calidad en uso para este estudio, de modo que el indicador elemental, presentado en el código de jerarquía 1.3 de la Tabla 6.11 (Nivel de Satisfacción), puede compararse directamente con los valores del rango de aceptabilidad para conocer el nivel de satisfacción expresado por los usuarios. El indicador relacionado se presenta a continuación.

6.2.3.3.1. Grado de Cumplimiento respecto a Satisfacción de Usuarios

El indicador de satisfacción (definido en 5.6.2.2.1.5) se obtuvo como el producto entre la métrica *Promedio del Grado de Satisfacción* (34,83, calculado en 6.2.2.3) y 2,5. El factor 2,5 surge de la necesidad de expresar el indicador como porcentaje y considerando que el valor máximo de la métrica involucrada puede ser 40 (cuarenta).

Se obtuvo **87,08%** como valor del indicador.

6.3.Implementar la Evaluación Global

El objetivo principal de la evaluación global es obtener el valor del indicador global de calidad en uso. Para ello, es necesario, como paso preliminar, obtener los indicadores parciales. En la sección 5.8 se mencionaron los operadores lógicos empleados en cada caso y también la justificación para la selección del modelo de agregación.

Para el cálculo de los indicadores parciales y globales se utilizó WebQEMTool [OLS01], una herramienta que facilita el cálculo y la interpretación de los datos, el seguimiento de la evaluación y que además tiene mecanismos de documentación y recomendación incorporados.

Los resultados correspondientes a los indicadores parciales y al indicador global de calidad en uso se presentan en la Tabla 6.11.

| Código de Jerarquía | Concepto Calculable | Característica | Atributo | Preferencia |
|---------------------|--------------------------------|------------------------|--|---------------|
| 1 | Nivel de Calidad en Uso | | | 57,45% |
| 1.1 | | Nivel de Eficacia | | 59,70% |
| 1.1.1 | | | <i>Grado de Cumplimiento Respecto a Proporción de Eficiencia de Tareas Completadas</i> | 54,20% |
| 1.1.2 | | | <i>Grado de Cumplimiento Respecto de Eficiencia de Tareas Considerando Eficacia</i> | 65,60% |
| 1.2 | | Nivel de Productividad | | 51,89% |
| 1.2.1 | | | <i>Grado de Cumplimiento Respecto a Proporción de Eficiencia de Tareas Completadas</i> | 49,80% |
| 1.2.2 | | | <i>Grado de Cumplimiento Respecto de Eficiencia de Tareas Considerando Eficacia</i> | 54,04% |
| 1.3 | | Nivel de Satisfacción | | 87,08% |
| 1.3.1 | | | <i>Grado de Cumplimiento respecto a Satisfacción de Usuarios</i> | 87,08% |

Tabla 6.11. Resultados de los indicadores elementales, parciales y del indicador global de Calidad en Uso.

WebQEMTool permite seleccionar entre dos clases de modelo de agregación (aditivo y LSP) y una vez ingresados los pesos de cada atributo, los operadores lógicos y los puntajes de los indicadores elementales, realiza automáticamente el cálculo de los indicadores parciales y del indicador global.

6.4. Analizar los Resultados, Elaborar las Conclusiones y Documentar²³

En esta sección se presenta la información obtenida a partir de las actividades de análisis y comparación de los indicadores de calidad

²³ Corresponde a las *Recomendaciones* del Marco de Medición y Evaluación

(elementales, parciales y globales). También se discuten brevemente las posibles causas de los resultados obtenidos.

6.4.1. Análisis de Resultados y Recomendaciones

Una vez realizado el cálculo de preferencia global, a partir de los indicadores parciales, es posible analizar el grado de cumplimiento respecto a la hipótesis elaborada para el estudio. También es factible considerar el cumplimiento de objetivos parciales, esto es a nivel de características.

Considerando los resultados obtenidos, presentados en la Tabla 6.11, se observa que el cálculo del puntaje del indicador global de calidad en uso resultó por debajo del valor mínimo de aceptabilidad del 70,00%, arrojando un 57,45%. Además, teniendo en cuenta el puntaje a nivel características, esto es: *Eficacia* (59,70), *Productividad* (51,89) y *Satisfacción*: (87,08) no se cumple con la hipótesis de que “al menos a nivel de características, el sitio está por encima de un umbral de satisfacción aceptable”, ya que solo *Satisfacción* está por encima del nivel de aceptabilidad esperado del 70%.

A nivel de características, es alto el nivel de *Satisfacción* que demostraron los usuarios (87,08%), pero marginal (51,89%) el nivel de *Productividad* y algo mejor, pero igualmente debajo del mínimo aceptable (59,70%), el nivel de *Eficacia*. Consecuentemente, se puede inferir que la percepción subjetiva que tienen los usuarios del producto es muy favorable, a pesar de que no resultaron productivos ni altamente eficaces en la realización de las tareas propuestas.

Profundizando el análisis de indicadores elementales se encuentra que es marginal el puntaje obtenido en *Complejidad de Tareas* (54,20%) pero en aquellas que fueron completadas correctamente, que afectan el nivel de eficacia *Eficacia*, se obtuvo un nivel casi satisfactorio (65,60%). A partir de estos datos se puede decir que una parte importante de las tareas no pudieron ser concluidas, pero de entre las que sí lo fueron, la mayor parte fueron concluidas correctamente. Por otro lado están claramente en área marginal las calificaciones obtenidas para los atributos relacionados con *Productividad*, esto es *Eficiencia en relación Complejidad* (49,80%) y *Eficiencia en relación a Eficacia* (54,04%).

Dado que eficiencia está asociada al consumo de recursos para llevar adelante una tarea y en este caso el recurso mensurado fue tiempo, se puede concluir que, o bien hay ciertos mecanismos que no son lo suficientemente fáciles de aprender, o bien no son sencillos de utilizar para un usuario intencional pero no experto. Parece lógico entonces que sea necesaria una gran cantidad de tiempo para poder finalizar una tarea correctamente, si el usuario no tiene el entrenamiento necesario.

Al analizar las causas de los puntajes alcanzados, se debe recordar que el grupo de usuarios que participó del caso de estudio empleó “Qplus Campus Virtual” prácticamente sin entrenamiento. En la mayoría de los

casos el producto les fue presentado brevemente cuando se presentaron a inscribirse como alumnos de la Facultad de Ingeniería de la UNLPam, en diciembre de 2003.

Finalmente, es oportuno resaltar que la elección del modelo de agregación LSP para el cálculo del indicador global otorga gran flexibilidad a distintos enfoques de la evaluación. Permite, si se emplean los operadores adecuados, penalizar o relativizar los puntajes obtenidos para cada atributo y subconcepto. Llegado el caso, por ejemplo para una población de usuarios expertos, se podrían utilizar los operadores de tal manera que si tan sólo uno de los subconceptos no alcanza el mínimo de aceptabilidad (70% en este estudio) tampoco se pueda obtener un puntaje aceptable para el concepto calculable de más alto nivel.

6.5. Conclusiones

Las conclusiones respecto a la implementación del caso de estudio se consideran desde dos perspectivas: por un lado respecto de los resultados del caso de estudio en sí y, por otro, en relación a los metadatos obtenidos para las métricas e indicadores de calidad en uso.

Respecto a los resultados del estudio, una vez realizado el análisis (6.4.1), sólo queda por sintetizar que la causa principal por la que el producto evaluado no resultó eficaz y productivo, en la medida de lo esperado, fue la falta de un entrenamiento indispensable. Un WbIS es una aplicación Web con cierto grado de sofisticación y si bien está al alcance de un usuario intencional, requiere de un entrenamiento básico para ser utilizada de un modo eficaz y productivo.

Por otro lado, en relación a la percepción subjetiva y según los resultados obtenidos, los usuarios se sintieron satisfechos empleando el producto, por encima del nivel esperado. Esto revela aspectos positivos del producto en cuanto a la disposición favorable que generó entre ellos, ya sea porque la mayoría manifestó la decisión de continuar utilizándolo, o porque también indicó que aprendió a usarlo sin un gran esfuerzo y que lo recomendaría a otros usuarios de características semejantes.

En relación a la segunda perspectiva, se puede afirmar que los metadatos obtenidos, o sea la definición exhaustiva de atributos, métricas e indicadores, abren otra dimensión de análisis para otros estudios de calidad en uso de otros WbIS.

Estos metadatos pueden ser empleados para estudios subsiguientes de calidad en uso sobre el mismo producto o para productos similares, considerando hipótesis elaboradas a propósito de cuestiones específicas, por ejemplo: otros perfiles de usuario, algún servicio señalado como deficitario por los usuarios a partir de encuestas, etc.

Además, también es factible reutilizar el proceso de diseño de las métricas e indicadores, que para este estudio se inició con la consideración del marco de medición y evaluación, luego prosiguió con la selección y adaptación de un modelo para calidad en uso y finalizó con el diseño y especificación de los atributos, métricas e indicadores. Cada etapa se realizó teniendo como referencia las distintas fases de la metodología de evaluación WebQEM, quedando reflejado, a su vez, el vínculo existente entre la metodología y el marco de medición y evaluación.

Capítulo 7: Consideraciones Finales y Trabajos Futuros

7.1 Consideraciones Finales

El objetivo fundamental del trabajo detallado en esta tesis fue proponer un enfoque sistemático y disciplinado para evaluar calidad en uso de aplicaciones Web.

Para ello, se siguió un proceso riguroso de discusión y análisis del estado del arte y de la práctica, se evaluó y adoptó un marco de medición y evaluación y una metodología, se diseñaron métricas e indicadores para un caso de estudio y, en cada etapa, se extrajeron conclusiones de distinto alcance.

A lo largo de los capítulos 2, 3 y 4, se presentaron los insumos necesarios, tanto desde el punto de vista conceptual como metodológico, para establecer las bases de la propuesta. Luego, en los capítulos 5 y 6 se presentaron el diseño y los resultados de un caso de estudio donde se puso en práctica la iniciativa.

La discusión en las secciones iniciales del capítulo 2, acerca del significado y las relaciones entre términos como calidad y calidad en uso, ofrece una idea clara de que la definición de calidad no es un concepto absoluto y que existen diferentes perspectivas sobre calidad, teniendo en cuenta un producto de software o bien un producto de software en uso. En todo caso, se dijo que el objetivo final de un proceso de especificación, medición y evaluación de calidad es alcanzar un nivel aceptable de calidad en uso.

También se revisó la literatura y la práctica relacionadas, trazando una analogía en la evolución del significado de términos clave como usabilidad y calidad en uso. Esa evolución ha sido reflejada por las normas de estandarización, que actualmente indican la necesidad de evaluar no sólo la calidad de un producto de software en funcionamiento –calidad externa-, artefactos intermedios del proceso de desarrollo –calidad interna-, sino también la calidad percibida por usuarios reales en un contexto real de uso, o sea la calidad en uso.

Otra cuestión importante que surgió de este análisis comparativo fue que el término usabilidad, utilizado frecuentemente en la literatura, es empleado como sinónimo de calidad en uso, escapando en cierta medida al significado dado por el estándar ISO 9126-1 que lo ubica como una de las características de alto nivel para calidad interna y externa.

Además, se halló que en la opinión de varios autores, las actividades propias de la medición y evaluación de calidad en uso están siendo tenidas en cuenta como una disciplina de soporte que aporta sistemáticamente al proceso de desarrollo, apuntando a diseñar la calidad en uso de un producto

de software desde etapas tempranas, proactivamente, tal como se hace con la calidad interna y externa.

