



UNIVERSIDAD DE GRANADA

TESIS DOCTORAL

PROGRAMA DE DOCTORADO EN TECNOLOGÍAS DE LA INFORMACIÓN LA
COMUNICACIÓN

**Modelo de evaluación de la usabilidad de entornos
web basado en las metodologías de Computing with
Words y Design Thinking. Caso de uso aplicado al
Decision Support System.**

Autor

Noe Zermeño Mejía

Directores

Francisco Herrera Triguero & Rosana Montes Soldado

Granada, 25 de abril de 2024

Editor: Universidad de Granada. Tesis Doctorales
Autor: Noe Zermeño Mejía
ISBN: 978-84-1195-440-2
URI: <https://hdl.handle.net/10481/94970>

Derechos de autor

Este trabajo está licenciado bajo una licencia  CC BY-NC-ND 4.0 DEED.

Modelo de evaluación de la usabilidad de entornos web basado en las metodologías de Computing with Words y Design Thinking. Caso de uso aplicado al Decision Support System.

Noe Zermeño Mejía

Palabras clave: Evaluación de la usabilidad, Toma de decisiones Multi Experto Multi Criterio, Modelo lingüístico computacional de 2-Tuplas, Validación de cuestionarios, *A/B testing*, Sistemas de soporte para decisiones (DSS)

Resumen

Realizar la evaluación de la usabilidad y la accesibilidad web resulta relevante en una época donde la constante es el aumento del uso de sitios web con distintos fines. Con ello, la atención se dirige cada vez más hacia la mejora de la satisfacción del usuario con respecto a las interfaces de usuario y la interacción humano-máquina (HCI, por sus siglas en inglés). Así, los procesos de Toma de decisión Multi Experto Multi Criterio se aplican en situaciones donde se requiere evaluar distintas alternativas a través de un conjunto de criterios, por ejemplo, elegir el sitio web mas usable entre varios con el mismo objetivo. En esta investigación se propone una metodología centrada en el usuario (LDM4WUE) aplicado a la evaluación de la usabilidad y la accesibilidad web, en donde también se consideran los elementos de *Design Thinking*, como un enfoque centrado en el usuario para hacer la propuesta más aplicable en contextos HCI.

La aplicación de LDM4WUE provee un resultado lingüístico altamente interpretable alineado al estándar de la escala adjetiva del cuestionario Escala de Usabilidad de Sistema (SUS, por sus siglas en inglés). En esta investigación, LDM4WUE utiliza datos de orígenes heterogéneos, de tal manera que realiza un proceso de unificación entre cada una de las escalas provenientes de los cuestionarios estandarizados para la evaluación de la usabilidad. Para el proceso de unificación sin pérdida de información se utiliza el modelo computacional

de 2-Tuplas. Finalmente, se agrega la información para obtener un resultado final dentro de un conjunto de términos no balanceados en la escala adjetiva de SUS.

Se diseñó y creó un cuestionario *Emotional Linguistic Usability Test* (ELUT, por sus siglas en inglés) propio, que tuvo una validación de contenido por medio de expertos, el cual fue incluido como componente del modelo. Para evaluar la usabilidad de un sitio web, la técnica *A/B testing* facilitó la comparación de datos entre dos diseños. Ampliar el alcance de las pruebas para incluir los diseños evaluados, junto con la participación de usuarios reales y ficticios, presenta un desafío para el cual pocas herramientas en línea ofrecen apoyo.

Posteriormente, se aplicaron varias pruebas de usabilidad con usuarios reales, quienes tienen algún tipo de discapacidad, para lograr una evaluación más completa y cercana a situaciones del mundo real, lo que añade un nivel de profundidad y autenticidad a los resultados obtenidos. Sobre la ampliamente reconocida Escala de SUS, se fusionaron opiniones colectivas para presentar resultados sobre la usabilidad de cada sitio web alternativo de manera comprensible.

Además, integramos LDM4WUE en la herramienta de *A/B testing* llamada *USE-AB-Tool DSS* (Sistema de Soporte de Decisiones, por sus siglas en inglés). Se evaluaron dos versiones de un mismo sistema, aplicando nuestro modelo LDM4WUE. Finalmente, mediante la técnica TOPSIS se realizó un ranking de alternativas de los sitios web, para elegir el más usable basado en un enfoque centrado en el usuario. Por lo tanto, realizar estas pruebas de usabilidad es una tarea intensiva que se ha simplificado significativamente con el apoyo de esta herramienta DSS.

Usability evaluation model for web environments based on Computing with Words and Design Thinking methodologies. Use Case applied to the Decision Support System.

Noe Zermeño Mejía

Keywords: Usability assessment, Multi Expert Multi Criteria Decision-Making, The 2-Tuple linguistic computational model, Questionnaire validation, A/B testing, Decision Support Systems.

Abstract

Web usability and accessibility assessment is relevant in an era of increasing use of websites for different purposes. With this, the focus is increasingly directed towards improving user satisfaction with user interfaces and human-computer interaction (HCI). Thus, Multi Expert Multi Criteria Decision Making processes are applied in situations where it is required to evaluate different alternatives through a set of criteria, for example, choosing the most usable website from several sites with the same objective. This research proposes a Fuzzy Linguistic Model (LDM4USE) applied to web usability and accessibility assessment, where elements of Design Thinking are also considered, as a user-centred approach to make the proposal suitable for HCI contexts.

The LDM4USE application provides a highly interpretable linguistic result aligned to the Adjective System Usability Scale (SUS). In this study, LDM4USE operates with heterogeneous source data, in order to perform a unification process between each of the scales coming from the standardized questionnaires for usability assessment. To perform the unification process without loss of information, the 2-Tuple computational model is used. Finally, the information is aggregated to obtain a final result within a set of unbalanced terms known as Adjective SUS.

A proprietary Emotional Linguistic Usability Test (ELUT) questionnaire is designed and created, which underwent content validation by experts and is included as a component of the model. To assess the usability of a website, the A/B testing technique facilitated the comparison of data between two designs. Extending the scope of testing to include the

designs being evaluated, along with the participation of real and fictitious users, presents a challenge for which few online tools offer support.

Subsequently, several usability tests are applied with real users, who have some kind of disability, to achieve a more comprehensive evaluation closer to real-world situations, which adds a level of depth and authenticity to the results obtained. On the widely recognised SUS scale, collective opinions are merged to present results on the usability of each alternative website in a comprehensible way. Furthermore, we integrate LDM4WUE into the A/B testing tool called *USE-AB-Tool* DSS (Decision Support System). Two versions of the same system are evaluated, applying our LDM4WUE methodology. Finally, using the TOPSIS technique, a ranking of website alternatives was performed to choose the most usable one based on a user-centred approach. Performing the usability assessment is an intensive task that has been significantly simplified with the support of this DSS tool.

Agradecimientos

Con gran emoción y humildad expreso mi agradecimiento a todos aquellos que han hecho contribuciones significativas para la realización de este trabajo de investigación doctoral. Su apoyo, aliento y orientación son fundamentales en cada paso de este viaje académico.

En primer lugar, me gustaría expresar mi sincero agradecimiento a mis asesores de tesis, el Dr. Francisco Herrera Triguero y Dra. Rosana Montes Soldado, por su dedicación, sabiduría y guía experta durante este difícil proceso. Su compromiso con la excelencia académica y el apoyo continuo son fundamentales para el éxito del programa. Mi más sincero agradecimiento por sus valiosas contribuciones y su inquebrantable paciencia.

Agradezco profundamente a mis asesores por su compromiso y asesoramiento durante el desarrollo de este estudio. Sus perspicaces comentarios y su disposición a compartir sus experiencias han sido invaluable para mí.

Quiero expresar mi gratitud a la Universidad de Granada por brindarme los recursos, el ambiente académico y el apoyo institucional necesarios para llevar a cabo esta investigación. La excelencia académica y el compromiso con la investigación han sido rasgos distintivos de esta institución, los cuales han contribuido significativamente a mi crecimiento académico y profesional.

Además, comparto mi más sincero agradecimiento con la Universidad de Guadalajara por el generoso obsequio de esta beca. El altruismo de la universidad y su énfasis en ser un centro de aprendizaje, investigación, innovación y crecimiento profesional han desempeñado un papel fundamental en mi trayectoria académica. Esta beca no solo me ha liberado de la tensión financiera que conlleva obtener una educación, sino que también ha sido una fuente de apoyo emocional y motivación, animándome a perseguir mis objetivos con pasión. Es con gran respeto y lealtad que valoraré la confianza y el apoyo continuo brindado

por la Universidad de Guadalajara y, a través de un desempeño académico sobresaliente, continuaré representando su visión mientras apoyamos aún más las necesidades de nuestra comunidad.

Quiero expresar mi más profundo agradecimiento a mis amigos y a todas las personas que han brindado su apoyo incondicional durante mi doctorado, Lucas, Cristina, Daniela, Victoria, Claudia, Jeronimo, Jeovani, Liliana, Luis Ángel y Servando, su presencia, ánimo y comprensión, han sido un sostén inestimable en este viaje lleno de desafíos y descubrimientos. Cada palabra de aliento, cada gesto de solidaridad, ha sido un recordatorio constante de que no estaba solo en este camino. A cada uno de ustedes, desde lo más profundo de mi corazón, ¡gracias!. Además, agradecer especialmente a mi hermana Karla Aida Zermeño Mejía, a mi tío Gerardo Alberto Mejía Pérez y al profesor Antonio Luzón, por sus palabras de apoyo en cada etapa de este proceso.

A mi familia, les estoy profundamente agradecido por su amor incondicional, su aliento constante y su sacrificio para hacer posible esta realización. A cada persona de mi familia les debo todo lo que soy y lo que he logrado en la vida. Su apoyo inquebrantable y sus palabras de ánimo fueron mi fuente de fortaleza en los momentos más desafiantes.

Índice general

1	Introducción	3
1.1	Planteamiento del problema	3
1.2	Hipótesis de la investigación	8
1.3	Objetivos de la investigación	8
1.4	Estructura de la tesis	9
2	Toma de decisiones aplicada a problemas del mundo real	11
2.1	Modelo de Toma de Decisión	11
2.2	Contribuciones de la Toma de Decisión Lingüística	13
2.2.1	Variable lingüística	14
2.2.2	Modelo Computacional Lingüístico de 2-Tuplas	17
2.2.3	Información multigranular balanceada	19
2.2.4	Información multigranular no balanceada	21
2.3	Técnicas de ranking para el enfoque lingüístico	22
2.3.1	Proceso Analítico Jerárquico Difuso Extendido	22
2.3.2	Fuzzy TOPSIS	25
2.4	Resumen de capítulo	26
3	La importancia de la evaluación de la usabilidad	29
3.1	Design Thinking como metodología de Diseño Centrado en el Usuario . . .	29
3.1.1	Principios y fases del <i>Design Thinking</i>	30
3.1.2	Técnicas de Ideación y Empatía	32
3.2	Accesibilidad web	33
3.2.1	Marco teórico	33
3.2.2	Normatividad	35
3.2.2.1	Normativas ATAG	35

3.2.2.2	Normativas de accesibilidad WCAG 2.2	36
3.2.3	Evaluación de sitios accesibles	36
3.3	Usabilidad	38
3.3.1	Marco teórico	39
3.3.2	Normativas ISO	41
3.3.3	Evaluación de la usabilidad	43
3.3.3.1	Métodos de evaluación de la usabilidad	43
3.3.3.2	Cuestionarios estandarizados	47
3.4	Resumen de capítulo	49
4	Diseño de un cuestionario para la evaluación de la usabilidad	51
4.1	Aspectos deseables para la calidad de un cuestionario	52
4.2	Procedimiento para el diseño de un cuestionario	53
4.3	Solución Software para la validación de constructo: el modelo lingüístico 2-Tuplas Fuzzy Delphi	54
4.3.1	Definición del cuestionario	58
4.3.2	Aplicación del método 2TFLLD: Primera ronda	59
4.3.3	Aplicación del método 2TLFD: Segunda ronda	64
4.4	Cuestionario ELUT	67
4.4.1	Validación de la confiabilidad del cuestionario ELUT	68
4.4.2	Cuestionario final Q_1	71
4.5	Resumen de capítulo	73
5	Un modelo de toma de decisiones lingüísticas multiexperto y multicri- terio para la evaluación de la usabilidad	75
5.1	Descripción general y diagrama de flujo de la metodología LDM4WUE. . .	76
5.2	Fase 1: Descripción del Problema	77
5.3	Fase 2. Empatía y role-playing	80
5.4	Fase 3: Recolección de la información de los usuarios	82
5.5	Fase 4. Agregación colectiva	91
5.6	Fase 5. Explotación de datos	94
5.7	Resumen de capítulo	96
6	Herramienta web <i>USE-AB-Tool Decision Support System</i>	97