Como aporte concreto, las discusiones del capítulo 2 permitieron establecer algunos prerrequisitos para diseñar y evaluar calidad en uso (tópico 2.3.5.3) y precisiones acerca de qué se entiende por calidad de aplicaciones Web (2.4), ya que la propuesta de la tesis está dirigida específicamente a evaluar calidad en uso de aplicaciones Web.

En el capítulo 3 se discutió brevemente acerca de cierta falta de consenso en el uso de términos clave en el área de medición y evaluación, aunque se reconocieron aportes significativos de varios autores. La consideración y elección de una ontología y un marco de medición y evaluación permitió salvar las divergencias y afrontar el diseño de la propuesta principal de la tesis sobre una base conceptual consistente.

La ontología de métricas e indicadores tiene como propósitos básicos, entre otros, los siguientes: facilitar la comunicación entre los interesados de un proyecto de medición y evaluación y , también, conformar la base de conocimiento para el diseño y construcción de un marco de medición y evaluación.

El marco de medición y evaluación, por su parte, está basado en la ontología de métricas e indicadores y promueve un enfoque ingenieril de la evaluación de calidad. O sea, el empleo sistemático, disciplinado y cuantificable de principios y prácticas reconocidas de ingeniería a la medición y evaluación de calidad.

La metodología para llevar a cabo el proceso de medición y evaluación, donde se especifican las fases y pasos necesarios, debía ser consistente con el marco adoptado. La elección fue WebQEM, una metodología probada en la evaluación de producto software para la Web sobre distintos dominios (e-commerce, museos, académicos), con una correspondencia directa entre cada una de sus fases con los cuatro componentes principales del marco de medición y evaluación.

El desafío de esta elección provino de adaptar la metodología para evaluar calidad en uso. Por ejemplo, articular la interacción entre evaluadores expertos (para lo cual WebQEM se diseñó originalmente) y usuarios finales. A las actividades propuestas en el paso 1 "*Definición de las metas de evaluación y selección del perfil de usuario*", se añadió el diseño de las tareas que debían llevar a cabo los usuarios y la especificación del contexto en el que se llevarían a cabo esas tareas. Para ello, se consideraron como actividades propias de los expertos, la selección de tareas típicas para el perfil de usuario seleccionado y la definición de las características del contexto (las instalaciones, el equipamiento, etc.).

Para poner a prueba la propuesta se diseñó e implementó un caso de estudio, presentado en el capítulo 5. El caso de estudio se diseñó con dos objetivos, donde uno de ellos estaba en relación a la aplicación evaluada y el otro, de carácter general, era “validar una propuesta de evaluación sistemática de calidad en uso para aplicaciones Web”. Este último es relevante en relación al objetivo de la tesis y, para alcanzarlo, la primer tarea fue la adopción y diseño de un conjunto de métricas e indicadores para evaluar calidad en uso de aplicaciones Web.

Cada métrica e indicador tiene una definición precisa y en conjunto constituyen una instancia del modelo de calidad en uso. Las definiciones permitieron registrar y almacenar los valores obtenidos en el proceso de medición y, sobre todo, los metadatos asociados de manera que puedan ser utilizados en otros proyectos.

Estos metadatos constituyen entonces el valor agregado a los resultados propios del caso de estudio. Evaluaciones de calidad en uso sobre el mismo producto o para productos similares podrían partir de una línea base contando con estos metadatos como insumos básicos, ya que se trata de definiciones exhaustivas, debidamente validadas.

Se observó además, que la metodología y el marco de medición tienen un grado de fortaleza y flexibilidad adecuado como para evaluar calidad en uso en forma metódica y sistemática. Cada fase y las actividades dentro de cada fase, tienen sentido y son apropiadas para diseñar una evaluación para calidad en uso, además de los casos donde ya fueron probados para calidad de producto. La flexibilidad se visualizó en la facilidad de integrar actividades propias de la evaluación de calidad en uso, tales como la definición de tareas y la especificación del contexto. También en la posibilidad de emplear un modelo específico de calidad en uso (distinto al de calidad de producto), en la fase de definición y especificación de los requerimientos de calidad.

Analizando el proceso de diseño de las métricas e indicadores llevado a cabo se concluyó en que también es factible de ser reutilizado. Este proceso comenzó con la consideración del marco de medición y evaluación, siguió con la selección y adaptación de un modelo para calidad en uso; luego con el diseño y especificación de los atributos, métricas e indicadores para el caso de estudio y por último con la ejecución y conclusión. Cada etapa se realizó siguiendo las distintas fases de la metodología WebQEM, con las adaptaciones necesarias para evaluar calidad en uso.

Como síntesis de esta contribución es posible afirmar que el diseño, la medición y evaluación de calidad en uso de una aplicación Web son procesos complejos e interrelacionados. Incluyen aspectos objetivos y subjetivos, interacción entre expertos y usuarios finales, y no son generalizables a contextos diferentes. Para obtener resultados confiables y útiles es necesario

afrontar esos procesos con un enfoque ingenieril, o sea sistemático, disciplinado y consistente. La propuesta de esta tesis se desarrolló con el objetivo de ofrecer una solución en ese sentido, evitando atajos y soluciones de compromiso. Los resultados del caso de estudio representan en este sentido, aunque no puedan tomarse como definitivos, un incentivo valioso para continuar trabajando en esta línea.

7.2 Trabajos Futuros

Los trabajos futuros relacionados a la propuesta de esta tesis se presentan en dos niveles, uno en relación al dominio y al producto evaluados y el otro, respecto al aporte metodológico y objetivo principal de la tesis.

7.2.1. Trabajos Futuros relacionados al Dominio del Producto Evaluado

Uno de los aspectos que resultó valioso en el proceso de medición y evaluación fue integrar aspectos de evaluación subjetiva a otros, de carácter objetivo, de uso frecuente en evaluación de producto. Para el caso de estudio, de carácter exploratorio, se empleó un instrumento validado estadísticamente, pero de uso general para producto software. Pero dado que el perfil de los usuarios del dominio de e-Learning es generalmente específico, sería importante contar con un instrumento apropiado –un cuestionario en este caso- para la evaluación de satisfacción.

Por ello se propone el diseño y construcción de un cuestionario, donde la pertinencia de los ítems a ser incluidos sea evaluada por un grupo de expertos teniendo en cuenta aspectos relevantes del dominio. Dicho cuestionario deberá ser válido estadísticamente y podría tener una versión para cada tipo de usuario dentro del perfil: administradores, alumnos y profesores o tutores. Además, podrá diseñarse considerando cada ítem o grupo de ítems relacionados a un atributo del subconcepto *Satisfacción*. El método de medición o cálculo del ítem se corresponderá con el de la métrica asociada al atributo, y un indicador basado en esa métrica representará el grado de satisfacción del requerimiento elemental alcanzado por cada usuario o grupo de usuarios.

Por otro lado, para poder obtener conclusiones significativas respecto al producto evaluado es necesario reiterar el estudio de calidad en uso, tanto para el perfil de usuario ya evaluado como para otros perfiles relevantes, tales como profesores y administradores. En el primer caso, para obtener conclusiones que resulten consistentes a lo largo del tiempo y se realicen mejoras en base a las recomendaciones; en el segundo caso, para analizar las otras dimensiones que presenta un WbIS, ya sea para gestionar contenidos, personas, recursos y servicios de comunicación, destinados a los otros interesados del proceso de enseñanza/aprendizaje.

7.2.2. Trabajos Futuros en relación al Enfoque para Medir y Evaluar de Calidad en Uso

En el capítulo 2 se ha señalado que las aplicaciones Web exceden a lo que son las aplicaciones de software tradicionales, en cuanto a que deben considerarse, además de aspectos estrictamente funcionales y no funcionales, contenido y presentación. Así, la calidad en uso adquiere una importancia adicional para una cantidad y variedad de usuarios que son, en ciertos casos, difíciles de precisar. En este sentido, si bien existen iniciativas que permiten tanto la interoperabilidad como la indexación y recuperación de contenidos, se visualiza un campo de trabajo para desarrollar una ontología para e-Learning, que refuerce o amplíe el potencial de los estándares existentes, y permita no sólo la búsqueda e intercambio de contenidos entre plataformas heterogéneas sino también dotar de significado de aceptación general a los términos relevantes del dominio, tal como se ha visto, por ejemplo, con el modelo de métricas e indicadores.

Referencias

- [ALE99]. J. Alexander, M. Tate; *Web Wisdom: How to Evaluate and Create Information Quality on the Web*, Lawrence Erlbaum Pub. 1999.
- [AVG01]. P. Avgeriou, A. Papasalouros, S. Retalis; "*Learning Technology Systems: issues, trends, challenges*". Proceedings of the 1st International Organization for Science and Technology Education (IOSTE) Symposium in Southern Europe: Preparing future citizens, Paralimni, Cyprus, 29 April - 2 May, 2001.
- [AVG03]. P. Avgeriou, A. Papasalouros, S. Retalis, Skordlakis; "*Towards a Pattern Language for Learning Management Systems*" <<http://ifets.ieee.org/periodical/6-2/2.html>>, 2003 (ISSN 1436-4522).
- [BAS01]. R. Baskerville, L. Levine, J. Pries-Heje, S. Slaughter; "*How Internet Software Companies Negotiate Quality*". IEEE Computer, May, 2001.pag. 51-57.
- [BASS01]. L. Bass, B. John y J. Kates; *Achieving usability through software architecture*, Technical report SEI/CMU, 2001.
- [BEV91]. Nigel Bevan, Jurek Kirakowski y Jonathan Maissel. "*What is Usability*". Proceedings of the 4th International Conference on HCI, Stuttgart, September 1991.
- [BEV95a]. Nigel Bevan; "*Measuring usability as quality of use*". Software Quality Journal,4, 1995, pags. 115-150.
- [BEV95b]. Nigel Bevan. "*Usability is Quality in Use*", Proceedings of the 6th International Conference on Human Computer Interaction, Yokohama, Julio 1995. Anzai & Ogawa (eds), Elsevier.
- [BEV97]. Nigel Bevan, Motoei Azuma. "*Quality in use: Incorporating Human Factors into the Software Engineering Lifecycle*". IEEE, 1997, paginas 169-179.
- [BEV99]. Nigel Bevan. "*Quality in use: meeting user needs for Quality*". Journal of Systems and Software, 49(1), 89-96, 1999. Disponible en: <<http://www.usability.serco.com/papers/qiuse.pdf>>
- [BOE76]. B. W. Boehm , J. R. Brown, M.Lipow; "*Quantitative evaluation of software quality*". Proceedings of the 2nd international conference on Software engineering, San Francisco, California, United States, 1976. Pags.: 592 – 605.
- [BOE78]. B. W. Boehm, J. R. Brown, M. Lipow, J.L. MacLeod, M.J. Merrid "*Characteristics of Software Quality*". Elseiver Nort Holland, 1978.
- [COV02]. G. Covella , L. Olsina; "*Specifying Quality Attributes for Sites with E-Learning Functionality*". 2nd Ibero American Conference on Web Engineering (ICWE'02), Santa Fe, Argentina, pp 154-167, ISSN 1666-6526 (2002).
- [COV02a]. Guillermo COVELLA; Facultad de Ingeniería (UNLPam). "*Soporte via Internet para aspirantes alumnos con herramientas e-learning*". Resolución Concejo Directivo n° 107/02, Fecha: 02/11/2002. URL: <http://www.ingreso.ing.unlpam.edu.ar>.

- [DAV03]. J. Davies; D. Fensel; F. Van Harmelen; *Towards the Semantic Web: Ontology-driven Knowledge Management*. John Willey & Sons (2003)
- [DUJ96]. J.J. Dujmovic; "A Method for Evaluation and Selection of Complex Hardware and Software Systems". *22nd Int'l Conference for the Resource Management and Performance Evaluation of Enterprise CS*. CMG 96 Proceedings, Vol. 1, pp. 368-378 (1996).
- [FER01]. Xavier Ferré, Natalia Juristo, Helmut Windl, Larry Constantine; "Usability Basics for Software Developers". *IEEE SOFTWARE* January / February 2001.
- [FOL04]. Eelke Folmer, Jan Bosch; "Architecting for Usability: a Survey". *Journal of Systems and Software* 70(1/2) 61-78 , 2004.
- [GAR84]. Garvin, *What does "product quality" really mean?*, Sloane Mgmt. Review, fall 1984.
- [GIL69]. T. Gilb; "Weighted Ranking by Levels", *IAG Journal*, Vol 2 (2), pp. 7-22, 1969.
- [GOO00]. Peter Goodyear; "eLearning, knowledge work and working knowledge". *IST2000 Event, eLearning Futures session*, Nice, November 2000.
- [GRO04]. Heinz L.Grob, Frank Bensberg, Blasius Dewanto; "Developing, Deploying, Using and Evaluating an Open Source Learning Management System". *26th Intl. Conference on Information Technologies Interfaces*, Cavtat/Dubrovnik, 7-10 de junio de 2004.
- [HAM01]. M. Hammar Cloyd. "Designing User-Centered Web Applications in Web Time". *IEEE Software* jan-feb, 2001.
- [Hix93]. D. Hix, H.R. Hartson; *Developing User Interfaces: Ensuring Usability Through Product and Process*. John Wiley & Sons, New York, 1993.
- [HUA99]. K. Huang; Y.W. Lee, R.Y. Wang; *Quality information and Knowledge*. NJ: Prentice Hall, 1999.
- [ISO9000]. ISO 9000 Norma Internacional. *Sistemas de Gestión de la Calidad- Conceptos y Vocabulario*. Traducción certificada, 2000.
- [ISO90003]. ISO/IEC 90003. *Software Engineering- Guidelines for the application of ISO 9001:2000 to Computer Software*, 2004.
- [ISO13407]. ISO DIS 13407: *User Center Design process for interactive systems*, 1997.
- [ISO14598-1]. ISO/IEC 14598-1. *Information Technology-Software Product Evaluation-Part 1: General overview*, 1999.
- [ISO14598-5]. ISO/IEC 14598-5. "Information technology -- Software product evaluation - Part 5: Process for evaluators", 1998.
- [ISO15939]. ISO/IEC 15939. "Software Engineering - Software Measurement Process" (2002).
- [ISO8402]. ISO DIS 8402. *Quality Vocabulary* (1994).
- [ISO9126]. ISO/IEC 9126. "Information technology - Software product evaluation - Quality characteristics and guidelines for their use (1991).

- [ISO9126-1]. ISO/IEC 9126-1:2001. Software Engineering— Software Product Quality—Part 1: Quality Model, Int'l Org. For Standardization, Geneva, 2001.
- [ISO9126-4]. ISO/IEC DTR 9126-4. Software Engineering –Product Quality - Part 4: Quality In Use Metrics.
- [ISO9241-11]. ISO 9241-11. “Ergonomic requirements for office work with visual display terminals (VDT)s - Part 11 Guidance on Usability” (1998).
- [IVO01]. Melody Y. Ivory; “*An Empirical Foundation for Automated Web Interface*”, capítulo 5 “Web Interface Measures”. Tesis doctoral en Ciencias de la Computación de la UNIVERSITY of CALIFORNIA at BERKELEY. 2001.
- [JUR01]. N. Juristo, H. Windl, L. Constantine; “*Introducing Usability*”. IEEE Software jan-feb, 2001.
- [KIR98]. J. Kirakowski, N. Claridge, R. Withehand; “*Human Centered Measure of Success in Web Site Design*”. 4th Conference on Human Factors & the Web, New Jersey, EEUU, junio de 1988.
- [KIT01]. B.A. Kitchenham, R.T. Hughes, S.G. Linkman; “*Modeling Software Measurement Data*”. IEEE Transactions on SoftwareEngineering, 27(9), 2001, pp. 788-804.
- [MAR03]. M. A. Martín; L. Olsina; “*Towards an Ontology for Software Metrics and Indicators as the Foundation for a Cataloging Web System*”, *proceed. of IEEE Computer Society (1st Latin American Web Congress)*, Santiago de Chile, pp 103-113, ISBN 0-7695-2058-8 (2003).
- [MAR04]. María de los Ángeles Martín; “*Sistema de Catalogación de Métricas e Indicadores con Potencia de Web Semántica*”. Tesis de Magister en Ingeniería de Software presentada en la Facultad de Informática de la UNLP. La Plata, setiembre de 2004.
- [MAT02]. M. Matera, M. F. Costabile, F. Garzotto, P. Paolini; “*SUE Inspection: An Effective Method for Systematic Usability Evaluation of Hypermedia*”. IEEE Transactions on System, Man and Cybernetics- Part A: Systems and Humans, Vol. 32, January 2002, pags. 93-103.
- [MCC77]. J. A. McCall, P.K. Richards, G.F. Walters; “*Factors in software quality, Vol I, II, III: Final Technical Report, RADC-TR-77-369*”. Rome Air Development Center, Air Force System Command, Griffith Air Force Base. NY 1977.
- [MIL03]. Veljko Milutinovic, Nikola Skundric; “*Will Distance Learning Create a Global University?*”. IEEE Computer, Vol 36, nº 5. Mayo de 2003.
- [MOL05]. Hernán Molina; “*Soporte al proceso de Medición para el Aseguramiento de Calidad en Proyectos de Software y Web*”. Presentación de Tesis de Licenciatura en la Facultad de Ciencias Exactas de la UNLu. Luján, mayo de 2005.
- [MUR01]. S. Murugesan; Y. Deshpande; S. Hansen; A. Ginige; “*Web Engineering: A New Discipline for Development of Web-based Systems*”. *LNCS 2016 of Springer, Web Engineering: Managing University and Complexity of Web Application Development.*, San Murugesan, Yogesh Deshpande Eds., pp. 3-13 (2001).
- [NIE93]. J. Nielsen; *Usability engineering*. Academic press, San Diego, CA, 1993.

- [NIE00]. J. Nielsen; *Designing Web Usability*. New Riders- Indianapolis, Ind. 2000. ISBN:156205810X.
- [NIE01]. J. Nielsen, R. Molich, C. Snyder, S. Farrell; *E-Commerce User Experience*. NN Group (2001).
- [NIE04]. J. Nielsen; *The Alertbox column*, <http://www.useit.com/alertbox/> (1995-2005).
- [OLS99]. L. Olsina; D. Godoy; G.J. Lafuente; G. Rossi; "Assessing the Quality of Academic Websites: a Case Study". *New Review of Hypermedia and Multimedia (NRHM) Journal*, Taylor Graham Publishers. UK/USA Vol. 5, pp. 81-103 (1999).
- [OLS00a]. L. Olsina; G.J. Lafuente; G. Rossi; "E-commerce Site Evaluation: a Case Study". LNCS 1875 of Springer, 1st International Conference on Electronic Commerce and Web Technologies, EC-Web 2000. London, UK, pp. 239-252 (2000).
- [OLS00b]. L. Olsina; *Metodología Cuantitativa para la Evaluación y Comparación de la Calidad de Sitios Web*. Presentación de Tesis Doctoral en la Facultad de Ciencias Exactas de la UNLP. La Plata, abril de 2000.
- [OLS01]. L. Olsina, M.F. Papa, M.E. Souto, G. Rossi; "Providing Automated Support for the Web Quality Evaluation Methodology". 4th Workshop on Web Engineering, at the 10th International WWW Conference, Hong Kong, pp. 1-11 (2001)
- [OLS02]. Olsina L.; Rossi G.; "Measuring Web Application Quality with WebQEM". *IEEE Multimedia*, 9(4), pp. 20-29 (2002).
- [OLS04a]. L. Olsina, M. A. Martín; "Ontology for Software Metrics and Indicators". *Journal of Web Engineering*, Rinton Press, US, Vol 2 N° 4, pp. 262-281, ISSN 1540-9589 (2004).
- [OLS04b]. Luis Olsina, Guillermo Covella, Gustavo Rossi; "Web Quality". Capítulo del libro "Web Engineering: Theory and Practice of Metrics and Measurement for Web Development". Emilia Mendes y Nile Mosley Editores. A publicarse por Springer Verlag en 2005. ISBN: 3-540-28196-7.
- [OLS05a]. Luis Olsina, Fernanda Papa, Hernán Molina; "Ontological Support for a Measurement and Evaluation Framework", a ser publicado en una edición especial del Int'l Journal of Intelligent Systems (Spring 2006).
- [OLS05b]. L. Olsina, F. Papa, H. Molina; "Organization-Oriented Measurement and Evaluation Framework for Software and Web Engineering Projects". 5th International Conference, ICWE 2005, Sydney, Australia, July 27-29, 2005, Proceedings Series: Lecture Notes in Computer Science, Vol. 3579, David Lowe, Martin Gaedke (Eds.) 2005, XXII, p. 633, ISBN: 3-540-27996-2.
- [POW98]. T.A. Powell; *Web Site Engineering: Beyond Web Page Design*. Prentice Hall (1998).
- [PSA04]. Y. Psaromiligkos, S. Retalis, K. Siassiakos; "DEPTH: A Methodology for Usability Evaluation of Web Sites Based on Design Patterns and Heuristics Criteria". Fourth Int'l Workshop on Web-Oriented Software Technologies (IWWOST'04) in conjunction with ICWE'04, pp. 44-53 (2004).

- [RET02]. Symos Retalis, Paris Avgeriou; *“Modelling Web-Based Instructional Systems”*. Journal of Information Technology Education, Vol. 1, Nº 1, páginas: 26-41, Glenn Lowry Editor, 2002. En línea en <http://jite.org/documents/Vol1/v1n1p025-042.pdf> .
- [SHA91]. B. Shackel; *Usability –context, framework, design and evaluation. Human Factors for Informatics Usability*. Cambridge University Press, Cambridge, pages 21-38. B. Shackel, S.Richardson eds., 1991.
- [SUS96] John Brooke; *“SUS –A ‘quick and dirty’ usability scale”*. Digital Equipment Co Ltd., Reading, United Kingdom.
En línea en <http://www.usability.serco.com/trump/documents/Suschapt.doc> .
- [TRI04]. L. Triacca, D. Bolchini, L. Botturi, A. Inversini; *“MiLE : Systematic Usability Evaluation for E-Learning Web Applications”*. World Conference on Educational Multimedia, Hypermedia and Telecommunications, Vol. 2004, Issue.1, 2004, pags. 4398-4405.
- [YIP05]. Michael Yip, Emilia Mendes; *“Web Usability Measurement: Comparing Logic Scoring Preference to Subjective Assessment”*. 5th International Conference on Web Engineering, ICWE 2005, Sydney, Australia, July 27-29, 2005, Proceedings Series: Lecture Notes in Computer Science, Vol. 3579 David Lowe, Martin Gaedke (Eds.). 2005, XXII, 633 p., ISBN: 3-540-27996-2, páginas 53-62.
- [ZUS98]. H. Zuse; *A Framework of Software Measurement*. Walter de Gruyter, Berlín-NY, 1998.