6.1	Diseño de una herramienta para A/B testing	98
6.2	Aspectos técnicos de <i>USE-AB-Tool</i>	99
6.3	Manual de uso para LDM4USE	101
6.3.1	Ingreso al sistema	101
6.3.2	Creación de un proyecto de evaluación de usabilidad	105
6.3.3	Realización de los test por los evaluadores	109
6.3.4	Obtención del informe de usabilidad	110
6.4	Resumen de capítulo	113
7	Caso de estudio: Aplicación web de <i>USE-AB-Tool</i>	115
7.1	Caso de estudio: Descripción de <i>USE-AB-Tool</i>	115
7.2	Fase 1. Estructuración del problema	116
7.3	Fase 2. Empatía y role-playing	120
7.4	Fase 3. Solicitud de los datos en el caso de uso	121
7.5	Fase 4. Agregación colectiva del caso de uso	124
7.6	Fase 5. Explotación de la información del caso de uso	126
7.7	Resumen de capítulo	129
8	Conclusiones y trabajo Futuro	131

Índice de figuras

2.1	Esquema general para la resolución de Problemas de Toma de Decisión Lingüístico.	14
2.2	Conjunto de Términos Lingüísticos Balanceado.	16
2.3	Representación gráfica de información granular no balanceada.	17
2.4	El valor α representa la traslación al término mas cercano.	18
2.5	Traslación de LTS en S^3 , S^5 o S^7 al nivel t^* , para el caso S^{13} al utilizar EHL.	21
2.6	Comparación de las valoraciones adjetivas con la puntuación SUS.	22
3.1	Etapas del <i>Design Thinking</i>	31
3.2	Niveles de cumplimiento para A, AA y AAA, según el estándar WCAG 2.2.	36
3.3	Resumen de resultados de la herramienta Wave al evaluar la página principal de un sitio web.	38
3.4	Norma ISO/IEC 25000 y sus divisiones.	42
3.5	Estructura de los modelos de calidad en la norma ISO 25010.	44
4.1	Diagrama del método Delphi Lingüístico Difuso de 2-Tuplas.	55
4.2	Ejemplo de una tabla de respuestas para un cuestionario en 2TLFD-DSS.	58
4.3	Opción de filtrado en 2TFLD-DSS.	58
4.4	Opción de recorte en 2TFLD-DSS.	59
4.5	Herramienta para ocultar items según el nivel de consenso del panel de expertos en 2TLFD-DSS.	59
4.6	Correo electrónico enviado a los jueces que evaluaron el cuestionario.	61
4.7	Resultados finales del 2TLFD-DSS para Q_0	64
4.8	Resultados finales del 2TLFD-DSS para Q_1	66
5.1	Diagrama de flujo que representa las fases de la aplicación.	78
5.2	<i>Roleplaying</i> que permite empatizar con usuarios.	78
5.3	Equivalencias para escala SUS.	83

5.4	Conjunto de términos lingüísticos S_{SUS} de acuerdo a la jerarquía $LH = S^3 \cup S^5 \cup S^9$ marcada con círculos.	87
5.5	Representación semántica de escala S_{SUS}	87
6.1	Modelo Aplicación-Servidor utilizado para el desarrollo de <i>USE-AB-Tool</i>	101
6.2	Formulario de registro de usuario.	102
6.3	Vista de <i>Log in</i>	103
6.4	Vista de inicio.	103
6.5	Vista de la sección <i>Projects</i>	104
6.6	Página de la sección <i>Test</i>	104
6.7	Vista para crear un proyecto nuevo.	105
6.8	Configuración de las alternativas.	106
6.9	Configuración de los criterios.	106
6.10	Configuración de la importancia de los criterios.	107
6.11	Configuración del peso de la opinión de expertos y usuarios finales.	108
6.12	Configuración de los roles para la evaluación de la usabilidad.	108
6.13	Configuración de la importancia de los roles.	109
6.14	Vista del informe de usabilidad en <i>USE-AB-Tool</i>	112
7.1	Vista de la página de inicio de las dos versiones del sistema <i>USE-AB-Tool</i>	117

Índice de tablas

Tabla 2.1	Notación de la granularidad según el nivel de la jerarquía lingüística.	20
Tabla 2.2	Términos lingüísticos para FAHP.	23
Tabla 3.1	Disciplinas del HCI con énfasis en la usabilidad.	39
Tabla 3.2	Cuestionario SUS en español.	48
Tabla 4.1	Conjunto de preguntas para la primera ronda sobre Q_0 de ELUT.	60
Tabla 4.2	Panel de expertos.	61
Tabla 4.3	Valoraciones en escala original de los jueces para el ítem I_4 .	62
Tabla 4.4	Valoraciones de los 13 jueces unificada a S^{13} .	63
Tabla 4.5	Valoraciones en escala original de los jueces para el ítem I_4 para la segunda ronda sobre Q_1 de ELUT.	65
Tabla 4.6	Valoraciones de los 13 jueces unificada a S^{13} para el ítem I_4 en la ronda 2.	66
Tabla 4.7	Tabla comparativa de resultados lingüísticos S^7 , consenso CS y confianza RS para cada ítem I_r en ronda 1 y 2.	67
Tabla 4.8	Prueba de usabilidad para redes sociales.	69
Tabla 4.9	Test de Usabilidad Lingüístico Emocional.	70
Tabla 4.10	Estadísticas de fiabilidad.	70
Tabla 4.11	Estadísticas de total de elemento.	71
Tabla 4.12	Conjunto de preguntas para el cuestionario ELUT tras la ronda dos.	72
Tabla 7.1	Valores lingüísticos asignados por el moderador para obtener la importancia de cada criterio.	118
Tabla 7.2	La matriz CP se configura con las evaluaciones del moderador para cada criterio.	118
Tabla 7.3	Cálculo de <i>Fuzzy sintetic extension</i> .	119
Tabla 7.4	Cálculo del <i>possibility degree</i> .	119

Tabla 7.5 Vista de las respuestas de $C_1 \cong SUS$ para cada A_i 122

Tabla 7.6 Vista de las respuestas de $C_2 \cong NPS$ con NPS_LTR_score para cada A_i 122

Tabla 7.7 Un usuario U_4 del conjunto E evalúa C_3 jugando el rol R_1 122

Tabla 7.8 Vista de las respuestas de $C_4 \cong ACC$ con etiquetas lingüísticas correspondientes a Accesibilidad. 123

Tabla 7.9 Vista de las respuestas de $C_5 \cong ELUT$ con $ELUT_score$ para cada A_i . 123

Tabla 7.10 Prueba de usabilidad C_3 para evaluar $USE-AB-Tool$ 124

Tabla 7.11 Elementos de las matrices $ID^{k,l}$, representadas como 2-Tuplas. 125

Tabla 7.12 Matrices de Decisión Unificadas (UID) expresadas en S^9 125

Tabla 7.13 *Role playing* derivado en \mathbf{ucd}_i^1 , el vector 2-Tuplas con la evaluación de la usabilidad para cada alternativa. 126

Tabla 7.14 Para cada alternativa, el modelo LDM4WUE puede calcular e informar puntajes combinados de usabilidad basados en la importancia de los roles en cada test. 126

Tabla 7.15 Ranking de alternativas por roles. 127

Tabla 7.16 Valores utilizados para ordenar las alternativas en $Ranking^{global}$ 128

Tabla 7.17 Puntuaciones finales del informe dado bajo escala adjetiva S^{SUS} 128

Índice general

Índice de figuras

Capítulo 1

Introducción

En este apartado se describe de manera general los ejes centrales de la investigación y se divide en cuatro secciones: La sección 1.1 da a conocer el planteamiento del problema, así como la formulación del problema de investigación. La sección 1.2 detalla la hipótesis planteada al inicio de la investigación. La sección 1.3 presentan los objetivos de la investigación. Finalmente, la sección 1.4 describe la estructura de la memoria de tesis.

1.1. Planteamiento del problema

En la actualidad, el uso de Internet es intrínseco en la vida moderna, ya que permite a las personas acceder a una variedad de información mundial que se genera al instante y actualiza rápidamente los servicios en línea, como comunicación, entretenimiento, trabajo y educación. A lo largo de los años, el número de usuarios de Internet ha aumentado significativamente y, en la actualidad, es una de las tecnologías ampliamente utilizadas en todo el mundo [1, 2]. Según un informe de la Unión Internacional de Telecomunicaciones (ITU) de 2022 [3], se estima que dos tercios de la población mundial tiene acceso a Internet. Además, según un estudio de Statista de febrero de 2024 [4], alrededor de 5,200 millones de usuarios activos de social media de la población mundial.

Este número sigue creciendo constantemente y se espera que aumente aún más en los próximos años, a medida que más personas adopten la tecnología. Además, la evolución de las tecnologías de la información ha dado lugar a nuevas formas de acceso a Internet, así como disciplinas que han hecho más fácil y conveniente para las personas conectarse en

cualquier momento y lugar [5, 6].

La disciplina que se encarga de evaluar y mejorar, a través de técnicas y herramientas, la interacción entre los humanos y los sistemas de información, es la HCI (*Human Computer Interaction*, por sus siglas en inglés) [7]. Sin embargo, con el tiempo, se reconoció que la eficacia y eficiencia no eran suficientes para garantizar la satisfacción del usuario y su aceptación del producto. Por lo tanto, se desarrolló la *User Experience* (UX) [8] como un enfoque más amplio que incluye no sólo la interacción, sino también aspectos emocionales, sociales y culturales de la experiencia del usuario.

Una problemática de la UX es que los sitios web no siempre consideran las necesidades y expectativas de los usuarios, lo cual implica una baja satisfacción y poca lealtad. Asimismo, si los usuarios tienen una mala experiencia, las posibilidades de que compartan comentarios negativos del sitio web aumenta, lo que reduce el tráfico de datos [9, 10].

Para analizar los aspectos antes mencionados se utiliza la usabilidad [11], ya que es una parte fundamental de la UX, este se enfoca en cómo los usuarios interactúan con un producto o servicio digitales, y las emociones que se generan al usar las UX. Si la usabilidad es insuficiente, o no adecuada, para los usuarios, pueden experimentar frustración y dificultades para completar tareas, lo que lleva a una experiencia negativa y reduce la eficacia del producto o servicio digital. Al centrarse en la usabilidad desde la perspectiva de la UX, se mejora significativamente la experiencia del usuario y, la efectividad y éxito del sitio web.

Los problemas de la usabilidad son variados y complejos, en algunos casos, el diseño del producto es complicado o confuso para los usuarios, lo que dificulta que logren sus objetivos rápida y efectivamente. Otro problema, es la falta de retroalimentación o la información poco clara sobre el estado de la tarea en curso, esto puede llevar a una disminución en la eficacia de uso, en especial, en entornos educativos.

Agregando, la accesibilidad se ha reconocido como un atributo de la usabilidad, dado que se enfoca en hacer accesible un recurso web, ya sea que la persona tenga o no una discapacidad [12].

Con el objetivo de medir el grado de facilidad de uso y satisfacción que los usuarios tienen al interactuar con un producto o sistema, se desarrollaron técnicas de evaluación de la usabilidad y la accesibilidad. La evaluación de la usabilidad, en sistemas web, desempeña

un papel fundamental en la mejora de la experiencia del usuario y es esencial para identificar y abordar posibles obstáculos que los usuarios encuentran al interactuar con un sitio web. La necesidad de evaluar la usabilidad viene de años atrás; un estudio realizado por Pearson (2008), menciona que existe un impacto directo entre las características de cualquier individuo y los componentes de la usabilidad. La norma ISO 9621/25010 relaciona la eficacia, eficiencia y satisfacción como los atributos principales de la usabilidad. Por su parte, Jakob Nielsen propuso cinco dimensiones para su evaluación: 1) facilidad de aprendizaje, 2) facilidad de memorizar, 3) eficiencia, 4) manejo de errores y, 5) satisfacción del usuario.

Los elementos mencionados en el párrafo anterior son evaluados desde la visión de expertos en disciplinas del Human Computer Interaction (HCI), y pocas veces se obtiene la opinión de los usuarios que en realidad hacen uso del sistema. Debido a lo anterior, los usuarios finales podrían tener una opinión totalmente distinta a los expertos. Además, para evaluar los atributos de usabilidad, se han creado distintos instrumentos de recolección de datos estandarizados tales como: *System Usability Scale* (SUS), *Net Promoter Score* (NPS), *Questrionnaire for user Interaction Satisfaction* (QUIS), *Web site Analysis and MeasureMent Inventory* (WAMMI), entre otros. Sin embargo, se debe considerar que al utilizar escalas numéricas o tipo Likert, es difícil obtener un resultado aproximado al razonamiento humano. Además, en virtud de que la usabilidad se compone de distintos criterios, se puede considerar un problema de toma de decisiones.

Mediante la realización de un *A/B testing* se observa la usabilidad y la efectividad de un diseño, a través de la comparación de dos versiones diferentes (A y B) de diseño de un producto o servicio [13]. Aunque esta técnica es ampliamente utilizada algunas limitaciones [14], debido a lo complejo de realizar una evaluación de más de dos alternativas, ya que no proporciona información sobre por qué una versión funciona mejor. Para obtener información más profunda, se recomienda utilizar simultáneamente otras técnicas de evaluación, como las entrevistas de usuario.