Apéndice A: Métricas del Caso de Estudio

Se presentan a continuación las métricas diseñadas para el caso de estudio, de manera detallada y siguiendo un esquema jerárquico top-down. Por razones de espacio, se han presentado en el capítulo 5 solamente las métricas vinculadas directamente a los atributos del árbol de requerimientos, de modo que este apéndice es complementario de esa información.

Para cada concepto calculable, se presentan en un primer nivel los atributos del árbol de requerimientos y la métrica relacionada, luego las métricas indirectas necesarias para el método de cálculo de la métrica presentada en primer término, precedidas por el atributo correspondiente. Por último, siempre en el contexto del concepto calculable de más alto nivel, se detallan las métricas directas. Este esquema se repite tanto para *Eficacia* (A.1) como para *Productividad* (A.2) y *Satisfacción* (A.3).

A.1. Concepto Calculable (Característica): Eficacia

A.1.1. Atributo: *Eficacia de Tarea*

Definición: correctitud en la realización total o parcial de las tareas, para todos los usuarios.

Objetivo: permite conocer en que medida el producto puede ser utilizado eficazmente por los usuarios en la realización de tareas típicas. Se considera que una tarea completada parcialmente otorga igualmente un cierto grado de eficacia, en la medida en que las metas parciales se cumplen correctamente.

A.1.1.1. Métrica(Nombre) : Promedio de la Proporción de Tareas Completadas por Todos los Usuarios, P_{ETT_u} .

Interpretación: $0 \leq P_{ETT_u} \leq 1$, cuanto más próximo a 1 mejor

Objetivo: Calcular el promedio general de la proporción de los objetivos completados correctamente, o sea de todos los usuarios en relación a todas las tareas propuestas.

Método de Cálculo(Fórmula):
$$P_{ETT_u} = \frac{\sum_{j=1}^{j=n} P_{ET1u}(j)}{n},$$

n :número de usuarios que intervienen en el estudio, P_{ET1u} es la métrica A.1.1.2.1.1.

Escala: Numérica.

Tipo de Escala: Proporción.

Unidad (tipo, descripción): tarea, promedio de la proporción de tareas realizadas correctamente por todos los usuarios.

A.1.1.2. Métricas Indirectas Relacionadas

A.1.1.2.1. Atributo: Proporción de Tareas Completadas por *un* Usuario.

A.1.1.2.1.1. Métrica (Nombre): Promedio de la Proporción de Tareas Completadas por *un* Usuario, P_ET1u.

Interpretación: $0 \leq P_ET1u \leq 1$, cuanto más próximo a 1 mejor.

Objetivo: Calcular, para un usuario, en relación a todas las tareas, el promedio de la proporción de los objetivos que han sido completados correctamente.

Método de Cálculo(Fórmula): $P_ET1u = \frac{\sum_{j=1}^{j=m} ET(j)}{m}$, m : nº de tareas propuestas, ET se describe en A.1.1.2.2.1.

Tipo de Escala: Numérica

Escala: Proporción

Unidad (tipo, descripción): tarea, promedio de la proporción de tareas realizada correctamente por un usuario.

A.1.1.2.2. Atributo: Submetas Alcanzadas Para una Tarea.

A.1.1.2.2.1. Métrica (Nombre): Proporción de Tarea Completada Correctamente, ET.

Interpretación: $0 \leq ET \leq 1$, cuanto más próximo a 1 mejor.

Objetivo: Conocer, para un usuario y una tarea, qué proporción de los objetivos de la tarea son completados correctamente.

Referencias: sección 8.1 “*Effectiveness Metrics*”, Anexo E.2.1.2 de [ISO9126-4]²⁴ y Apéndice C (asignación de pesos a las submetas de las tareas).

Método de Cálculo(Fórmula): $ET = 1 - \left| \sum A_i \right|$,

donde A_i es el peso asignado a la parte proporcional de la tarea *no* realizada o realizada incorrectamente. En caso de que la sumatoria de los pesos de A_i exceda 1, a la métrica se le asignará el valor 0.

Escala: Numérica.

Tipo de Escala: Proporción.

²⁴ “Las metas pueden ser descompuestas en submetas que especifiquen componentes de una meta general y los criterios que deberán satisfacerse para alcanzar las submetas”

Unidad (tipo, descripción): tarea, proporción de cada tarea realizada correctamente.

A.1.2. Atributo: Completitud de Tareas.

Definición: proporción de las tareas propuestas completadas por los usuarios.

Objetivo: permite conocer en que medida las tareas propuestas han sido completadas por los usuarios. A diferencia del atributo *Eficacia de Tarea*, aquí se toman en cuenta sólo las tareas que han sido completadas correctamente en su totalidad.

A.1.2.1.Métrica(Nombre): Promedio de la proporción de tareas completadas sobre tareas propuestas para todos los usuarios (n usuario \leftrightarrow m tareas).

Interpretación: $0 \leq P_{CT} \leq 1$, cuanto más próximo a 1 mejor,

Objetivo: calcular, para todos los usuarios de un perfil, el promedio de la proporción de tareas completadas en relación a las tareas propuestas.

Método de Cálculo(Fórmula): $P_{CT} = \frac{\sum_{j=1}^{j=n} CT(j)}{n}$,

n : cantidad de usuarios, CT:Completitud de tareas, descrita en A.1.2.2.1.1.

Escala: numérica

Tipo de Escala: proporción

Unidad (tipo, descripción):Tarea, proporción de tareas completadas sobre tareas propuestas.

A.1.2.2. Métricas Indirectas Relacionadas

A.1.2.2.1. Atributo: Tareas Completadas en relación a Tareas Propuestas

A.1.2.2.1.1. Nombre de la Métrica: Proporción de Tareas Completadas sobre Tareas Propuestas, CT.

Interpretación: $0 \leq CT \leq 1$, Cuanto más próximo a 1 mejor,

Objetivo: calcular la proporción de tareas completadas en relación a las totalidad de tareas propuestas para un usuario.

Método de Cálculo(Fórmula): $CT = \frac{T_c}{T_p}$, T_c : N° de tareas completadas (métrica A.1.2.3.2.1), T_p : N° de tareas propuestas (métrica A.1.2.3.1.1).

Referencias: ver sección 8.1 "Effectiveness Metrics" de [ISO9126-4].

Escala: Numérica.

Tipo de Escala: Proporción.

Unidad (tipo, descripción): Tarea, proporción de tareas completadas sobre tareas propuestas.

A.1.2.3. Métricas Directas Relacionadas

A.1.2.3.1. Atributo: Tareas Propuestas, Tp.

Definición: número de tareas propuestas a los usuarios.

Objetivo: proveer un atributo adicional que permita establecer el valor de la métrica para el atributo Completitud de Tareas (A.1.2).

A.1.2.3.1.1. Métrica (Nombre): N° de tareas propuestas, Tp.

Interpretación: $0 < Tp$, la cantidad de tareas propuestas debe ser mayor que 0.

Objetivo: Conocer el número de tareas propuestas para que realicen los usuarios.

Referencias: *Apéndice C.*

Tipo de método: Objetivo, manual.

Método de Medición: Conteo del n° de tareas propuestas.

Escala: Numérica.

Tipo de Escala: Intervalo.

Unidad(tipo, descripción): Tarea, número de tareas propuestas a los usuarios.

A.1.2.3.2. Atributo: Tareas Completadas, Tc.

Definición: número de tareas completadas totalmente por un usuario.

Objetivo: proveer un atributo adicional que permita establecer luego los valores de la métrica para el atributo Completitud de Tareas (A.6.2.1.1.2).

A.1.2.3.2.1. Nombre de la Métrica: número de tareas completadas correctamente.

Interpretación: $0 \leq Tc$.

Objetivo: Conocer el n° de tareas completadas correctamente por un usuario.

Referencias: *apéndice C.*

Tipo de método: objetivo, manual.

Método de Medición: conteo de las tareas completadas correctamente.

Escala: numérica.

Tipo de Escala: proporción.

Unidad(tipo, descripción): tareas.

A.2. Concepto Calculable (Característica): **Productividad**

A.2.1. **Atributo** *Eficiencia en relación a Eficacia.*

Definición: productividad en las tareas completadas parcial o totalmente.

Objetivo: conocer el nivel de productividad obtenido por los usuarios, considerando aquellas tareas que fueron parcial o totalmente completadas.

A.2.1.1. Métrica(Nombre): Promedio de eficiencia de tareas para todos los usuarios considerando Eficacia.

Interpretación: $0 \leq P_EFtTu$, cuanto más grande, mejor.

Objetivo Calcular el promedio de eficiencia general, para todos los usuarios, en la realización de las tareas propuestas.

Método de Cálculo(Fórmula):
$$P_EFtTu = \frac{\sum_{j=1}^{j=n} P_EFt1u(j)}{n}$$

n : número de usuarios intervinientes en el test, P_EFt1u se detalla en la sección A.2.1.2.1.1.

Escala: numérica.

Tipo de Escala: proporción.

Unidad: 1/Minutos.

A.2.1.2. Métricas Indirectas Relacionadas

A.2.1.2.1. Atributo: Eficiencia de un Usuario en relación a Eficacia

A.2.1.2.1.1. Métrica(Nombre): Promedio de Eficiencia de Tareas para un Usuario considerando Eficacia, P_EFt1u .

Interpretación: $0 \leq P_EFt1u$, cuanto más grande, mejor.

Objetivo: Calcular el promedio de eficiencia de un usuario en la realización de las tareas propuestas.

Método de Cálculo(Fórmula):
$$P_EFt1u = \frac{\sum_{j=1}^{j=T1} Eft(j)}{Tp}$$

El numerador, a partir de la métrica Eft (sección A.2.1.2.2.1), representa la sumatoria de la proporción de Eficiencia de Tarea considerando Eficacia para cada tarea, y la unidad es tarea/minutos. Tp , el denominador, es el número de tareas propuestas (métrica A.1.2.3.1.1).

Escala: numérica.

Tipo de Escala: Proporción.

Unidad: 1/minutos.

A.2.1.2.2. Atributo: Eficiencia de Tarea en Relación a Eficacia

A.2.1.2.2.1. Métrica (Nombre): Proporción de Eficiencia de Tarea considerando Eficacia, Eft.

Interpretación: $0 \leq Eft$, cuanto más grande, mejor.

Objetivo: Calcular cuan eficiente es un usuario en la realización de una tarea. Expresándose el resultado como la proporción de una tarea completada correctamente en cierta unidad de tiempo (típicamente minutos).

Método de Cálculo(Fórmula): $Eft = \frac{ET}{Tt}$

ET: Proporción de Tarea Completada Correctamente (métrica A.1.1.2.2.1),

Tt: Tiempo de tarea.

Escala: Numérica.

Tipo de Escala: Proporción.

Unidad(tipo, descripción): Tarea Completada Correctamente/minutos.

A.2.1.3. Métricas Directas Relacionadas

A.2.1.3.1. Atributo: Tiempo Empleado por un Usuario para llevar a cabo una Tarea.

Definición: tiempo que emplea un usuario en completar una tarea.

Objetivo: medir el tiempo utilizado por un usuario para realizar una tarea.

A.2.1.3.1.1. Nombre de la Métrica: Tiempo Empleado por un Usuario para Completar una Tarea, Tt.

Interpretación: $0 < Tt$, cuanto más pequeño mejor.

Objetivo: Conocer el tiempo que le insume a un usuario realizar una tarea.

Tipo de método: Objetivo/semiautomático.

Método de Medición: computar, para cada usuario que participó del estudio, el tiempo que le insumió realizar una tarea. El cómputo se realiza desde el momento en que el usuario inicia la tarea hasta que indica que la finalizó.

Escala: Numérica.

Tipo de Escala: Proporción.

Unidad: Tiempo (Minutos).

A.2.2. Atributo: Eficiencia en relación a Completitud de Tareas.

Definición: productividad de los usuarios en las tareas completadas totalmente.

Objetivo: conocer el nivel de productividad obtenido por los usuarios, considerando sólo aquellas tareas que fueron totalmente completadas.

A.2.2.1 Nombre de la Métrica: Promedio de la Proporción de Eficiencia de Tareas Completadas.

Interpretación: $0 < P_EFtRcTu$, cuanto mayor, mejor.

Objetivo: Calcular el promedio de eficiencia general, para todos los usuarios intervinientes, en la realización de las tareas propuestas.

Método de Cálculo(Fórmula):
$$P_EFtRcTu = \frac{\sum_{j=1}^{j=n} EFtRc(j)}{n}$$

n : número de usuarios que intervienen en el test, $EFtRc$ es la métrica A.2.2.2.1.1.

Escala: numérica.

Tipo de Escala: intervalo.

Unidad: tareas/minutos.

A.2.2.2. Métricas Indirectas Relacionadas

A.2.2.2.1. Atributo: Eficiencia en las Tareas Completadas.

A.2.2.2.1.1. Nombre de la Métrica: Proporción de Eficiencia de Tareas Completadas, $EFtRc$.

Interpretación: $0 < EFtRc$, cuanto mayor, mejor. A medida que el tiempo empleado disminuye (el denominador se hace más pequeño), aumenta el nivel de eficiencia. Además, dado que las tareas consideradas son aquellas completadas totalmente, el numerador (la métrica CT, A.1.2.2.1.1) es siempre mayor que 0.

Objetivo: Calcular la eficiencia de un usuario que al menos ha completado correctamente una tarea.

Método de Cálculo(Fórmula):
$$EFtRc = \frac{CT}{TtCc}$$

CT: es Completitud de Tarea y TtCc: Tiempo de Tareas Completadas Correctamente.

Escala: numérica.

Tipo de Escala: intervalo.

Unidad: tarea/minutos.

A.2.2.2.2. Atributo: Tiempo Total de Tareas Completadas.

A.2.2.2.2.1. Métrica(Nombre): Tiempo Total de Tareas Completadas, TtCc.

Interpretación: $0 < TtCc$, cuanto menor, mejor. A menor tiempo insumido, mayor productividad en la realización de *todas* las tareas completadas.

Objetivo: Calcular el tiempo total que insumen todas las tareas que completó correctamente un usuario.

Método de Cálculo(Fórmula): $TtCc = \sum_{i=1}^{Tc} TtC(i)$

TtC: tiempo de tarea completada.

Escala: numérica.

Tipo de Escala: intervalo.

Unidad: tiempo (minutos o fracción para este caso).

A.2.2.3. Métricas Directas Relacionadas

A.2.2.3.1. Atributo: Tiempo Necesario para Completar una Tarea.

A.2.2.3.1.1. Métrica(Nombre): Tiempo de Tarea Completada, TtC.

Interpretación: $0 < TtC$, cuanto menor, mejor. A menor tiempo insumido para completar una tarea, mayor productividad en la realización de la tarea.

Objetivo: Conocer el tiempo que insumió a un usuario completar correctamente una tarea.

Tipo de método: Objetivo/Semiautomático.

Método de Medición: computar, para cada usuario que participó del caso de estudio, el tiempo que le insumió completar correctamente una tarea.

Escala: numérica

Tipo de Escala: intervalo.

Unidad: tiempo.

A.3. Concepto Calculable (Característica): *Satisfacción*

A.3.1. Atributo: *Satisfacción*.

Definición: es el nivel de satisfacción expresado por los usuarios en relación al producto evaluado.

Objetivo: permite conocer en que medida los usuarios se sintieron satisfechos empleando el producto. Si bien representa el aspecto subjetivo de la evaluación de calidad en uso, es posible cuantificarlo. La evaluación subjetiva toma en cuenta aspectos del producto tales como: facilidad para ser aprendido y utilizado, adecuación del diseño y presentación de la interfaz, etc.

A.3.1.1. Nombre de la Métrica: Promedio del grado de Satisfacción, P_GS.

Interpretación $0 \leq P_{GS} \leq 40$, cuanto más próximo a 40, mejor.

Objetivo Calcular el promedio del grado de satisfacción para todos los usuarios.

Método de Cálculo(Fórmula):
$$P_{GS} = \frac{\left(\sum_{j=1}^{j=n} GS(j) \right)}{n}$$

n: número de usuarios que intervienen en el test.

Escala: numérica.

Tipo de Escala: proporción.

Unidad: grado de satisfacción.

A.3.1.2. Métricas Indirectas Relacionadas

A.3.1.2.1. Atributo: Satisfacción para un Usuario.

A.3.1.2.1.1. Nombre de la Métrica: Satisfacción para un Usuario, GS.

Interpretación: $0 \leq GS \leq 40$, cuanto más próximo a 40, mejor. El 40 como valor máximo del cuestionario no es arbitrario sino que responde a características del diseño de la encuesta.

Objetivo: Conocer el grado de satisfacción de un usuario, usando el producto evaluado.

Referencias: [SUS96], Apéndice D.

Método de Cálculo(Fórmula):
$$GS = \left(\sum_{k=1}^{k=\frac{m}{2}} \text{item}_{2k} + \sum_{k=1}^{k=\frac{m}{2}} \text{item}_{2k-1} \right)$$

item_k : k-esimo ítem del cuestionario, m: cantidad de ítems del cuestionario.

Cada ítem puede ser respondido seleccionando una de 5 (cinco) posibles respuestas, oscilando entre “*Totalmente en Desacuerdo*” y “*Totalmente de Acuerdo*” de una escala tipo Likert.

El puntaje para los ítems impares se calcula como la posición de la respuesta menos 1 (uno) y el puntaje para los ítems pares como 5 (cinco) menos la posición de la respuesta. De este modo, el puntaje máximo que puede obtenerse para cada ítem es 4 y el mínimo 0, y para cada cuestionario el máximo puntaje cuarenta (40).

Escala: numérica.

Tipo de Escala: proporción.

Unidad: grado de satisfacción.

Apéndice B: Ontología de Métricas e Indicadores

B.1. Introducción

Este apéndice presenta la definición exhaustiva de los conceptos de la Ontología de Métricas e Indicadores. Los conceptos más importantes se mencionaron en el capítulo 3. La fuente original de esta información es el trabajo “*Ontological Support for a Measurement and Evaluation Framework*” [OLS05a]. Luego, esa versión publicada fue revisada por los autores y las modificaciones y novedades están incluidas en esta apéndice.

En la sección B.2, la tabla B.1 contiene la definición de los conceptos incluidos en la Ontología de Métricas e Indicadores. En la sección B.3. están los modelos, con notación UML, que representan gráficamente los conceptos y sus relaciones en distintas instancias del marco de medición y evaluación. Finalmente, en B.4 están las tablas con la definición de los atributos (tabla B.2) y las relaciones (tabla B.3) detalladas en los modelos.

B.2 Definición de Conceptos

Tabla B.1. Definición de Conceptos

| Concept | Synonyms | Definition |
|--------------------------|------------------------------|--|
| InformationNeed | | Insight necessary to manage objectives, goals, risks, and problems. |
| EntityCategory | | Object category that is to be characterized by measuring its attributes. |
| Entity | Object | A concrete object that belongs to an entity category. |
| CalculableConcept | MeasurableConcept | Abstract relationship between attributes of entities categories and information needs. |
| ConceptModel | | The set of sub-concepts and the relationships between them, which provide the basis for specifying the concept requirement and its further evaluation or estimation. |
| Attribute | Property, Feature | A measurable physical or abstract property of an entity category. |
| Metric | | The defined measurement or calculation method and the measurement scale. |
| DirectMetric | Single, Base Metric | A metric of an attribute that does not depend upon a metric of any other attribute. |
| IndirectMetric | Hybrid, Derived Metric | A metric of an attribute that is derived from metrics of one or more other attributes. |
| Function | Formula, Algorithm, Equation | Algorithm or formula performed to combine two or more metrics. |
| Scale | | A set of values with defined properties. |

| Concept | Synonyms | Definition |
|----------------------------|-----------------------------------|--|
| CategoricalScale | | A scale where the measured or calculated values are categories, and cannot be expressed in units, in a strict sense. |
| NumericalScale | | A scale where the measured or calculated values are numbers that can be expressed in units, in a strict sense. |
| Unit | | Particular quantity defined and adopted by convention, with which other quantities of the same kind are compared in order to express their magnitude relative to that quantity. |
| Method | Procedure | Logical sequence of operations and possible heuristics, specified generically, for allowing the realisation of an activity description. |
| SoftwareTool | Software Instrument | A tool that automates partially or totally a measurement or calculation method. |
| MeasurementMethod | Counting Rule, Protocol | The particular logical sequence of operations and possible heuristics specified for allowing the realisation of a metric description by a measurement. |
| CalculationMethod | | The particular logical sequence of operations specified for allowing the realisation of a formula or indicator description by a calculation. |
| Measurement | | Activity that uses a metric definition in order to produce a measure's value. |
| Measure | | The number or category assigned to an attribute of an entity by making a measurement. |
| Indicator | Criteria | The defined calculation method and scale in addition to the model and decision criteria in order to provide an estimate or evaluation of a calculable concept with respect to defined information needs. |
| ElementaryIndicator | Elementary Preference or Criteria | An indicator that does not depend upon other indicators to evaluate or estimate a calculable concept. |
| ElementaryModel | | Algorithm or function with associated decision criteria that model an elementary indicator. |
| DecisionCriteria | | Thresholds, targets, or patterns used to determine the need for action or further investigation, or to describe the level of confidence in a given result. |
| Range | | Threshold or limit values that determine the acceptability levels. |
| GlobalIndicator | Global Preference or Criteria | An indicator that is derived from other indicators to evaluate or estimate a calculable concept. |
| GlobalModel | Aggregation Model or Function | Algorithm or function with associated decision criteria that model a global indicator. |
| Calculation | Computation | Activity that uses an indicator definition in order to produce an indicator's value. |
| IndicatorValue | | The number or category assigned to a calculable concept by making a calculation. |

B.3. Modelos UML para los conceptos de la Ontología de Métricas e Indicadores

Los conceptos de la Ontología de Métricas e Indicadores están organizados en tres módulos, presentados en las Figuras 1, 2 y 3 de acuerdo a las fases de la metodología WebQEM, como una forma de desagregar y gestionar la complejidad del modelo completo.