Por lo que las técnicas clásicas de evaluación de la usabilidad, como las pruebas de usabilidad, las inspecciones heurísticas y las encuestas de satisfacción del usuario, han sido ampliamente utilizadas y han demostrado su eficacia para identificar problemas de usabilidad, sin embargo, también presentan algunas limitaciones y desafíos tales como: variedad de usuarios, contexto de uso, subjetividad, costo y tiempo, medición de métricas

cualitativas, herramientas y metodologías limitadas [15].

Por su parte, al evaluar la accesibilidad, se utilizan principalmente herramientas web, sin embargo, para tener una mejor perspectiva, es necesaria la inspección humana del código HTML, principalmente por parte de un experto, lo que ocasiona que sea complicada y realizada solo por especialistas.

Por lo que así, se hacen evidentes algunas de las limitaciones de las técnicas clásicas de la evaluación de la usabilidad y la accesibilidad, como el reclutamiento de participantes, ya que se realiza a través de expertos, o, se debe conseguir una muestra significativa y representativa de la población objetivo. Por lo tanto, el costo se muestra implícito tanto en el pago a los expertos, como en el reclutamiento para conseguir participantes adecuados, lo que limita la validez de los resultados obtenidos. Aunado a lo anterior, el diseño de la prueba de usabilidad, con escenarios y tareas realistas para los participantes, requiere tiempo y esfuerzo. Además, una vez completada la prueba de usabilidad, es necesario analizar y procesar los datos para obtener conclusiones útiles. El análisis requiere la contratación de expertos en usabilidad o el uso de software especializado, lo que genera costos adicionales.

Por otra parte, el sesgo del evaluador afecta la objetividad de los resultados, especialmente en técnicas como las inspecciones heurísticas, donde el evaluador utiliza su propio criterio para identificar problemas de usabilidad. Agregando que las pruebas de usabilidad y otras técnicas clásicas de evaluación se realizan en un entorno controlado y podrían no reflejar las condiciones reales en las que los usuarios utilizan el servicio digital.

Pocas técnicas de evaluación de la usabilidad y la accesibilidad utilizan enfoques centrados en los usuarios, lo que tiene un impacto negativo en la evaluación final. Por ello, en la evaluación de la usabilidad se utilizan las disciplinas UX, como complemento en la evaluación de la satisfacción del usuario, al identificarse como una medida crítica del éxito del sistema digital. Sin embargo, la satisfacción es en gran medida subjetiva. Las encuestas, las entrevistas y los grupos de discusión proporcionan una comprensión más profunda de las opiniones y perspectivas de los usuarios, mientras que la observación de la interacción del usuario con el sistema provee una comprensión más objetiva de la facilidad de uso y eficacia de la experiencia [16]. Sin embargo, las evaluaciones clásicas de la usabilidad y cuestionarios de satisfacción se basan mayormente en escalas numéricas o tipo Likert [17].

Para centrarse en el usuario se requieren metodologías más cercanas al razonamiento humano, como *Computing with Words* (CW) [18, 19]. CW es una metodología que opera con las percepciones de las personas en lugar de con medidas numéricas aportando flexibilidad en la interpretación de los resultados al expresarse en lenguaje natural y no mediante números [18, 19].

CW es una metodología que se enfoca al procesamiento de información, en la cual, la información y el conocimiento se representan y manipulan en términos lingüísticos o expresiones verbales en lugar de en términos numéricos o simbólicos tradicionales. Si consideramos la evaluación de la usabilidad de un sitio desde dos versiones distintas, o la comparación de distintos sitios web con el mismo objetivo (alternativas), a través de un conjunto de criterios, es posible considerarla como un problema de toma de decisiones (DM, del inglés *Decision Making*) [20].

Diversos estudios han incorporado propuestas de Toma de Decisiones Multicriterio Multi-Experto (MEMCDM) para evaluar la usabilidad [21]. Eldrandaly et. al [22], utilizaron el método WASPAS MCDM bajo conjuntos neutrosóficos de tipo 2 y un total de 31 subcriterios para evaluar sitios web académicos desde un enfoque de UX a través de expertos. Por su parte, Agrawal et al. [23], proponen un enfoque MCDM basado en la usabilidad y la accesibilidad (WCAG 2.0) utilizando la Técnica de Orden de Preferencia por Similitud a la Solución Ideal (TOPSIS), el Proceso Analítico Jerárquico (AHP) y la herramienta WAVE para la evaluación de varios sitios web de aerolíneas indias.

Además, la evaluación de la usabilidad puede ser simple o compleja según la necesidad: sólo realizado por expertos, con expertos y personas (ficticias), o con usuarios finales reales (siendo esta opción la más adecuada pero la más cara de todas). En vista de la falta de una herramienta informática que asista la evaluación de la usabilidad en todo el proceso, se observa la necesidad de desarrollar y aplicar un sistema web que recolecte la información de todo tipo de usuarios desde uno o distintos lugares mejorando tiempos, costos, accesibilidad y satisfacción de las personas involucradas.

La evaluación de la usabilidad web mediante modelos lingüísticos difusos es de suma importancia debido a que la usabilidad web es un concepto multidimensional que abarca factores como la navegación, el diseño, la accesibilidad, eficiencia, eficacia y satisfacción, entre otros. Estos aspectos no siempre se logran medir con precisión mediante métodos

cuantitativos tradicionales, ya que la percepción del usuario es subjetiva y puede variar. Las Metodologías Lingüísticas Difusas de Toma de Decisiones [24] permiten recolectar esta ambigüedad en las respuestas de los usuarios, ofreciendo una visión más completa de cómo interactúan con el sitio web y su percepción en su uso. Al utilizar enfoques difusos, los diseñadores y desarrolladores identifican áreas de mejora para ajustar la usabilidad y así satisfacer las expectativas de los usuarios. Cuando un sistema tiene una buena usabilidad, mejora la experiencia del usuario, además que aumenta la retención de visitantes y la fidelidad al sitio web. Por lo tanto, la aplicación de modelos lingüísticos difusos de toma de decisiones conduce a una evaluación más completa de problemas ambiguos y subjetivos.

1.2. Hipótesis de la investigación

Con el propósito de orientar el trabajo para alcanzar los objetivos de investigación, se plantea la siguiente hipótesis:

La evaluación de la usabilidad y la accesibilidad, a través de los razonamientos y opiniones de usuarios calculadas desde un Modelo Lingüístico Difuso de Toma de Decisiones, proporciona información que permite medir la usabilidad de un sitio web con el objetivo de mejorar la experiencia de usuario.

1.3. Objetivos de la investigación

Debido a la importancia que tiene la satisfacción del usuario en la evaluación de la usabilidad, el objetivo general de la presente investigación es:

Proporcionar una solución software para medir la usabilidad y la accesibilidad, por medio de las opiniones de los usuarios tras la experiencia en el uso de los sitios web, a través de un modelo lingüístico difuso de toma de decisiones, considerando al usuario final como el centro de la propuesta.

Objetivo específicos.

Para cumplir el objetivo general se deben cumplir los siguientes objetivos específicos:

1. Realizar una caracterización de las métricas de evaluación de la usabilidad web.

2. Diseñar un test que cubra los aspectos de la medición de la usabilidad web que no hallan sido considerados en cuestionarios estándar.
3. Diseñar un Modelo de Toma de Decisiones que calcule la variable lingüística usabilidad a partir de la evaluación de diferentes test de usabilidad, utilizando diferentes características de usuarios para cumplir el paradigma de diseño universal y permita las opiniones tras una prueba de usabilidad genérica para sitios web.
4. Implementar una herramienta que soporte al modelo de Toma de Decisión Lingüística para favorecer el manejo de las evaluación de la usabilidad web y sea de utilidad para ingenieros y diseñadores de interfaces de usuario.
5. Valorar la usabilidad de la solución software como parte de la calidad para el apoyo al uso práctico del modelo de Toma de Decisión Lingüística del mismo y como propio Caso de Uso para el modelo de TD propuesto.

1.4. Estructura de la tesis

A continuación, se describe el contenido de cada capítulo que compone esta memoria de tesis.

- **Capítulo I.** En este capítulo se muestra el planteamiento del problema, pregunta de investigación, objetivos e hipótesis, estableciendo los cimientos del presente trabajo.
- **Capítulo II.** El capítulo de Toma de Decisiones, describe las bases de esta investigación, además de técnicas básicas para aplicar un enfoque del paradigma lingüístico.
- **Capítulo III.** En éste capítulo se desarrollan los conceptos de *Desing Thinking*, como un enfoque para comprender al usuario; Además cubre conceptos sobre la accesibilidad web y su normativa. Finalmente, se detallan los conceptos sobre usabilidad web.
- **Capítulo IV.** Con respecto al capítulo de cuestionario, se describen las características principales que debe contener un instrumento de esta naturaleza, así como los pasos para realizarlo. Posteriormente, se desarrolla el proceso para la validación de un cuestionario de elaboración propia, para la evaluación de la usabilidad, por medio

del modelo lingüístico de Toma de Decisiones [25], diseñado exprofeso para apoyar en la validación de esta tarea.

- **Capítulo V.** En este capítulo, se formaliza matemáticamente la recolección de datos, para la evaluación de la usabilidad, por medio del Modelo de Toma de Decisiones Lingüísticas Multiexperto y Multicriterio (LDM4WUE), aplicado para evaluar las alternativas.
- **Capítulo VI.** Cubre las fases de análisis, diseño, implementación de la solución software como soporte al modelo LDM4WUE. Incluye el manual de usuario y acceso al software, su propósito es ser una herramienta libre y gratuita para *A/B Testing*, utilizado para evaluar la usabilidad.
- **Capítulo VII.** En el capítulo de caso de uso aplicado, se utiliza el modelo LDM4WUE, para realizar la evaluación de la usabilidad y accesibilidad de la solución software.
- **Capítulo VIII.** Finalmente, se presentan las conclusiones de la memoria de tesis, resaltando los puntos principales, hallazgos y aportaciones de la misma.

Capítulo 2

Toma de decisiones aplicada a problemas del mundo real

La toma de decisiones desempeña un papel central en diversos ámbitos de la vida, abarcando desde las elecciones diarias a nivel personal hasta las estratégicas dentro del contexto empresarial y organizacional.

Este capítulo se distribuye de la siguiente manera: la sección 2.1, describe las características principales que conforman los Modelos de Toma de Decisiones. La sección 2.2, presenta el enfoque lingüístico en los Modelos de Toma de Decisiones. La sección 2.3, expone dos técnicas para realizar ranking de alternativas desde un enfoque lingüístico difuso. Finalmente, la sección 2.4 presenta los puntos principales de este capítulo.

2.1. Modelo de Toma de Decisión

En un marco general, un problema de toma de decisiones (*decision making* o DM, por sus siglas en inglés) es una situación en la que una persona o grupo enfrenta la necesidad de elegir entre varias opciones disponibles, cada una con sus propias ventajas, desventajas, riesgos y consecuencias. Este tipo de problema puede surgir en una amplia gama de contextos, desde decisiones personales y profesionales hasta situaciones organizacionales y sociales.

Los problemas de toma de decisiones suelen involucrar la selección del curso de acción más adecuado para alcanzar un objetivo específico o resolver un problema particular. Esto

puede implicar considerar factores como la información y los recursos disponibles, las restricciones externas, los valores personales o grupales y, las preferencias individuales o colectivas.

La Toma de Decisiones Multicriterio (MCDM, por sus siglas en inglés) es una metodología que aborda la complejidad de evaluar alternativas a través de múltiples criterios o dimensiones al tomar una decisión. En este enfoque, los decisores deben considerar evaluar un conjunto de alternativas A a través de un conjunto de criterios C que pueden o no estar sujetos a diferentes pesos o importancias.

La MEMCDM implica la integración de las opiniones y evaluaciones de múltiples expertos en un proceso de decisión. En este contexto, los problemas lingüísticos surgen cuando los expertos expresan sus juicios y conocimientos utilizando vocabulario y terminología técnica con incertidumbre o imprecisión.

Por lo general, un proceso de toma de decisiones implica los siguientes pasos:

1. **Identificación del problema u objetivo:** El proceso comienza con la identificación clara y precisa del problema que se debe resolver o del objetivo que se quiere alcanzar.
2. **Recopilación de información:** Se recopila información relevante y pertinente relacionada con el problema u objetivo. Esto puede incluir datos, hechos, opiniones, experiencias pasadas y cualquier otro tipo de información que pueda ayudar en la toma de decisiones.
3. **Identificación de alternativas:** Se generan y se evalúan diversas alternativas posibles para abordar el problema o alcanzar el objetivo. Es importante considerar una variedad de opciones antes de tomar una decisión.
4. **Evaluación de alternativas:** Se analizan y se comparan las diferentes alternativas en función de ciertos criterios como eficacia, viabilidad, costo, riesgo, etc.
5. **Toma de decisión:** Se elige la mejor alternativa después de evaluar todas las opciones disponibles. La decisión tomada debe ser coherente con los objetivos y valores de la persona o del grupo que toma la decisión.