Módulo de Requerimientos

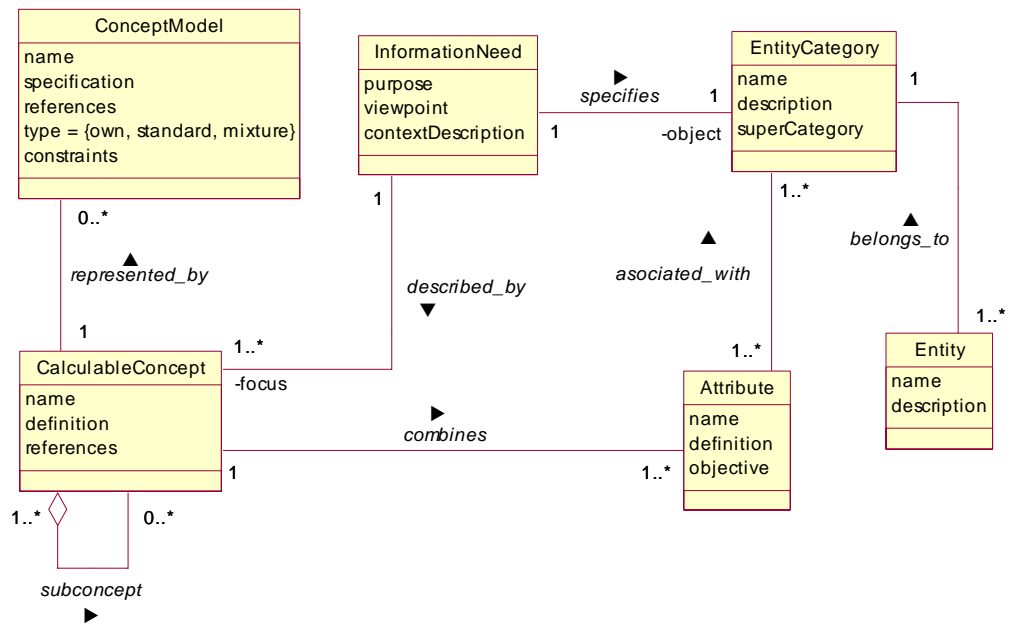


Figura B.1. Conceptos involucrados en la fase de Definición y Especificación de Requerimientos.

Módulo de Medición

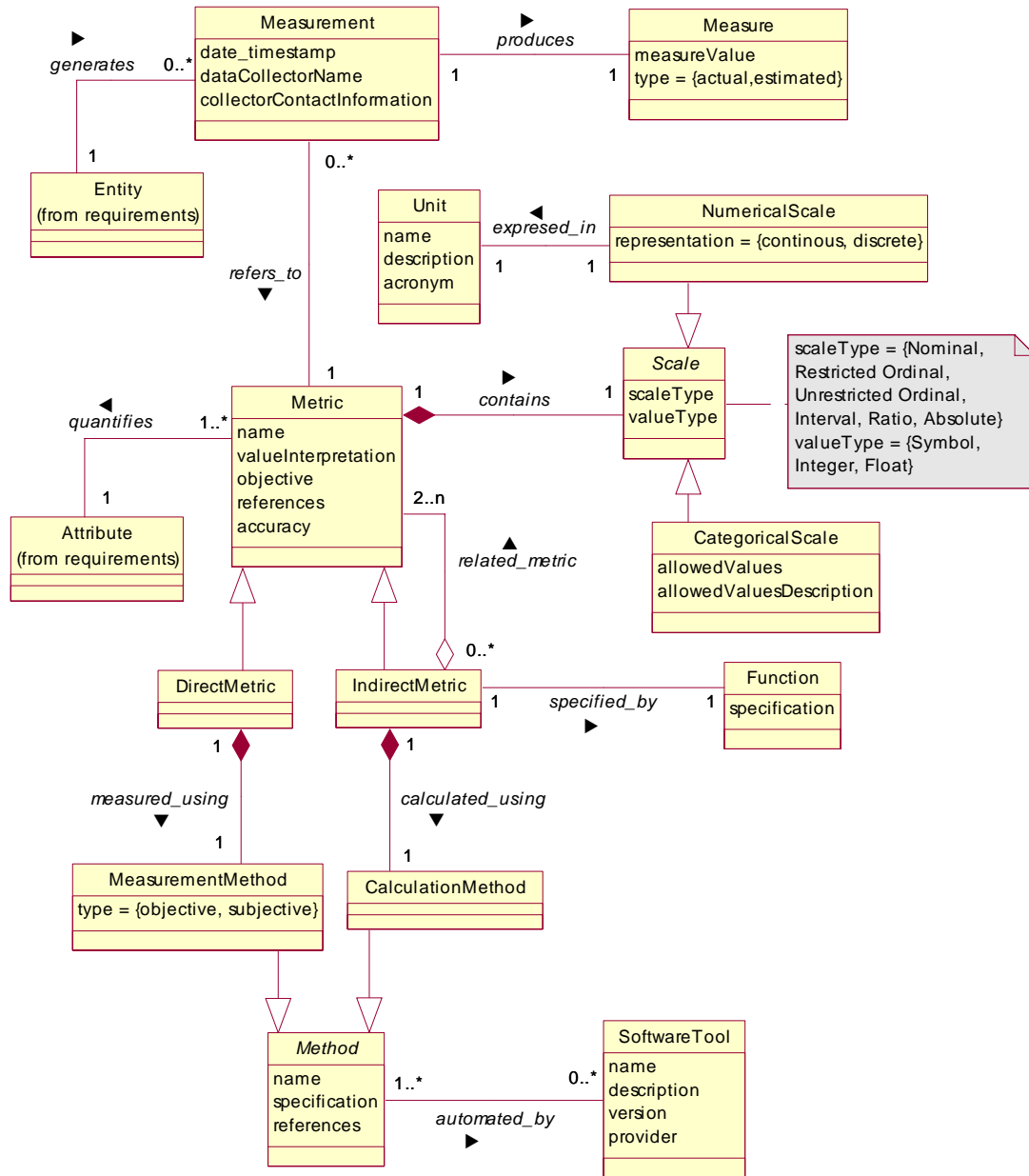


Figura B.2. Conceptos involucrados en la fase de Diseño de la Medición e Implementación.

Módulo de Evaluación

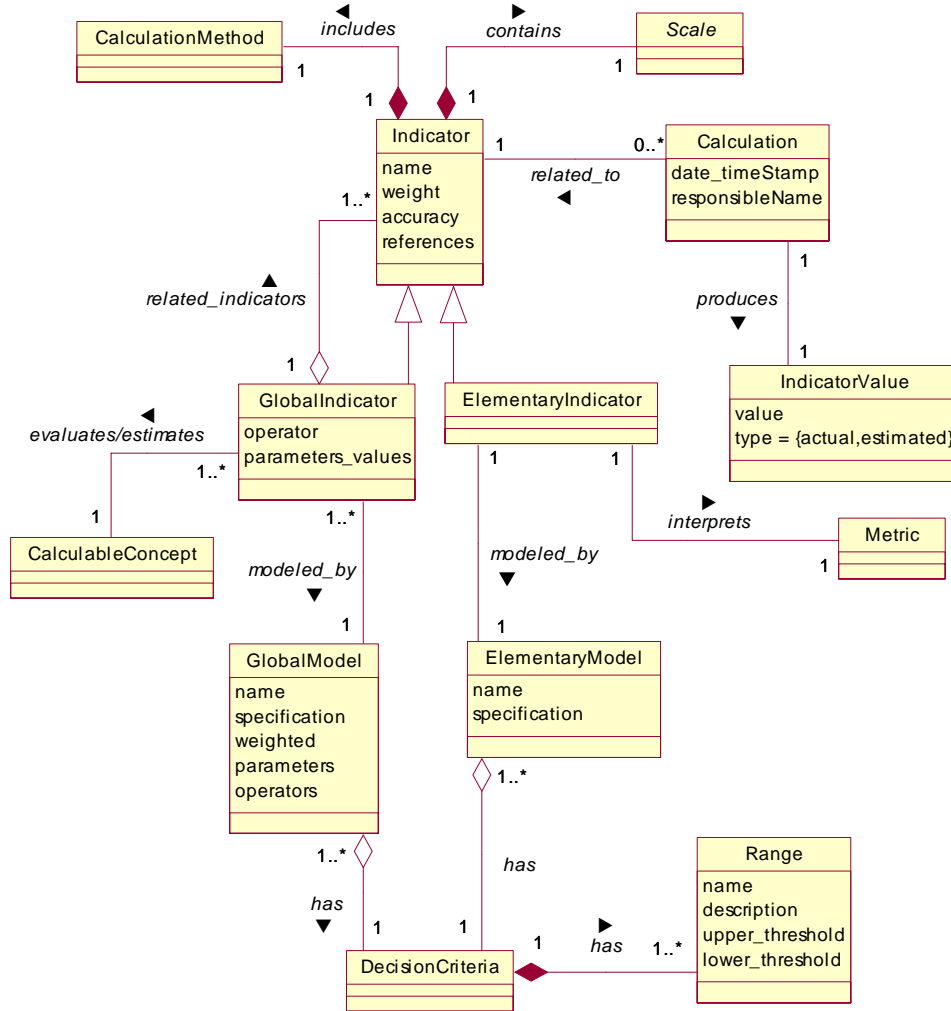


Figura B.3. *Conceptos involucrados en la fase de Diseño de la Evaluación e Implementación.*

B.4. Descripción de Atributos y Relaciones

Tabla B.2. Descripción de Atributos de la Ontología de Métricas e Indicadores.

| Concept | Attribute | Description |
|-----------------------------|--------------------------|--|
| Attribute | Name | Name of an attribute to be identified. |
| | Definition | An unambiguous description of the attribute meaning |
| | Objective | Goal or purpose to measuring this attribute |
| Calculable Concept | Name | Name of a calculable concept to be identified. |
| | Definition | An unambiguous description of the calculable concept meaning. |
| | References | References to bibliographical or URL resources, where additional and authoritative information of a given calculable concept can be consulted. |
| Calculation | date_timeStamp | Instant when a calculation is performed. |
| | responsibleName | |
| Categorical Scale | allowedValues | List of literals indicating the valid values of a categorical scale. |
| | allowedValuesDescription | An unambiguous description of the allowed values meaning. |
| Concept Model | Name | Name of a concept model to be identified. |
| | Specification | A formal or semiformal representation of a concept model. |
| | References | References to bibliographical or URL resources, where additional and authoritative information of a given concept model can be consulted. |
| | Type | Indicates the type of concept model, it can be own, standard or mixture. |
| Decision Criteria | | |
| Elementary Indicator | | |
| Elementary Model | Name | Name of an elementary model to be identified. |
| | Specification | A formal or semiformal representation of an elementary model. It can be e.g. a mathematical or logical representation |
| Entity | Name | Name of an entity to be identified. |
| | Description | An unambiguous description of the entity meaning |
| Entity Category | Name | Name of an entity category to be identified. |
| | Description | An unambiguous description of the entity category meaning |
| | superCategory | |
| Function | Specification | A formal or semiformal representation of a function. Synonymous: formula. |
| Global Indicator | operator | |
| | Parameters_values | |
| Global Model | Name | Name of a global model to be identified. |
| | Specification | A formal or semiformal representation of a global model. It can be e.g. a mathematical or logical representation |
| | weighted parameters | |
| | operators | |
| Indicator | Name | Name of an indicator to be identified. |
| | weight | |
| | Accuracy | A quantification of the accuracy level inherent to the way of producing the value of an indicator. |
| | References | References to bibliographical or URL resources, where additional and authoritative information of the given indicator can be consulted. |
| Indicator Value | value | Numerical or categorical result assigned to an indicator [10]. Synonymous: data. |
| | type | Indicates the type of indicator value, it can be actual or estimated. |
| Information Need | purpose | |
| | viewpoint | |
| | contextDescription | |