Sin embargo, los modelos clásicos asumen que la información disponible es completa y precisa, lo cual no siempre ocurre en entornos complejos y dinámicos. Los Modelos

Lingüísticos, como la lógica difusa y la teoría de conjuntos borrosos, permiten manejar la incertidumbre y la imprecisión inherentes al lenguaje natural humano, lo que los hace más adecuados para problemas donde la información es vaga o incompleta.

2.2. Contribuciones de la Toma de Decisión Lingüística

Los problemas de Toma de Decisiones Lingüísticas (LDM) representan un dominio basado en la comprensión del lenguaje natural como elemento clave en el proceso de elección. Estos problemas surgen en una variedad de contextos, y se centran en la interpretación y evaluación de la información expresada en forma de texto o discurso para permitir una toma de decisiones eficaz.

Entonces, la representación lingüística de la información considera el lenguaje en sus distintas expresiones, incluye desde estructuras sintácticas simples hasta complejas, incluyendo palabras, frases, símbolos, gráficos, diagramas, entre otros. Una de las áreas de aplicación es en el Procesamiento del Lenguaje Natural (PLN), una disciplina interdisciplinaria que combina la computación con palabras, la inteligencia artificial y la ciencia de la computación. Estas disciplinas se encargan de comprender e interpretar el lenguaje humano de forma automatizada.

Algunos ejemplos concretos de aplicación de la representación lingüística de la información en problemas del mundo real incluyen: recuperación de información, análisis de sentimientos, traducción automática, generación de resúmenes automáticos, asistentes virtuales y chatbots, análisis de opiniones [26].

Para facilitar la resolución de problemas de LDM, Herrera y Herrera-Viedma [27] proponen una serie de pasos para la solución de problemas MCDM que se detalla a continuación (ver figura 2.1):

1. *Fase de recolección:* se transforman las expresiones lingüísticas seleccionadas por su equivalente en 2-Tuplas con la finalidad de no perder información al momento de operar con estas.
2. *Fase de agregación:* en esta fase, se selecciona un operador simbólico que agregara las

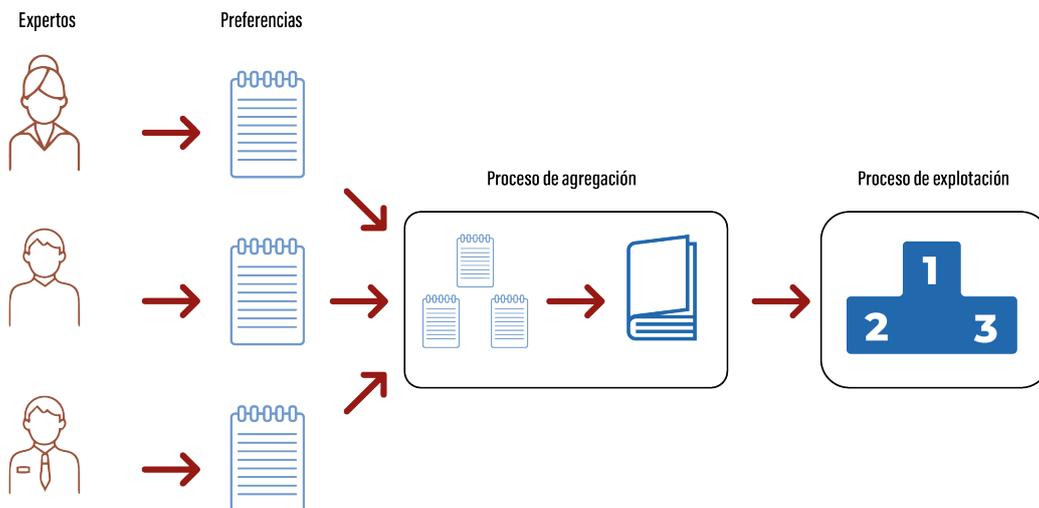


Figura 2.1: Esquema general para la resolución de Problemas de Toma de Decisión Lingüístico. Fuente: Elaboración propia.

opiniones recolectadas. Este operador se aplica a ambos lados del intervalo lingüístico.

3. *Fase de explotación:* finalmente, se ordena el conjunto de alternativas dando la solución al problema DM.

Esto se relaciona con las variables lingüísticas, ya que permiten manejar de mejor forma la información subjetiva por medio de la computación con palabras o la lógica difusa, con la finalidad de adaptarse a situaciones complejas en la toma de decisiones.

2.2.1. Variable lingüística

Para complementar el uso de computación con palabras (CW) y LDM, es necesario utilizar variables lingüísticas [28]. El enfoque lingüístico difuso manipula información cualitativa a través de variables lingüísticas para la representación de valores, que es imprecisa, pero gracias a los modelos lingüísticos es posible calcular con este tipo de información.

Las variables lingüísticas juegan un papel importante en algunas aplicaciones donde es difícil expresar una opinión por medio de valores numéricos, especialmente cuando el

mensaje proviene del lenguaje cotidiano. Dicha información es la que se trabaja por medio de variables lingüísticas cuyos valores se derivan de la percepción y difieren significativamente de los datos exactos obtenidos mediante la medición [28, 29].

Definición 2.1. Una variable lingüística se caracteriza por una quintupla $(H, T(H), U, G, G)$, donde H es el nombre de la variable; $T(H)$ es el conjunto de términos de H , donde cada valor es una variable difusa denotada como X y que varía a lo largo del universo del discurso U , el cual está asociado con una variable base u ; G es una regla sintáctica para generar los nombres de los valores de H ; y H es una regla semántica para asociar significado $M(X)$, a cada elemento de H , el cuál es un conjunto difuso de U .

Ejemplo 2.1. Se considera la variable lingüística *aires*, entonces:

H : es el nombre de la variable lingüística *aire*.

$T(H)$: es el conjunto difuso de términos lingüísticos asociados con la temperatura del aire, como **frío**, **templado** y **caliente**.

U : es el conjunto difuso de reglas que relacionan los términos lingüísticos con los valores de temperatura.

G : es el conjunto difuso de reglas gramaticales que rigen la formación de frases o expresiones lingüísticas sobre la temperatura del aire.

Por ejemplo, se pueden tener las siguientes reglas difusas en el conjunto U :

- Si la temperatura es baja, entonces es **frío**.
- Si la temperatura es moderada, entonces es **templado**.
- Si la temperatura es alta, entonces es **caliente**.

Además, el conjunto G se puede conformar con las siguientes reglas gramaticales:

- Una frase sobre la temperatura puede estar compuesta por un adjetivo de temperatura seguido de la palabra **aire**.
- Un adjetivo de temperatura puede ser **frío**, **templado** o **caliente**.

Las variables lingüísticas pueden representarse mediante números difusos; esos valores, están distribuidos dentro de un intervalo, que está establecido con una sintaxis y semántica previamente definida.

El concepto de gránulo es un grupo de objetos, físicos o mentales que comparten similitud, funcionalidad o proximidad. Cuando sus límites están claramente definidos, se conoce como granulo nítido, en caso contrario se conoce como granulo difuso [30].

Por ejemplo, la edad se granula nítidamente en números naturales, mientras que difusamente en etiquetas lingüísticas como muy joven, joven, adulto, viejo.

Dependiendo del LDM, la granularidad de las variables lingüísticas puede ser pequeña (3 términos), o bastante amplia (13 términos), siendo el conjunto de términos lingüísticos (LTS, por sus siglas en inglés) impar con una parte media.

Un LTS balanceado es aquel en el que los términos se relacionan de manera equilibrada y recíproca. Esto significa que si uno de los términos lingüísticos implica cierta medida de pertenencia, su término simétrico implicará la misma medida de pertenencia en el sentido opuesto. Por ejemplo, si tenemos un conjunto de términos lingüísticos como *Muy bajo*, *Bajo*, *Medio*, *Alto* y *Muy alto*, un conjunto simétrico aseguraría que si algo es *Bajo* con cierta medida de pertenencia, entonces su opuesto, *Alto*, tendría la misma medida de pertenencia, pero en dirección contraria. Esto ayuda a mantener la coherencia y equilibrio en la representación de la información lingüística (ver figura 2.2).

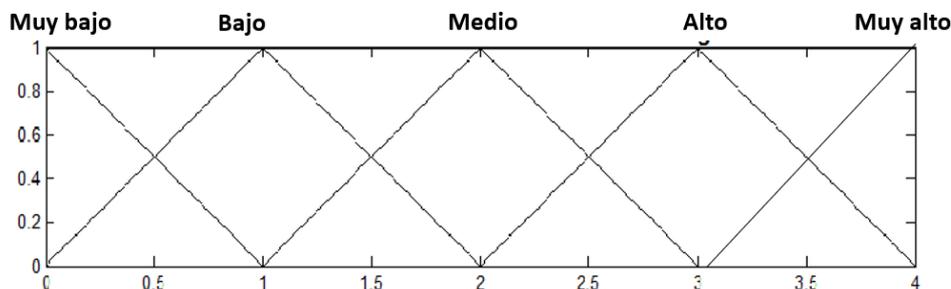


Figura 2.2: Conjunto de Términos Lingüísticos Balanceado.

Por otro lado, un LTS no balanceado serían las calificaciones utilizadas en el sistema educativo de Estados Unidos. En este sistema, las calificaciones pueden incluir letras (A, B, C, D, F) o números, por ejemplo, en un dominio de $[0, 100]$. Estos términos no son simétricos en el sentido de que no representan una relación equilibrada y recíproca entre ellos. Por ejemplo, la transición de una calificación A a un B puede no representar la misma magnitud de cambio que la transición de un D a un C . De la misma manera, la calificación F que expresa la peor evaluación, representa un gran porcentaje del total de la evaluación. En la figura 2.3, se observa claramente la asimetría de esta escala.

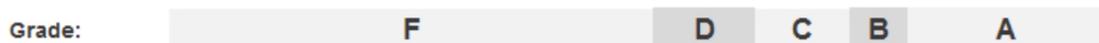


Figura 2.3: Representación gráfica de información granular no balanceada. Fuente: [31].

2.2.2. Modelo Computacional Lingüístico de 2-Tuplas

Los Modelos de Toma de Decisiones son esenciales para resolver computacionalmente situaciones de decisión en muchas áreas como la sanidad, los negocios y la educación. Requieren una gestión eficaz de las evaluaciones de expertos para clasificar las alternativas con calidad. Existen múltiples representaciones diferentes para resolver DM, entre las cuales la Representación Lingüística de 2-Tuplas tiene un gran impacto [32]. Sin embargo, al agregar las variables lingüísticas, la representación lingüística clásica no tiene suficiente precisión ya que existe pérdida de información. Debido a lo anterior, surgió la Representación lingüística de 2-Tuplas.

La Representación lingüística de 2-Tuplas [33] ha sido ampliamente utilizada en los últimos años por diversos autores [32, 34, 35] y aborda la incertidumbre utilizando términos lingüísticos para prevenir la pérdida de datos mediante un dominio continuo dentro del marco de la teoría de conjuntos difusos.

Herrera *et al.* define las funciones Δ y Δ^{-1} para transformar valores numéricos en 2-Tuplas y viceversa. Este modelo representa la información lingüística a través de un par de valores (s_g, α) , donde s_g es la etiqueta lingüística y α es el valor numérico que representa el valor de la traducción simbólica (ver figura 2.4).

Proposición 1. [36] Sea $S^{g+1} = \{s_0, \dots, s_g\}$ un conjunto de términos lingüísticos y $\beta \in [0, g]$ el resultado de una operación de agregación. La función Δ transforma un valor β en una 2-Tupla de información equivalente mediante la Ecuación 2.1.

$$\Delta : [0, g] \rightarrow S^{g+1} \times [-0.5, 0.5]$$

$$\Delta(\beta) = (s_i, \alpha), \text{ con } \begin{cases} s_i & i = \text{round}(\beta), \\ \alpha = \beta - i \end{cases} \quad (2.1)$$

donde $\text{round}(\cdot)$ es el operador de redondeo común.

Proposición 2. [36] La función inversa Δ^{-1} transforma una 2-Tupla a su valor equivalente

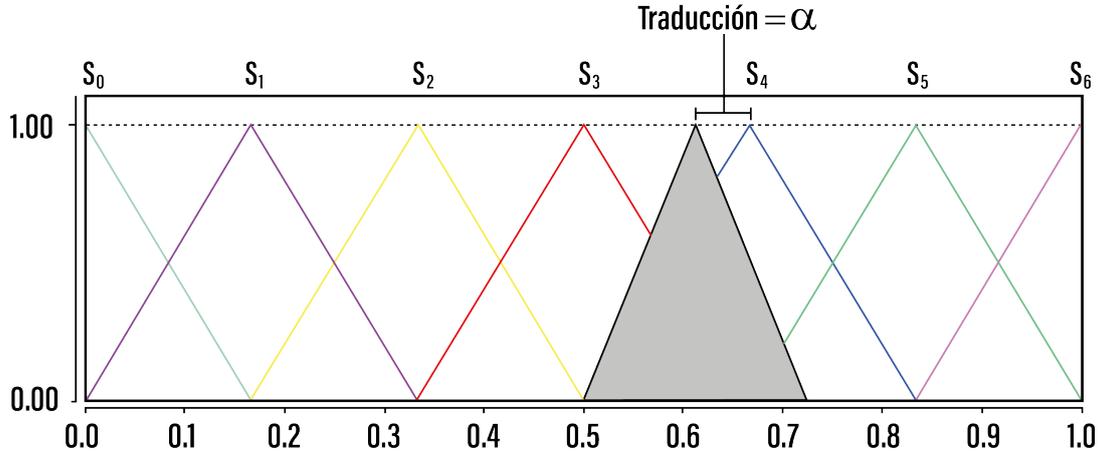


Figura 2.4: El valor α representa la traslación al término mas cercano. Fuente: Elaboración propia.

numérico $\beta \in [0, g]$ con la ecuación 2.2.

$$\Delta^{-1}(s_i, \alpha) = i + \alpha = \beta \quad (2.2)$$

La agregación de un conjunto de valores lingüísticos en una 2-Tupla debe igualmente constituir una 2-Tupla que resuma dicho conjunto. En diversos estudios se han propuesto varios operadores de agregación lingüística [37, 38, 39, 40] con el fin de facilitar y flexibilizar el proceso de agregación de información lingüística.

Sea $x = \{(s_0, \alpha_1), \dots, (s_{g-1}, \alpha_{g-1})\} = \{\beta_0, \dots, \beta_{g-1}\}$ un conjunto de valores lingüísticos 2-Tupla, $W = \{w_i | i = 1, \dots, g - 1\}$ un vector de pesos, y W' su versión normalizada como $\sum_{i=0}^n w'_i = 1$. La media aritmética ponderada extendida \bar{x}^e se define como:

$$[ht]\bar{x}^e(x) = \Delta \left(\frac{\sum_{i=1}^n \Delta^{-1}(s_i, \alpha_i) \cdot w_i}{\sum_{i=1}^n w_i} \right) = \Delta \left(\frac{1}{n} \sum_{i=1}^n \beta_i w'_i \right). \quad (2.3)$$

El Operador de Agregación con Peso Ordenado (*OWA*, por sus siglas en inglés) es una herramienta esencial en la teoría de la toma de decisiones bajo incertidumbre y en la lógica difusa. Este operador se utiliza para fusionar múltiples valores de entrada, asignando pesos

a cada valor y ordenándolos de acuerdo con su importancia relativa.

El OWA de 2-Tuplas es una técnica flexible que puede adaptarse a una variedad de escenarios de toma de decisiones donde se requiere fusionar información proveniente de diferentes fuentes o perspectivas, asignando pesos para reflejar su importancia relativa.

2.2.3. Información multigranular balanceada

El concepto de jerarquías lingüísticas se introdujo en [41] para diseñar sistemas jerárquicos de reglas lingüísticas. La estructura de jerarquías lingüísticas ha sido ampliamente utilizada en CW para mejorar el proceso en el contexto de información lingüística multigranular, donde los expertos realizan la evaluación de criterios basándose en la diferencia de conocimientos y experiencia con el criterio seleccionado.

Una Jerarquía Lingüística (LH, por sus siglas en inglés) [42] se compone de diferentes conjuntos de términos lingüísticos en los que cada uno tiene una granularidad diferente a los demás. El nivel jerárquico de cada LTS es $l(t, n(t))$, siendo t el número que indica el nivel de la jerarquía y $n(t)$ la granularidad de t . A la hora de definir el LH, se considera que las funciones de pertenencia de los términos son triangulares, uniformes y simétricas en un dominio de $[0, 1]$. Además, el valor del término central representa el valor de *indiferencia*.

Los niveles pertenecientes a la jerarquía lingüística se ordenan según la granularidad de cada uno de ellos, por ejemplo, para dos niveles consecutivos t y $(t + 1)$, $n(n(t + 1)) \geq n(t)$. De acuerdo con lo anterior, LH es la unión de todos los niveles t : $LH = \cup_t l(t, n(t))$ que componen la jerarquía lingüística. Cada nivel en el orden jerárquico conlleva un aumento de la granularidad del LTS con respecto al nivel anterior quedando de la siguiente manera:

$$LH = \cup_t S^{n(t)}(t), \quad t \in \{1, \dots, h\}$$

$$S^{n(t)}(t) = \{s_0^{n(t)}, \dots, s_{\delta_t}^{n(t)}\} \quad \text{donde } \delta_t = n(t) - 1, \quad \delta_t \in \mathbb{N}.$$

La generación de LH debe satisfacer las siguientes reglas básicas [41].

1. Preservar todos los puntos modales de las formas para las funciones de pertenencia de cada término lingüístico de un nivel al siguiente.
2. Para realizar una transición suave, hay que añadir puntos intermedios para la

construcción de $S^{n(t+1)}$ entre las etiquetas de nivel t .

Esto puede formalizarse de la siguiente manera:

$$l(t, n(t)) \rightarrow l(t + 1, 2 \cdot n(t) - 1)$$

La tabla (2.1) muestra la granularidad requerida en cada nivel t en función del valor $n(t)$.

	<i>Nivel 1</i>	<i>Nivel 2</i>	<i>Nivel 3</i>	<i>Nivel 4</i>
$l(t, n(t))$	$l(1, 3)$	$l(2, 5)$	$l(3, 9)$	$l(4, 17)$
$l(t, n(t))$	$l(1, 7)$	$l(2, 13)$		

Tabla 2.1: Notación de la granularidad según el nivel de la jerarquía lingüística.

Definición 2.2. [43] Sea $LH = \cup_t S^{n(t)}(t)$, se considera que la representación lingüística difusa 2-Tupla realiza procesos CW en contextos de información lingüística multigranular sin pérdida de información. La función de transformación recursiva se generalizó para transformar términos lingüísticos entre cualquier nivel lingüístico en LH. Además, se ha definido de forma no recursiva, $TF_{t'}^t : l(t, n(t)) \rightarrow l(t', n(t'))$ tal que:

$$TF_{t'}^t(s_i^{n(t)}, \alpha^{n(t)}) = \Delta_{t'} \left(\frac{\Delta_t^{-1}(s_i^{n(t)}, \alpha^{n(t)}) \cdot (n(t') - 1)}{n(t) - 1} \right)$$

LH presenta una restricción dado que, por ejemplo, no es posible obtener S^7 del nivel 2 utilizando S^5 . La solución radica en la utilización de Jerarquías Lingüísticas Extendidas (ELH) [44], las cuales pueden abordar cualquier LTS sin restricciones.

Las ELH comprenden una serie de conjuntos de términos lingüísticos $S^{n(t)}(t)$ asociados a un nivel t , cada uno con una granularidad $n(t)$ distinta. Un nuevo conjunto de términos lingüísticos en la ELH con $t^* = m + 1$ conserva todos los puntos modales previos.

Existe una función de transformación $TF_{t^*}^t$ para transformar un término $s_i \in S^{n(t)}$ en su término equivalente s_j expresado en $S^{n(t^*)}$ con $t < t^*$:

$$TF_{t^*}^t(s_i^{n(t)}, \alpha_j) = \Delta \left(\frac{\Delta^{-1}(s_i^{n(t)}, \alpha_i) \cdot (n(t^*) - 1)}{n(t) - 1} \right) = (s_j^{n(t^*)}, \alpha_j) \quad (2.4)$$

La figura 2.5 demuestra gráficamente la traslación de LTS en S^3 , S^5 y LTS en S^3 , S^5 hacia S^{13} .

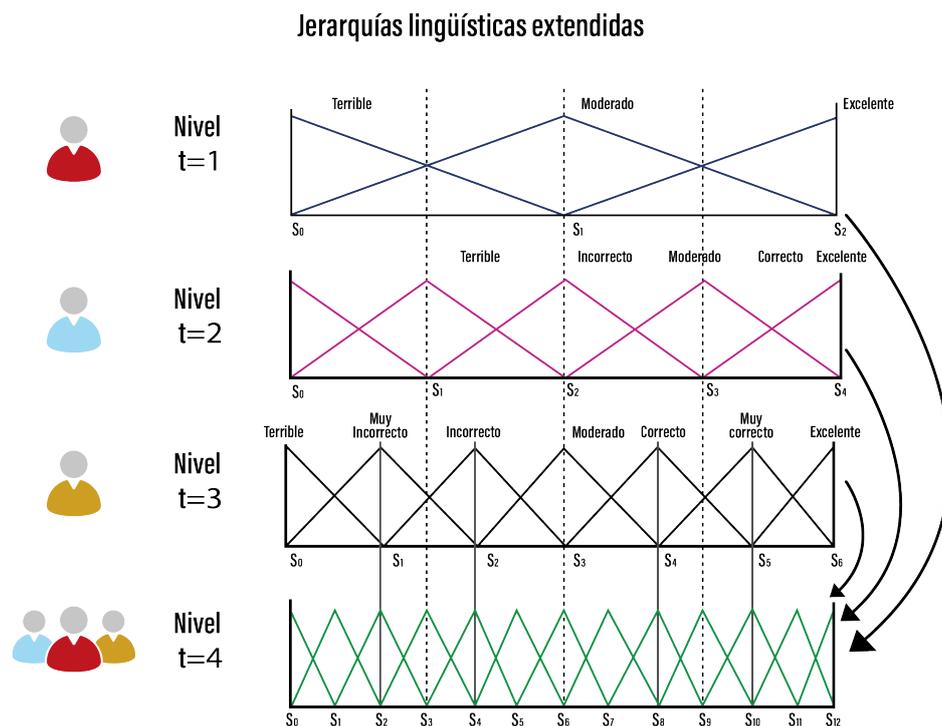


Figura 2.5: Traslación de LTS en S^3 , S^5 o S^7 al nivel t^* , para el caso S^{13} al utilizar EHL. Fuente: Elaboración propia.

2.2.4. Información multigranular no balanceada

La representación lingüística no balanceada propuesta por Herrera [43] es una técnica utilizada en el ámbito de la lógica difusa y la modelización de sistemas basados en conocimientos lingüísticos. Esta técnica se diferencia de las representaciones lingüísticas balanceadas al permitir que los términos lingüísticos tengan diferentes rangos de variación o importancia en el contexto del problema que se está modelando.

La idea principal detrás de la representación lingüística no balanceada es que los términos lingüísticos pueden tener diferentes distribuciones de membresía en el conjunto difuso asociado. Es decir, no todos los términos lingüísticos tienen la misma influencia en la modelización de la incertidumbre o en la toma de decisiones.

Por ejemplo, en un problema de control de temperatura, los términos lingüísticos *frío*, *templado* y *caliente* pueden tener diferentes rangos de variación. La representación lingüística no balanceada permite capturar estas diferencias al asignar diferentes funciones de membresía a cada término lingüístico.

En la práctica, la representación lingüística no balanceada puede lograrse mediante el uso de funciones de pertenencia asimétricas o no simétricas, que reflejen la distribución real de los valores y la importancia relativa de los términos lingüísticos en el contexto del problema específico. Un claro ejemplo de esto es la escala evaluativa de SUS (ver figura 2.6).

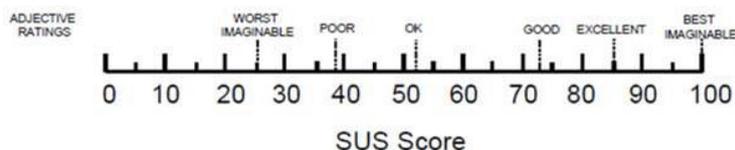


Figura 2.6: Comparación de las valoraciones adjetivas con la puntuación SUS. Fuente [45].

2.3. Técnicas de ranking para el enfoque lingüístico

Los problemas de DM han ido en aumento en la investigación científica. Para realizarlo se parte de 1 a N expertos en la materia, hasta la elección de los diversos criterios que importan en el problema específico. Hasta la fecha, se han desarrollado varios métodos y enfoques para estudiar este problema, entre los que se incluyen: TOPSIS [46, 47], VIKOR [48, 49], Analytic Hierarchy Process (AHP) [50], método difuso [51, 52], etc. En particular, AHP y TOPSIS son muy beneficiosos cuando un problema de DM es complicado. La razón es que AHP y TOPSIS pueden priorizar múltiples criterios de elección en una jerarquía mediante la evaluación de la importancia relativa de los criterios y, por lo tanto, pueden generar una clasificación general de las alternativas. [53].

2.3.1. Proceso Analítico Jerárquico Difuso Extendido

Propuesto por Saaty [50], el método AHP es una técnica cualitativa estructurada ampliamente utilizada en problemas de toma de decisiones [54] multicriterio. Para hacer frente a la imprecisión de los juicios lingüísticos, Chang [55] desarrolló un AHP difuso ampliado (FAHP) que se ha utilizado como método de investigación en varias áreas [56, 57].

Para tratar con aplicaciones reales, el FAHP se ha combinado con otras técnicas multicriterio. Una de las aplicaciones más comunes del FAHP es determinar los pesos relativos de los criterios de evaluación [58, 59]. Se recopilan los juicios por pares de los

responsables de la toma de decisiones para derivar escalas de prioridad con respecto a los criterios. Se complementa la matriz de preferencia CP con los números difusos triangulares (TFN) correspondientes a la tabla 2.3.1.