| Concept | Attribute | Description |
|---------------------------|-----------------------------|---|
| Measure | measureValue | Numerical or categorical result assigned to an attribute. Synonymous: data. |
| | type | Indicates the type of measure value, it can be actual or estimated. |
| Measurement | date_timeStamp | Instant when a measurement is performed. |
| | dataCollectorName | |
| | collectorContactInformation | |
| Measurement Method | Type | Indicates the type of measurement method that depends on the nature of the operations used to quantify an attribute. Two types may be distinguished: subjective (quantification involving human judgement), and objective (quantification based on numerical rules) [10]. |
| Method | Name | Name of a method to be identified. |
| | Specification | A formal or semiformal description of a method. |
| | References | References to bibliographical or URL resources, where additional and authoritative information of the given method can be consulted. |
| Metric | Name | Name of a metric to be identified. |
| | valueInterpretation | An unambiguous textual statement for helping stakeholders to understand the obtained value meaning, e.g. the closer to zero the better. |
| | Objective | Goal or purpose for applying the specific metric |
| | References | References to bibliographical or URL resources, where additional and authoritative information of the given metric can be consulted. |
| | Accuracy | A quantifier of the accuracy level inherent to the way of producing the value of a metric. |
| Numerical Scale | representation | A numerical scale can be continuous or discrete. |
| Range | name | Name of a range to be identified. |
| | description | An unambiguous description of the range meaning. |
| | upper_threshold | |
| | lower_threshold | |
| Scale | scaleType | The type of scales depends on the nature of the relationship between values of the scale [10]. These types of scales are commonly defined: nominal, ordinal (restricted or unrestricted), interval, ratio, and absolute. |
| | valueType | |
| Software Tool | Name | Name of a software tool to be identified. |
| | Description | An unambiguous description of a software tool. |
| | Version | Number that indicates a software tool version. |
| | Provider | Indicates the name (or URL) of a software tool supplier. |
| Unit | name | Name of a unit to be identified. |
| | description | An unambiguous description of the unit meaning |
| | acronym | |
| | | |

Tabla B.3. Descripción de las Relaciones incluidas en la Ontología de Métricas e Indicadores.

| Name | Description |
|-------------------------|--|
| associated_with | One or more measurable attributes are associated with one or more entity categories. |
| automated_by | One or more methods can be automated by none or several software tools. |
| Belongs to | |
| Calculated using | An indirect metric is calculated using a specific calculation method |
| combines | A calculable concept combines (associates) one or more measurable attributes. |
| contains | A metric or an indicator contains a specific scale. |
| described_by | |

Medición y Evaluación de Calidad en Uso de Aplicaciones Web

| Name | Description |
|----------------------------|--|
| evaluates/estimates | A global indicator evaluates/estimates a calculable concept. |
| expressed_in | A numerical scale must be expressed in a specific unit. (In a strict sense, there is no idea of unit for categorical scales) |
| generates | |
| has | A decision criteria has one or more ranges. |
| has | An indicator model has one or more decision criteria. |
| includes | An indicator includes a specific calculation method. |
| interprets | An elementary indicator may interpret none or one specific metric. |
| measured_using | An direct metric is measured using a specific measurement method |
| modeled_by | An elementary (or global) indicator is modeled by one elementary (or global) model. |
| produces | A measurement (or calculation) activity produces a specific measure (or indicator) value. |
| quantifies | One or more metrics can quantify an attribute. |
| refers-to | A measurement activity is related to a metric (description). None or several measurements can be made on the same metric. |
| related_indicators | A global indicator can be structured (aggregated) on the basis of two or more related indicators. |
| related_metrics | An indirect metric can be structured on the basis of two or more related metrics. |
| related-to | A calculation activity is related to an indicator (description). None or several calculations can be made on the same indicator. |
| represented_by | A calculable concept can be represented by none or several concept models. |
| specifies | |
| specified_by | An indirect metric is specified by a given function (or formula). |
| subConcept | A calculable concept may be composed of none or several sub-concepts, which are in turn calculable concepts. |

Apéndice C: Diseño y definición de las tareas

En este apéndice se presenta el diseño de las tareas que se propusieron a los estudiantes en el caso de estudio presentado en los capítulos 5 y 6.

Si bien las tareas están especificadas en el capítulo 6 aquí se ofrece una especie de hoja de ruta de lo que fue el proceso de diseño. Se trata en cierto modo de un aspecto relevante de la propuesta de la tesis, en tanto la definición de las tareas es una nueva actividad dentro de los pasos definidos para la metodología de evaluación adoptada.

En la tabla C.1 se presentan aquellos aspectos que se tuvieron en consideración para la definición de las tareas. De esta tabla se derivaron: la redacción de la propuesta a los usuarios para que realizaran las tareas y parte del texto donde se presenta el caso de estudio.

| Nombre Tarea | Descripción | Texto Presentación | Notas |
|-------------------------------------|--|---|--|
| Navegación de Contenidos | El usuario deberá, una vez franqueado el acceso al sitio, seleccionar el enlace correspondiente a la Unidad 1, Clase 3 titulada Los Números Reales y luego encontrar, en la Topico 2 la Actividad 1, debiendo hacer clic sobre el link correspondiente | Navegá dentro el sitio Web hasta encontrar y poder desplegar la Actividad 1 correspondiente a la clase titulada "Los Números Reales" | El hecho de <i>realizar</i> la actividad puede implicar tiempos distintos de acuerdo al nivel de conocimiento de la asignatura, y no necesariamente un problema de usabilidad del producto/servicio para aquellos que resulten mas lentos. Es por eso que no se solicita completar y responder. |
| Realizar una consulta a un Profesor | El usuario deberá enviar un mensaje al profesor, completando los campos obligatorios y , para considerarse bien realizada deberá seleccionar el profesor que desarrolló los contenidos visualizados en la Clase sobre Los Números Reales. | En relación a la clase sobre Los Números Reales, envíe la siguiente consulta al profesor que desarrolló el tema: <i>¿Por qué el cero no es un numero natural?</i> . | <i>Nota:</i> Para saber quien es el profesor que desarrolló el capítulo sobre el que se hace la consulta deberá prestar atención a la columna titulada De dentro de la pestaña PIZARRON, luego de hacer clic en el enlace Introducción a la Matemática, una vez validado como usuario en el campus virtual. |
| Participar en un foro de opinión | El usuario deberá localizar la funcionalidad para | Envíe la siguiente opinión sobre el tema | |

| Nombre Tarea | Descripción | Texto Presentación | Notas |
|---|--|--|---|
| | participar en un foro, titulada Opiniones , dentro del menú de navegación horizontal, ubicado a la izquierda de la ventana principal. Seleccionar el tema, completar los campos | denominado <i>Acceso a Internet</i> :“El acceso a Internet resultó adecuado para acceder al material de estudio del curso”. El tema está previamente establecido en el foro. Verifique que su opinión ha sido incluida junto a las existentes. | |
| Organizar el material de estudio en carpetas personales | Para pasar la Clase a la carpeta Finalizados debe hacer clic sobre uno de los botones incluidos al pie de la pagina identificado con el icono ✓ | Suponiendo que ya hayas concluido con la Actividad 1 sobre Los Números Reales, señalará esa clase como finalizada. | Para verificar si esta tarea está bien realizada (y completa) el usuario podrá verificar si la clase ya no está visible en el pizarrón y en cambio se halla en la carpeta <i>finalizados</i> , dentro del “Portafolio”. |

Tabla C.1. Aspectos relevantes para la definición de Tareas.

C.1. Definición de las Tareas para el Caso de Estudio

Tarea 1. Búsqueda de información (browsing)

Buscar y desarrollar la actividad 1 sobre Los Números Reales. Dicha actividad se encuentra en la sección 2/13 de la Unidad 1.

Tarea 2. Comunicación asincrónica (a. mensajería, b. participación en foros)

En relación a la clase sobre Los Números Reales, *enviar la siguiente consulta* al profesor que desarrolló el tema: *¿Por qué el cero no es un número natural?*

Tarea 3. Participación en foros

Enviar el siguiente comentario sobre el tema de *Acceso a Internet*, “[comentario detallado]”.

Una vez enviado el comentario compruebe que su opinión ha sido incluida junto a las ya existentes.

Tarea 4. Gestión del material de aprendizaje

Suponiendo concluida la actividad 1 sobre Los Números Reales, señalar esa clase como completada dentro de sus carpeta personales.(Portafolio).

C.2. Criterios para Evaluar Correctitud y Completitud de las Tareas

Se han considerado aspectos de completitud y correctitud de las tareas para asignarles una tasa porcentual de éxito. Teniendo en cuenta que hay ciertas tareas que pueden ser completadas pero en forma incorrecta, por Ej.: enviar una consulta (tarea 2) puede ser completada, pero enviada a un destinatario erróneo, o en el caso de participación en un foro, opinar en el foro equivocado.

Para evaluar la tasa de éxito alcanzado por cada usuario en particular y luego en forma global, o sea para todos los usuarios, se empleó para cada tarea un criterio multi-nivel definido como subconjunto.

Este criterio se define como un subconjunto de los números naturales (en una escala estrictamente ordinal). Una variable discreta X puede tomar más de dos valores, cada uno de los cuales se corresponde luego a un indicador de calidad.

Los valores para X para cada tarea son los siguientes:

Tarea 1 Navegación de Contenidos: esta tarea tiene cuatro pasos básicos, acceder a la página de la Unidad 3 sobre números reales, luego avanzar a la segunda clase y finalmente encontrar, al pie de dicha página la actividad 3 y pulsar el mouse sobre la misma, por eso podemos considerar las siguientes valores de la variable X .

0: el usuario no encontró la página con el enlace a la actividad para desplegar.

1: el usuario encontró y accedió a la página con un enlace, a su vez, a la página con la actividad para desplegar.

2: 1 + el usuario encontró la página con el enlace a la actividad para desplegar, pero no logró desplegar la actividad.

3: 2 + el usuario encontró la página con el enlace a la actividad para desplegar, y desplegó la actividad en la ventana activa, tal como especifican las instrucciones para la tarea.

Los valores para la variable X guardan correspondencia con los siguientes indicadores:

$0 \Rightarrow 0\%$, $1 \Rightarrow 33\%$, $2 \Rightarrow 66\%$ y $3 \Rightarrow 100\%$

Tarea 2: Comunicación Asincrónica: Esta tarea se puede realizar en distintas secuencias, cada una con un conjunto de pasos distintos, pero se consideró que el usuario está en condiciones de iniciarla una vez que se

despliega la ventana para seleccionar los destinatarios, de modo que se propusieron los siguientes valores posibles para la variable X .

0. El usuario no encontró el acceso (típicamente un botón para pulsar) para desplegar la lista de destinatarios.

1. El usuario desplegó la lista de destinatarios, e incluso seleccionó el destinatario correcto, pero canceló luego la acción sin enviar el mensaje.

2. El usuario seleccionó un destinatario incorrecto o todos los destinatarios y envió el mensaje con el *tema* y el *texto* indicados.