Numero difuso	Variable lingüística	Función de membresía
1	<i>Igualmente importante</i>	(1,1,1)
3	<i>Moderadamente importante</i>	(1,3,5)
4	<i>Muy importante</i>	(3,5,7)
5	<i>Fuertemente importante</i>	(5,7,9)
6	<i>Absolutamente importante</i>	(7,9,9)

Tabla 2.2: Términos lingüísticos para FAHP.

- **Paso 2.1. Obtención de juicios por pares sobre la importancia de los criterios.**

A partir del moderador, se recopila los juicios por pares de los responsables de la toma de decisiones para derivar escalas de prioridad con respecto a los criterios. Se completa la matriz CP de tamaño $m \times m$ con los Números Difusos Triangulares (*low, medium, upper*) = (l, m, u) (TFN, por sus siglas en inglés) como se presenta en la siguiente ecuación, con las valoraciones provistas por el moderador.

$$CP_{j,j'} = (l_{j,j'}, m_{j,j'}, u_{j,j'}), \forall j \leq j'; \quad j, j' = 1, \dots, m \quad (2.5)$$

y sus valores son completados a través de la ecuación 2.6:

$$CP_{j',j} = (l_{j',j}, m_{j',j}, u_{j',j}) = \left(\frac{1}{u_{j',j}}, \frac{1}{m_{j',j}}, \frac{1}{l_{j',j}} \right), \forall j > j'. \quad (2.6)$$

- **Paso 2.2. Calculo de la extensión difusa sintética.**

La extensión sintética difusa para el criterio C_j se calcula mediante las siguientes ecuaciones:

$$s_j = (l_j, m_j, u_j) = \sum_{j'=1}^m M_{C_j}^{j'} \otimes \left[\sum_{j=1}^m \sum_{j'=1}^m M_{C_j}^{j'} \right]^{-1}, \forall j = 1, \dots, m \quad (2.7)$$

siendo

$$\sum_{j'=1}^m M_{C_j}^{j'} = \left(\sum_{j'=1}^m l_{j,j'}, \sum_{j'=1}^m m_{j,j'}, \sum_{j'=1}^m u_{j,j'} \right), \forall j = 1, \dots, m, \quad (2.8)$$

$$\left[\sum_{j=1}^m \sum_{j'=1}^m M_{C_j}^{j'} \right]^{-1} = \left(\frac{1}{\sum_{j=1}^m \sum_{j'=1}^m u_{j,j'}}, \frac{1}{\sum_{j=1}^m \sum_{j'=1}^m m_{j,j'}}, \frac{1}{\sum_{j=1}^m \sum_{j'=1}^m l_{j,j'}} \right). \quad (2.9)$$

■ **Paso 2.3. Grado de posibilidad.**

Se obtiene el grado de posibilidad de los elementos s_j y $s_{j'}$, $\forall j, j' = 1, \dots, m, j \neq j'$ usando la siguiente ecuación:

$$V(s_j \geq s_{j'}) = \begin{cases} 1, & \text{Si } s_j \geq s_{j'} \\ 0, & \text{Si } l_{j'} \geq u_j \\ \frac{l_{j'} - u_j}{(m_j - u_j) - m_{j'} - l_{j'}}, & \text{en otro caso} \end{cases} \quad (2.10)$$

■ **Paso 2.4. Obtención de los vectores de pesos para el conjunto C .**

Por último, calculamos el peso de cada criterio C_j , ($j = 1, \dots, m$) a través de la siguiente ecuación:

$$WC'_j = \min[V(s_j \geq s_{j'})], \forall j, j' = 1, \dots, m, j \neq j', \quad (2.11)$$

y los normalizamos, obteniendo el peso final para cada criterio C_j , ($j = 1, \dots, m$) aplicando la siguiente ecuación:

$$WC_j = \frac{WC'_j}{\sum_{j=1}^m WC'_j}. \quad (2.12)$$

Es importante señalar que la consistencia de los resultados debe ser analizada. La aplicación de Fuzzy AHP puede ser difícil debido a los resultados inconsistentes. El Índice de consistencia debe ser inferior a 0,1 para un resultado fiable, y 0,2 es el nivel máximo tolerado, y se puede calcular de acuerdo con Saaty [50].

2.3.2. Fuzzy TOPSIS

Hwang y Yoon [60] propusieron inicialmente el método TOPSIS que es uno de los métodos multicriterio más populares utilizados en los enfoques matemáticos. Debido a su sencillez de aplicación, la técnica TOPSIS ha sido utilizada en varios estudios para clasificar alternativas competidoras en términos de su rendimiento global [61, 62, 63, 64]. A continuación se describen los pasos del algoritmo TOPSIS.

Sea D Una matriz normalizada compuesta de valores de 2-Tuplas (s_{ij}, α_{ij}) , donde $i = 1, \dots, n$ representa la alternativa evaluada y $j = 1, \dots, m$ su criterio asociado. Se asume que los criterios tienen igual importancia. El procedimiento de TOPSIS [65] determina la solución ideal positiva A^+ y la solución ideal negativa A^- como un vector de 2-Tuplas conformado por:

$$A^+ = [(r_1^+, \alpha_1^+), (r_2^+, \alpha_2^+), \dots, (r_m^+, \alpha_m^+)] \quad (2.13)$$

$$A^- = [(r_1^-, \alpha_1^-), (r_2^-, \alpha_2^-), \dots, (r_m^-, \alpha_m^-)] \quad (2.14)$$

donde

$$(r_j^+, \alpha_j^+) = \left\{ \max_i \{ (s_{ij}, \alpha_{ij}) \} \right\}, j = 1, \dots, m \quad (2.15)$$

$$(r_j^-, \alpha_j^-) = \left\{ \min_i \{ (s_{ij}, \alpha_{ij}) \} \right\}, j = 1, \dots, m \quad (2.16)$$

A continuación, se calculan las medidas de separación D_i^{+l} y D_i^{-l} de cada alternativa respecto a la solución ideal positiva y la solución ideal negativa basándose en la distancia Euclídea:

$$D_i^+ = \Delta \sqrt{\sum_{j=1}^m \left(\Delta^{-1}(s_{ij}, \alpha_{ij}) - \Delta^{-1}(r_j^+, \alpha_j^+) \right)^2} \quad (2.17)$$

$$D_i^- = \Delta \sqrt{\sum_{j=1}^m \left(\Delta^{-1}(s_{ij}, \alpha_{ij}) - \Delta^{-1}(r_j^-, \alpha_j^-) \right)^2} \quad (2.18)$$

Enseguida, se calcula el coeficiente de proximidad relativo a las soluciones ideales de

cada alternativa A_i con respecto a la solución ideal positiva A^{+l} :

$$RC_i = \Delta \left(\frac{\Delta^{-1}(D_i^-)}{\Delta^{-1}(D_i^+) + \Delta^{-1}(D_i^-)} \right) \quad (2.19)$$

donde $0 \leq \Delta^{-1}(RC_i) \leq 1$.

Por último, se establece la clasificación de las alternativas. Según el coeficiente de proximidad relativa a la alternativa ideal, cuanto mayor sea RC^{+l} , mejor es la alternativa A_i . Entonces, todas las alternativas $A_i (i = 1, 2, \dots, m)$ pueden clasificarse según un orden descendente de los valores de cercanía relativa.

Por lo mencionado anteriormente, el modelo Fuzzy TOPSIS es una herramienta útil para la toma de decisiones multicriterio en situaciones de incertidumbre e imprecisión, permitiendo a los tomadores de decisiones obtener soluciones más robustas y bien fundamentadas.

2.4. Resumen de capítulo

- Los Modelos de Toma de Decisiones brindan una estructura clara y organizada para abordar problemas complejos en la vida real, lo que facilita la comprensión del problema y sus componentes.
- Los MEMCLDM permiten la evaluación y consideración de múltiples criterios relevantes para la toma de decisiones, lo que ayuda a tener en cuenta diferentes aspectos y perspectivas.
- El Modelo Computacional Lingüístico de 2-Tuplas permite a los expertos expresar sus preferencias y opiniones de manera lingüística, lo que facilita la interpretación de los resultados y la participación de personas no especializadas en el proceso de toma de decisiones.
- La Representación multigranular balanceada permite una evaluación más precisa y completa de las alternativas, al considerar múltiples niveles de granularidad en la información.
- La Representación multigranular no balanceada ofrece flexibilidad al permitir que

diferentes alternativas tengan diferentes niveles de granularidad, lo que puede ser útil para problemas con requisitos y características heterogéneas.

- Las técnicas FAHP y TOPSIS, permiten el manejo de la incertidumbre y la imprecisión inherentes a muchos problemas de la vida real, lo que los hace más robustos y aplicables en situaciones reales.
- El análisis comparativo ayuda a identificar la opción más adecuada, dentro de un conjunto de alternativas, en función de los criterios establecidos y las preferencias de los involucrados.
- Mejorar la calidad de las decisiones, por medio de los modelos mencionados, permite que los resultados obtenidos sean interpretables y transparentes, lo que facilita la comunicación y la justificación de las decisiones tomadas. En general, la aplicación de Modelos de Toma de Decisiones Lingüísticas puede mejorar la calidad y la eficacia de las decisiones tomadas en situaciones complejas de la vida real, lo que conduce a mejores resultados.

Capítulo 3

La importancia de la evaluación de la usabilidad

Para garantizar la eficacia de los sistemas interactivos requiere metodologías centradas en el usuario como *Human-Computer Interaction* (HCI), Diseño de Interacción (IXD, por sus siglas en inglés), Interfaz Gráfica de Usuario (GUI, por sus siglas en inglés) y Diseño Centrado en el Usuario (UCD, por sus siglas en inglés). Es posible garantizar que un sistema sea fácil de usar y atractivo desde el comienzo del ciclo de vida del producto, lo que aumenta la eficacia, la eficiencia y la satisfacción del usuario del sistema, logrando una experiencia de usuario positiva y aumentando la aceptación entre los usuarios.

Este capítulo se distribuye de la siguiente manera: La sección 3.1, describe una metodología centrada en el usuario como *Design Thinking*. La sección 3.2, presenta los conceptos básicos de la accesibilidad. La sección 3.3, expone la normativa de usabilidad y métodos de evaluación. Finalmente, la sección 3.4 presenta los puntos principales de este capítulo.

3.1. Design Thinking como metodología de Diseño Centrado en el Usuario

El Diseño Centrado en el Usuario (UCD, por sus siglas en inglés) es un enfoque cuya finalidad es comprender y satisfacer las necesidades, deseos y expectativas de los usuarios a través del proceso de diseño [66]. Se basa en la idea de que la satisfacción del usuario es

30 3.1. Design Thinking como metodología de Diseño Centrado en el Usuario

la clave para el éxito de un producto o servicio, y que el diseño debe ser dirigido por las necesidades s en lugar de las características o especificaciones técnicas. Por lo tanto, se diseñan soluciones útiles, efectivas, eficientes y agradables para los usuarios [67]. Además de crear experiencias positivas y satisfactorias, fomentando su lealtad y fidelidad. Por lo tanto, el UCD propone un proceso iterativo que busca la mejora continua a través de la retroalimentación y el aprendizaje constante de los usuarios.

El Diseño Centrado en el Usuario es una parte fundamental de la metodología de *Design Thinking* (DT), ya que el enfoque en las necesidades de los usuarios es esencial para generar soluciones a las problemáticas de usabilidad. *Design Thinking* es un enfoque iterativo y colaborativo para la resolución de problemas que involucra a múltiples disciplinas, incluyendo el diseño, la tecnología y los negocios [68].

El DT [68] como enfoque de solución de problemas, ha sido desarrollado y perfeccionado por una serie de profesionales e instituciones a lo largo de los años. Sin embargo, su origen se puede rastrear hasta la década de 1950 y el trabajo del diseñador industrial alemán Dieter Rams [69] y su enfoque en la simplicidad y la funcionalidad en el diseño de productos.

DT combina la empatía con el usuario, la creatividad y la experimentación para llegar a soluciones innovadoras y sostenibles, aplicado a una amplia gama de problemas. Se trata de un proceso iterativo que permite a los equipos de desarrollo comprender las necesidades y deseos de los usuarios, idear y experimentar soluciones, y perfeccionar y validar las ideas.

El DT se aplica desde la creación de nuevos productos y servicios hasta la mejora de procesos y sistemas existentes, se basa en la idea de que la solución óptima a un problema se encuentra en la intersección entre la tecnología y la comprensión profunda de las necesidades y deseos de los usuarios.

Ademas, *Desing Thinking* también se aplica para diversas actividades en los espacios académicos, especialmente en proyectos y trabajo en grupo, ya que uno de los requisitos previos del DT es el trabajo en equipo y la comunicación abierta [70, 71].

3.1.1. Principios y fases del *Design Thinking*

En general, DT se divide en cinco fases: empatizar, definir, idear, prototipar, probar y perfeccionar. Cada una de estas fases permite al equipo de desarrollo progresar en la

comprensión del problema y en la búsqueda de soluciones efectivas y sostenibles. Las cinco fases del proceso de DT son [72]:

1. **Empatizar:** Esta fase implica comprender las necesidades, deseos y desafíos de los usuarios a través de la investigación y la observación. El objetivo es desarrollar una comprensión profunda de las personas a las que se pretende ayudar y sus experiencias cotidianas.
2. **Definir:** En esta fase, se sintetiza la información recopilada en la fase de empatizar para identificar el problema central que se desea resolver. Se definen las oportunidades y desafíos clave a partir de la comprensión de las necesidades de los usuarios.
3. **Idear:** En esta fase, el equipo genera una gran cantidad de ideas y soluciones posibles. El objetivo es encontrar la mayor cantidad de soluciones posibles para el problema definido, sin preocuparse por su viabilidad o factibilidad.
4. **Prototipar:** En esta fase, se construyen modelos o prototipos de las soluciones ideadas para poner a prueba las ideas y validar la comprensión de las necesidades de los usuarios.
5. **Probar y perfeccionar:** En esta fase, se evalúan los prototipos con los usuarios y se recopila retroalimentación para comprender qué funciona y qué no funciona. Se utiliza la retroalimentación para perfeccionar y ajustar las soluciones.



Figura 3.1: Etapas del *Design Thinking*. Fuente: Elaboración propia.

Como se mencionó previamente, el proceso de DT es iterativo, lo que significa que puede regresar y repetir cualquiera de las fases en cualquier momento durante el proceso, dependiendo de las necesidades y de los resultados. La flexibilidad y la adaptabilidad son

32 3.1. Design Thinking como metodología de Diseño Centrado en el Usuario

elementos clave del proceso de esta metodología, utilizando varias técnicas para llevar el proceso de DT.

3.1.2. Técnicas de Ideación y Empatía

Hay varias herramientas y estrategias que se utilizan en el proceso de DT para recopilar información sobre las necesidades de los usuarios, incluyendo [72]:

- **Investigación de usuarios:** Se realiza una investigación exhaustiva de los usuarios para comprender sus desafíos, necesidades y comportamientos. Esto puede incluir entrevistas, encuestas, grupos focales, análisis de datos y observación.
- **Personas ficticias:** Se crean personas ficticias o *arquetipos* que representan a los diferentes tipos de usuarios. Esto permite al equipo visualizar y comprender mejor las necesidades y desafíos de los usuarios.
- **Juegos de roles:** El uso del juego de roles ha evolucionado desde la teoría a las relaciones sociales y su interacción. El ser humano genera un rol asociado a un contexto, este rol determina su comportamiento social en función de lo aprendido, lo vivido con otras personas o las características del entorno [73]. El objetivo de utilizar esta técnica es generar empatía con otras personas, situándonos en las características de otros individuos. De esta forma, se identifica el nivel de posibilidades para realizar una actividad, la forma de reaccionar ante alguna situación concreta y las ventajas o limitaciones para ejecutar con éxito una tarea específica. Desde las primeras exploraciones, se utilizó el juego de Legos donde los usuarios simulan a los *stakeholders*, intercambiando roles entre expertos y usuarios [74]. También, se desarrolló una investigación donde se aplica la técnica de juego de roles para la educación [73], de la misma manera esta técnica es ampliamente utilizada en modelos de negocio [75, 76].
- **Mapas de empatía:** Se utilizan mapas visuales para representar las experiencias y las necesidades de los usuarios. Esto ayuda a sintetizar la información recopilada e identificar oportunidades clave.
- **Prototipado de baja fidelidad:** Se construyen modelos o prototipos de baja

fidelidad para probar y validar las ideas de solución. Esto permite a los usuarios experimentar y dar retroalimentación sobre las soluciones.

Estas son solo algunas de las herramientas y estrategias comúnmente utilizadas en el proceso de DT para recopilar información sobre las necesidades de los usuarios. Es importante destacar que el proceso es flexible y adaptable, por lo que puede utilizar una combinación de herramientas y estrategias para recopilar información en función de las necesidades específicas de cada proyecto.

Por lo tanto, DT se centra en buscar soluciones centradas en el usuario, lo cual comparte con la accesibilidad, dado que se crean productos y servicios inclusivos y accesibles para todos, incluyendo a personas con discapacidades.

3.2. Accesibilidad web

En esta sección, se revisa el concepto de accesibilidad para contextualizar su uso dentro del campo de la informática, además, se describen las normativas que se ven involucradas en el tema, así como las diversas maneras en que se puede evaluar la accesibilidad web.

3.2.1. Marco teórico

En el contexto de las Tecnologías de la Información y la Comunicación (TIC), la accesibilidad se define como un “arte” mediante el cual se puede garantizar que cualquier recurso, se encuentre disponible para cualquier persona, tenga o no alguna discapacidad [12, 77]. Existen discapacidades que limitan el uso de las TIC, principalmente encontramos la discapacidad motora, de uno o varios miembros, que imposibilita el uso de dispositivos y, la discapacidad , que dificulta la percepción de la información [78].

La accesibilidad informática permite a una persona que presente alguna limitación física llegar hasta él, así como alcanzarlo y utilizarlo para su propio bienestar [79]. En el año 2003, España fue el primer país que publicó unas normas definidas que contiene los requisitos necesarios y así garantizar la accesibilidad para personas con discapacidad y/o mayores [80].

Según Abril [81], es necesario seguir criterios de diseño orientados en la línea del **Diseño**

para Todos o **Diseño Universal**, para dar el beneficio tecnológico a las personas con diversidad funcional.

Finalmente, en 1982 se aprobó la Ley de Integración Social de los Minusválidos; para el año 1987, comenzó el proyecto COST 219 (*European Cooperation in the Field of Scientific and Technical Research*), hasta que en el año 1989 se creó el CEAPAT, centro que se encargó de llevar numerosas iniciativas, proyectos y propuestas en los diferentes ámbitos de la accesibilidad [82]. Como apoyo, se crearon ayudas técnicas tales como productos y tecnologías para los discapacitados [83], además, apoyos para los diseñadores y desarrolladores tecnológicos como pautas, técnicas y recomendaciones de accesibilidad; como referencia encontramos al europeo y estándar de facto, conformadas por la *Web Content Accessibility Guidelines 1.0* (WCAG.1.0) de la *Web Accessibility Initiative* (WAI) organismo dependiente del *World Wide Web Consortium* (W3C). Estas pautas describen los obstáculos que las personas con discapacidad física, visual, auditiva y/o cognitiva/neurológica pueden hallar en las páginas Web [84].

Una alternativa de permitir múltiples formatos en la recepción de información, es con el uso de las diferentes tecnologías de apoyo: auditiva, táctil, de amplificación, encontrando recursos hardware y software [81].

Para hardware, encontramos el monitor o pantalla; la línea Braille, que es un aparato externo para ser usado con el software de revisión de pantalla; el guión de voz, para lograr acceder a la información ubicada en la pantalla del ordenador [85]. En cuanto al software, para las personas con dificultad visual, hay programas amplificadores de pantalla así como programas conversadores texto-voz. Por otro lado, se tienen los navegadores parlantes para ayudar con la navegación por internet, para las personas con dificultad visual y/o con discapacidad motora [86].

Existe una gran cantidad de estudiantes con alguna discapacidad, ellos tienen derecho a la educación en todos los niveles, lo que implica que los entornos educativos, en especial los sistemas, sean accesibles para ellos. En general, las personas con algunas discapacidad, presentan barreras actitudinales, ambientales y técnicas [87].

Las barreras tecnológicas o electrónicas surgen cuando la tecnología no se puede convertir a otro formato al que se pueda acceder mediante dispositivos de asistencia, lo que tiene un impacto directo en su desarrollo académico.

El mejorar la accesibilidad, en los espacios académicos, incluye la capacidad de acceder a información del aula y recursos electrónicos a través del sitio web de la institución de educación superior. También que estén disponibles las instalaciones del campus, como el laboratorio de computación y la biblioteca.

Por lo anterior, en la Convención sobre los derechos de las personas con discapacidad de Naciones Unidas, se enumeran en sus diferentes artículos sobre la importancia que tiene la Accesibilidad [88]. Asimismo, la UNE 139803:2012 Requisitos de Accesibilidad para Contenidos en la Web [89], es una normativa dirigida a quienes desarrollan programas informáticos, diseñadores de los sistemas operativos o de páginas Web y fabricantes de ordenadores (periféricos).

La normativa de accesibilidad es fundamental en la calidad de los productos y servicios, para que al aplicarlas se eliminen las dificultades de acceso a la información.

3.2.2. Normatividad

Dentro de la normatividad, específicamente para Internet, el World Wide Web Consortium (W3C) [84] describe una serie de pautas de accesibilidad en las páginas Web, las que se deben cumplir para que estas puedan ser accesibles para todas las personas creando nuevas oportunidades para las personas con discapacidad. En las siguientes subsecciones se describen las dos normativas más importantes para la accesibilidad según W3C.

3.2.2.1. Normativas ATAG

El W3C es la comunidad internacional que trabaja en el desarrollo de protocolos Web comunes, su misión principal es guiar la Web hacia su máximo potencial, a través del desarrollo de protocolos y pautas que aseguren el crecimiento del futuro de la Web [90]. Las Pautas de Accesibilidad para Herramientas de Autor (*ATAG*, por sus siglas en inglés) proporcionan pautas para diseñar herramientas de auditoría de contenido web que sean mucho más accesibles para los autores con discapacidades. Estos software o servicios fueron creados por autores (desarrolladores web, escritores, diseñadores, etc.) quienes producen contenido web (páginas web estáticas, aplicaciones web dinámicas, etc.) cuya función es hacer que las herramientas de creación sean accesibles, es decir, que las personas con

discapacidad puedan crear contenido web [91, 92].

3.2.2.2. Normativas de accesibilidad WCAG 2.2

Las Pautas de Accesibilidad para el Contenido Web (WCAG) [93] presentan recomendaciones para diseñar contenido Web más accesible [94]. Actualmente, las *Web Content Accessibility Guidelines 2.2* (WCAG 2.2) son el marco de referencia. Las Pautas de Accesibilidad al Contenido Web (WCAG) clasifican los criterios de éxito en tres niveles progresivos de dificultad: A, AA y AAA (ver figura 3.2).

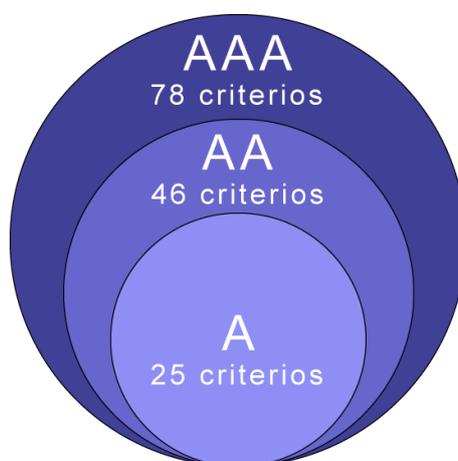


Figura 3.2: Niveles de cumplimiento para A, AA y AAA, según el estándar WCAG 2.2. Fuente: Elaboración propia.

Para España, a través de la Norma UNE 139803:2012 [89], los contenidos Web tendrán que cumplir con el nivel de adecuación AA de las WCAG 2.0 [89]. De esta manera, las WCAG 2.0 eran más una recomendación que una norma en sí misma, ahora por ley las pautas que se debían cumplir son las WCAG 2.0 aprobada como estándar internacional ISO/IEC 40500:2012 [95]. Para alcanzar los estándares esperados, es necesario llevar a cabo evaluaciones de la accesibilidad web.

3.2.3. Evaluación de sitios accesibles

Knighly [96] propuso cuatro tipos de evaluación: 1) Evaluación del esfuerzo, 2) Evaluación del proceso, 3) Evaluación de la eficacia y 4) Evaluación del impacto. Se conocen diversas propuestas y metodologías de evaluación de recursos digitales, a nivel global existen parámetros e indicadores para la evaluación de sitios web en diferentes ámbitos requeridos. Codina (como se citó en [96]), establece seis parámetros definidos como propiedades o

características de los recursos digitales que serán objeto de evaluación. Aun cuando se ha querido homogeneizar criterios, la discusión académica continúa tanto desde la óptica de autores independientes, como de la Organización Internacional para la Normalización.

Como se menciona en párrafos anteriores, la WAI es un organismo que pertenece al W3C para promover la asistencia entre las organizaciones de todo el mundo, a fin de crear estrategias para la creación de recursos y herramientas para que las Webs sean más accesibles [94].

López et al [97], comentan que la accesibilidad Web debe garantizar que los portales de la Internet logren ser visitados y manipulados satisfactoriamente, es indispensable valorar sus contenidos usando las herramientas apropiadas.