3. El usuario seleccionó el destinatario correcto y envió el mensaje con el *subject* y el *texto* indicados.

Los valores para la variable X guardan correspondencia con los siguientes indicadores :

0 => 0% , 1 => 33% , 2 =>66% y 3=> 100%

Tarea 3:*Participar en un foro de opinión.* Esta tarea se puede acometer de una única manera, pero pueden ocurrir algunos eventos que determinen que no se complete correctamente, por ej.: el usuario puede seleccionar un foro equivocado para emitir su opinión, o cancelar el envío del mensaje, aun cuando hubiera completado correctamente los campos del formulario correspondiente. Se propusieron los siguientes valores posibles para la variable X:

0. El usuario no halló el enlace al servicio titulado "Opiniones", necesario para enviar una opinión a un foro.

1. el usuario encontró y accedió a la página donde puede seleccionar el foro en el que participar

2. 1 + el usuario desplegó el formulario para enviar el mensaje, pero dentro de un foro que no es el correcto (propuesto para la tarea).

3. 1 + el usuario solamente desplegó el formulario para enviar el mensaje, dentro del foro correcto.

4. 2 + el usuario canceló la acción sin enviar el mensaje.

5. 3 + el usuario completó todos los campos del formulario para enviar opinión y lo envió al foro correspondiente.

Los valores para la variable X tienen correspondencia con los siguientes indicadores :

0 => 0% , 1 => 20% , 2 =>40%, 3=> 60%, 4=> 80% y 5=> 100%

Tarea4: *Gestión del material de aprendizaje.* Esta tarea se debía realizar pulsando sobre los botones que aparecen al pie de cada ventana con el contenido de las clases. En el caso de pulsarse el botón correcto la

clase desaparece del *Pizarrón* y aparece en la carpeta *Completados*, dentro de los Mensajes recibidos. Los valores posibles para la variable X son:

0. no encuentra la clase que deber ser pasada a completada
1. encontró la clase, pero no encontró los botones que le permitieran pasar la clase a completada.
2. encontró la clase pero pulsó sobre el botón equivocado
3. 2 + pulsó sobre el botón correcto

Los valores para la variable X guardan correspondencia con los siguientes indicadores :

0 => 0% , 1 => 33% , 2 =>66%, y 3=> 100%

C.3. Imágenes de la Realización del Caso de Estudio



Figura C.1. Uno de los usuarios en plena sesión de trabajo, filmado con una cámara digital.



Figura C.2. Sesión de trabajo abierta, seguida desde una notebook.

Apéndice D: El Caso de Estudio Presentado a los Usuarios y el Cuestionario de Satisfacción

En este apéndice se presentan las instrucciones entregadas a los participantes del caso de estudio para que llevaran adelante las tareas junto con cuestionarios para la evaluación de satisfacción. También se encuentra un cuestionario sobre información específica del usuario (sección D.3), para considerar al momento de la evaluación, cuestiones que pudieran haber infundido en el desempeño de los usuarios utilizando la aplicación Web evaluada.

El apéndice D está organizado de la siguiente manera: En la sección D.1 están las instrucciones para las tareas, en la sección D.2 el cuestionario sobre formación, preferencias, edad y género; en la sección D.3 está el cuestionario sobre *Satisfacción* y en la sección D.4 hay información acerca de cómo se usa y calcula el cuestionario para Satisfacción.

D.1. Instrucciones para los Participantes

Instrucciones Generales

Gracias por ayudarnos en esta evaluación.

El propósito de este trabajo es evaluar cuan fácilmente un alumno de la Facultad de Ingeniería de la UNLPam puede utilizar “Qplus Campus Virtual”, una aplicación Web que permite acceder al curso de Introducción a la Matemática 2004 vía Internet, estudiar en forma autónoma y eventualmente interactuar con profesores y compañeros en línea.

Para poder cumplir este propósito nosotros te propondremos realizar ciertas tareas, cuya secuencia se registrará por medio de un software de seguimiento de actividades sobre la computadora. Luego, y para facilitarnos la tarea de comprensión de los resultados, te solicitaremos que completes un cuestionario y respondas unas pocas preguntas acerca de tu formación y tu ambiente de estudio. No se te pedirán datos de identidad personal dado que no son necesarios en este estudio.

La meta de esta evaluación es colaborar en la evaluación de la Calidad en Uso del producto, y los resultados podrían ser utilizados para futuras modificaciones del producto actual.

Por favor recordá que estamos testeando “QPlus Campus Virtual” y no a vos en forma personal.

Instrucciones para llevar a cabo las tareas

Como ya has usado el sitio del curso de Introducción a la Matemática 2004 (<http://www.ingreso.ing.unlpam.edu.ar>) y, para poder evaluar la **Calidad en Uso** que ofrece te proponemos realizar las siguientes tareas:

1. Navegación de Contenidos

Medición y Evaluación de Calidad en Uso de Aplicaciones Web

2. Realizar una consulta a un Profesor
3. Organizar el material de estudio en carpetas personales
4. Participar en un foro de opinión

HACENOS SABER CUANDO ESTÉS LISTO PARA COMENZAR.

Tarea 1 – Navegación de Contenidos

Navegá dentro del sitio Web hasta encontrar y poder desplegar la Actividad 1 correspondiente a la clase titulada “Los Números Reales”.

AVISANOS CUANDO ESTES LISTO PARA CONTINUAR

Tarea 2– Realizar una consulta a un Profesor

En relación a la clase sobre Los Números Reales, envía la siguiente consulta al profesor que desarrolló el tema: ¿Por qué el cero no es un numero natural?.

AVISANOS CUANDO ESTES LISTO PARA CONTINUAR

Tarea 3– Participar en un foro de opinión

Envía la siguiente opinión sobre el tema denominado Acceso a Internet :”El acceso a Internet resultó adecuado para acceder al material de estudio del curso“.

Dicho tema ha sido previamente establecido en el foro. Tu opinión debería quedar incluida junto a las enviadas anteriormente.

AVISANOS CUANDO ESTES LISTO PARA CONTINUAR

Tarea 4 –Organizar el material de estudio en carpetas personales (Ultima tarea)

Suponiendo que hubieras terminado de estudiar sobre Los Números Reales señalará la clase correspondiente como finalizada.

AVISANOS CUANDO TERMINE.

D.2. Cuestionario sobre Formación, Preferencia, Edad y Género

Por favor, antes de iniciar las tareas completá el siguiente cuestionario:

- Experiencia usando Windows u otro sistema operativo con interfaz gráfica(años):
- Experiencia usando servicios de Internet (años):
- Experiencia en el uso de ambientes de aprendizaje vía Internet -semejantes al curso de Introducción a la Matemática- (meses):
- Actitud favorable hacia el uso de ambientes de aprendizaje vía Internet para estudiar (1-7):

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| | | | | | | |

1=Prefiero usar la computadora y "Qplus Campus Virtual" **todo lo posible** para acceder al material del curso *Introducción a la Matemática* y estudiar a partir de los contenidos del sitio Web, 7= Prefiero usar la computadora y "Qplus Campus Virtual" **lo menos posible** para acceder al material del curso *Introducción a la Matemática* y estudiar a partir de los contenidos del sitio Web.

- Genero (M-F):

- Grupo de edad: <2 20-24 26-30 31-35 >35

D.3. Cuestionario de Satisfacción

EVALUACION DE CALIDAD EN USO DEL SITIO WEB

www.ingreso.ing.unlpam.edu.ar

Esta encuesta está destinada a evaluar algunos aspectos de calidad del sitio Web del curso de Introducción a la Matemática 2004, con el objetivo de mejorar aquellos que resulten relevantes para futuras experiencias en éste y otros cursos o materias que se dicten en la Facultad de Ingeniería de la Universidad Nacional de La Pampa. Esta encuesta debe ser respondida marcando con una **X** un casillero dentro de la escala, indicando el grado de acuerdo que tenés respecto al concepto que se expresa en cada ítem. La escala tiene cinco puntos, que van desde *Totalmente en Desacuerdo* hasta *Totalmente de Acuerdo*. Por favor respondé a todos y cada uno de los ítems. Si pensás que en alguno de los ítems no podés responder marcá el punto central de la escala (**3**).

Medición y Evaluación de Calidad en Uso de Aplicaciones Web

1. Me gustaría usar el sistema en otras materias de la carrera

Totalmente en

Totalmente de

Desacuerdo

Acuerdo

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

2. El sistema me resultó complejo

Totalmente en

Totalmente de

Desacuerdo

Acuerdo

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

3. El sistema me resultó fácil de usar

Totalmente en

Totalmente de

Desacuerdo

Acuerdo

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

4. Necesitaría la ayuda de un experto para usar el sistema

Totalmente en

Totalmente de

Desacuerdo

Acuerdo

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

5. Las funciones del sistema estaban bien integradas

Totalmente en

Totalmente de

Desacuerdo

Acuerdo

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

6. Percibí que varias funciones del sistema estaban ausentes o no integradas.

Totalmente en

Totalmente de

Desacuerdo

Acuerdo

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

7. Pienso que la mayoría de los alumnos podrían aprender a usar el sistema rápidamente.

Totalmente en

Totalmente de

Desacuerdo

Acuerdo

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

8. El sistema me resultó pesado y complicado de usar.

Totalmente en

Totalmente de

Desacuerdo

Acuerdo

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

9. Me sentí confiado usando el sistema.

Totalmente en

Totalmente de

Desacuerdo

Acuerdo

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

10. Necesité detenerme para aprender varias cosas antes de poder avanzar usando el sistema

Totalmente en

Totalmente de

Desacuerdo

Acuerdo

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

D.4. Utilización y Cálculo del Cuestionario

El cuestionario está basado en una escala tipo *Likert*, con ítems de respuesta forzada, y donde los usuarios deben expresar el grado de acuerdo o desacuerdo con la afirmación expresada. Estos ítems han sido adaptados del cuestionario [SUS96]²⁵, que está expresado en términos de una escala de 5 (cinco) puntos.

D.4.1. Utilización del Cuestionario

El cuestionario fue respondido inmediatamente después de que los usuarios realizaron las tareas, antes de salir del ambiente de trabajo, sin haber dialogado con otros usuarios o con los responsables del caso de estudio. La propuesta sobre respuesta neutral en casos de duda, obedece a dos necesidades, en primer lugar cada ítem debe ser respondido sin un proceso de elaboración significativo, empleando mucho tiempo y, en segundo término, que todos los ítems deben ser respondidos.

D.4.2. Cálculo del Cuestionario

El cuestionario representa una única medida de la satisfacción de los usuarios, pero se trata de una métrica indirecta (A.3.1.1), el resultado de cada uno de los ítems, en forma aislada, no es significativo y no tiene ninguna utilidad.

Para el cálculo primero se suman los aportes de cada ítem. El puntaje de cada ítem está en un rango de 0 a 4. Para los ítems 1, 3, 5, 7 y 9 el puntaje se calcula la posición de la escala menos 1. Para los ítems 2, 4, 6, 8 y 10, el puntaje se calcula como 5 menos la posición de la escala. Multiplicando por 2.5 se obtiene el cálculo final, tal cual se expresa con el indicador *Grado de Cumplimiento respecto a Satisfacción de Usuarios* (6.2.2.3).

²⁵ Se agradece a los autores de [SUS96] la libre disponibilidad del cuestionario para su empleo con fines no comerciales.