Para identificar que el contenido Web sea accesible, puede ser de forma manual, a través de una verificación del funcionamiento del sitio Web, con la intervención de un experto [98], o de manera automática, usando una aplicación informática. Además, en Internet encontraremos alrededor de 90 herramientas validadas que hacen una evaluación automática de accesibilidad a un sitio Web. Esta acción, debe ser complementada con una revisión manual, sin obviar la evaluación del código HTML, CSS y XHTML, que evita errores, en lo que respecta a la distinguibilidad, adaptabilidad, legibilidad y navegación [78]. Por lo tanto, se concluye que para evaluar la accesibilidad Web se necesita una selección cuidadosa de las herramientas para realizar esta tarea, sin olvidar tener como referencia el número de criterios que abarca, con relación a las pautas WCAG 2.2 dictadas por el W3C [97].

El Wave Accessibility Tool¹ es una aplicación en línea que se emplea para evaluar la accesibilidad de los sitios web conforme a las directrices establecidas por la W3C, especialmente las Directrices de Accesibilidad al Contenido Web (WCAG). Su propósito principal es examinar las páginas web en busca de posibles obstáculos de accesibilidad que podrían dificultar la experiencia de una amplia variedad de usuarios, incluyendo aquellos con discapacidades visuales, auditivas, motoras y cognitivas. Esta herramienta ofrece diversas funcionalidades, como la detección de errores habituales de accesibilidad, como imágenes carentes de texto alternativo, estructuras de documentos inapropiadas y problemas de contraste de color. Además, proporciona recomendaciones y sugerencias para

¹<https://wave.webaim.org/>

subsanan estos inconvenientes y mejorar así la accesibilidad del sitio web (ver figura 3.3).

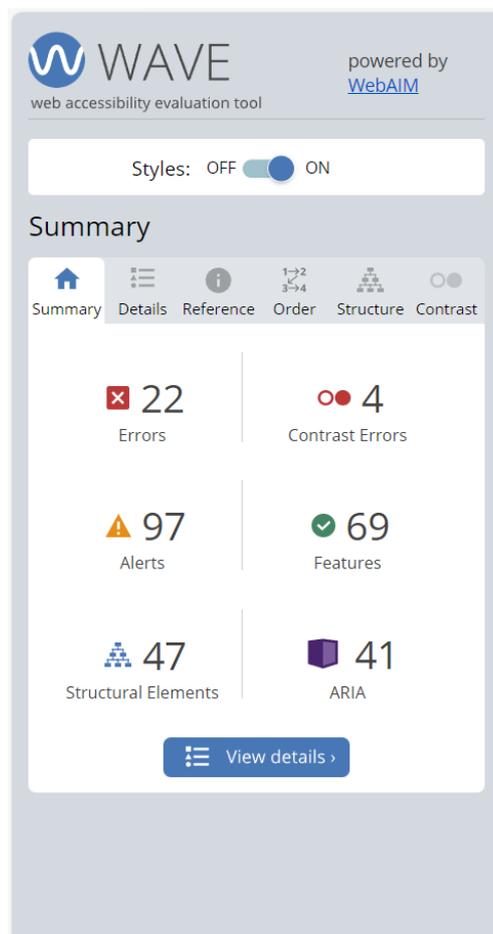


Figura 3.3: Resumen de resultados de la herramienta Wave al evaluar la página principal de un sitio web. Fuente: Elaboración propia.

Finalmente, es importante resaltar que la accesibilidad ha sido agregada como atributo de la usabilidad. Además, Rubin & Chisnell [99], proponen trabajar de manera paralela la usabilidad y la accesibilidad, tal que un producto sea usable para personas con discapacidad.

3.3. Usabilidad

En esta sección se explican brevemente las disciplinas para la evaluación de la Interacción Humano – Ordenador (HCI, por sus siglas en Inglés), se mencionan definiciones del término usabilidad, además, se detallan métodos y técnicas actuales para su evaluación incluyendo algunos cuestionarios estandarizados para medir la satisfacción del usuario.

3.3.1. Marco teórico

La usabilidad o *usability*, es reconocida como un atributo que proviene de la *Human – Computer Interaction* (HCI), ésta última surge a finales de la década de 1980 como un campo de investigación que trata acerca de la interacción que se da entre las personas y las computadoras [100] en la correcta consecución de las tareas definidas para sistemas y/o aplicaciones. Este es un campo en evolución debido a los rápidos cambios en la tecnología [21]. Además es un campo interdisciplinario, ya que incorpora a la psicología cognitiva, antropología, ingeniería industrial, las ciencias del comportamiento, sociología, las ciencias computacionales, las comunicaciones, entre otras [100]. La tabla 3.1, describe las disciplinas de HCI que abordan a la usabilidad.

Disciplina del HCI	Descripción	Relación con la Usabilidad
Diseño de Interfaces (IxD) [101]	Se centra en la creación de la apariencia visual y la disposición de los elementos en una interfaz de usuario.	Busca garantizar que la disposición y el diseño de la interfaz mejoren la experiencia del usuario, haciéndola más intuitiva y fácil de usar.
Interfaz Gráfica de Usuario (GUI) [102]	Diseño de interfaces que permiten la interacción con un sistema utilizando elementos gráficos como botones, menús y ventanas.	La GUI se centra en hacer que la interacción con el sistema sea más visual y accesible, lo que contribuye a una experiencia de usuario más satisfactoria y eficiente.
Experiencia de Usuario (UX) [103]	Se refiere a la experiencia general de un usuario al interactuar con un sistema, incluyendo aspectos emocionales, sensoriales y prácticos.	La UX busca garantizar que el diseño de la interfaz y la interacción con el sistema sean agradables, útiles y satisfactorios para el usuario, lo que implica una alta usabilidad.
Diseño Centrado en el Usuario (UCD) [104]	Enfoque de diseño que coloca las necesidades y deseos del usuario en el centro del proceso de desarrollo.	El UCD busca garantizar que el diseño del sistema sea orientado y validado por los usuarios, lo que conduce a productos finales que son altamente usables y adaptados a las necesidades del usuario.

Tabla 3.1: Disciplinas del HCI con énfasis en la usabilidad.

Dentro de el enfoque UCD, está la metodología de DT, que como se mencionó en la sección 3.1, se centra en comprender las necesidades y desafíos de los usuarios a través de la investigación y la empatía, así como desarrollar soluciones innovadoras y empáticas para satisfacer esas necesidades. Por otro lado, la usabilidad se enfoca en evaluar y mejorar la

calidad de uso de un producto o servicio, asegurándose de que sea intuitivo, fácil de usar y eficiente para los usuarios. Al incorporar pruebas de usabilidad en el proceso de DT, se pueden identificar y corregir problemas de usabilidad antes de la implementación, lo que resulta en soluciones más efectivas y eficientes.

Por su parte, Jacob Nielsen, considerado el padre de la usabilidad, extendió la definición a un “atributo relacionado con la facilidad de uso para la aceptación práctica y rentable del producto”. Además, en lo que respecta a software, la usabilidad es un atributo de calidad para medir cuán fáciles de usar son los sitios web [105]. Nielsen [106] propuso cinco componentes:

1. **Capacidad de aprendizaje:** Se refiere a la facilidad de aprender la funcionalidad y el comportamiento del sistema.
2. **Eficiencia:** Se refiere a el nivel de productividad alcanzable, una vez que el usuario ha aprendido el sistema.
3. **Capacidad de memorizar:** Se refiere a la facilidad de recordar la funcionalidad del sistema, de modo que el usuario ocasional pueda volver al sistema después de un período de inactividad, sin necesidad de volver a aprender a usarlo.
4. **Pocos errores:** Se refiere a la capacidad del sistema para presentar una tasa de error baja, ayudar a los usuarios que cometen pocos errores durante el uso del sistema y, en caso de que cometan errores, ayudarlos a recuperarse fácilmente.
5. **Satisfacción de los usuarios** Se refiere a la medida en que el usuario encuentra agradable el uso del sistema.

La ISO/IEC 9126 [107] refiere que la usabilidad es “la capacidad de un software de ser comprendido, aprendido, usado y ser atractivo para el usuario”, la ISO/IEC 9241 la define como “la eficiencia y satisfacción con la que un producto permite alcanzar objetivos específicos a usuarios específicos”, esta se centra en la calidad de uso, tiene más que ver con la manera en que el usuario hace sus tareas de forma específica con efectividad [105].

Por su parte, la norma ISO 9241-210 define la experiencia del usuario (UX, por sus siglas en inglés) como “las percepciones y respuestas de un individuo que resultan del uso

o el uso anticipado de un producto, servicio o sistema” [108]. Su finalidad es la de englobar aquellos aspectos deseables referentes al diseño de interacción que los ingenieros pueden construir [109].

La usabilidad también se refiere al conjunto de métodos y técnicas con la finalidad de mejorar esa facilidad de uso en el proceso de diseño, trata de que el usuario logre encontrar lo que busca dentro del menor tiempo posible, sin dejar de lado la estética y el contenido, como el ser novedoso, inteligente, entendible, atractivo y/o comprensible [105]. Asimismo, es importante considerar las emociones que emergen de los usuarios al realizar tareas o proceso [110].

El término de usabilidad se ha definido en entornos académicos, además, la usabilidad es clave para permitir la comunicación entre la universidad y sus usuarios. Los sitios web universitarios se consideran como puertas virtuales, tanto para los estudiantes como para otros usuarios de todo el mundo.

También existen ciertos problemas de usabilidad en los entornos de sitios web académicos, los cuales tienen impacto sobre el éxito y fracaso de los sistemas de software, tales como los costos asociados a problemas del usuario con el software [111].

3.3.2. Normativas ISO

La ISO/IEC 25000 [112], mejor conocida como SQuaRE (*System and Software Quality Requirements and Evaluation*), es un conjunto de normas con el objetivo de crear un marco de trabajo común para evaluar la calidad de un producto software, estas normas son el resultado de la evolución de otras normas, como las ISO/IEC 9126 e ISO/IEC 14598 [113]. Las normas ISO/IEC 25000 se componen de cinco divisiones (ver figura 3.4). Cada una de estas divisiones tiene un propósito y a su vez contienen determinadas características.

En ISO 25000, el término de usabilidad dentro de esta norma, la podemos encontrar en la ISO 25010, entendida como la capacidad que tiene un producto de software para entenderse, usarse, aprenderse y ser atractiva para los usuarios; dicha característica se subdivide en seis sub características:

- **Inteligibilidad o reconocibilidad de la adecuación:** Se encarga de que el usuario identifique si el sistema es adecuado para sus necesidades, y si cumple con



Figura 3.4: Norma ISO/IEC 25000 y sus divisiones. Fuente: Elaboración propia.

los requerimientos para satisfacer sus expectativas.

- **Aprendizabilidad:** Se refiere a la capacidad de un software para permitir que el usuario aprenda sobre su utilización. Se basa en la adquisición de nuevos conocimientos para interactuar con el sistema.
- **Operabilidad:** Se basa en la capacidad del software para que el usuario pueda controlar fácilmente el sistema, así como evaluar la eficiencia y eficacia con la que se puede utilizar el sistema.
- **Protección contra errores de usuario:** Se relaciona con la capacidad que tiene el software para evitar o proteger a usuarios de errores o, en su defecto, minimizar las consecuencias de los mismos.
- **Estética de la interfaz de usuario:** Tiene que ver con la posibilidad de que un sistema sea agradable y logre la satisfacción del usuario. Se relaciona con la

aparición visual y el diseño.

- **Accesibilidad:** Esta dimensión se basa en la posibilidad de un software de ser utilizado por usuarios con distintas capacidades o condiciones, como lo sería una discapacidad.

Todas las normas, previamente desarrolladas, así como sus características, aportan en la adecuada aplicación de la usabilidad.

3.3.3. Evaluación de la usabilidad

La usabilidad debe ser evaluada con la finalidad de cubrir los requisitos solicitados en la norma ISO 25000. Como mencionan Mex-Álvarez et al. [114], la usabilidad de un sistema o producto en esta norma puede verse desde dos perspectivas: la del software (o producto) y la del uso desde la perspectiva del usuario (ver figura 3.5).

Es así, que la usabilidad puede evaluarse desde un modelo de calidad del producto, compuesto por ocho características, que se relaciona con las propiedades estáticas de un software y a las dinámicas del sistema de una computadora, y puede aplicarse tanto a un producto de software como a un sistema; o desde un modelo de calidad en uso, que se compone de cinco características, que resulta de una interacción al utilizar un producto en un contexto particular, y aplicarse a cualquier sistema humano-computadora [114].

3.3.3.1. Métodos de evaluación de la usabilidad

Los métodos de evaluación de la usabilidad son técnicas para recolectar información y percepciones de los usuarios acerca de la interacción entre los usuarios y la interfaz del SI. La evaluación de la usabilidad puede ser clasificada en tres grupos: (1) inspección, (2) indagación y (3) test. Estas categorías se describen en las siguientes subsecciones.

- **Inspección**

En este método, los expertos tienen el conocimiento necesario para medir el grado de usabilidad de un SI probando la interfaz de usuario. Este tipo de método da buenos resultados en las primeras etapas del diseño de la interfaz [115]. A continuación, se presenta una lista de técnicas correspondientes al método de inspección:

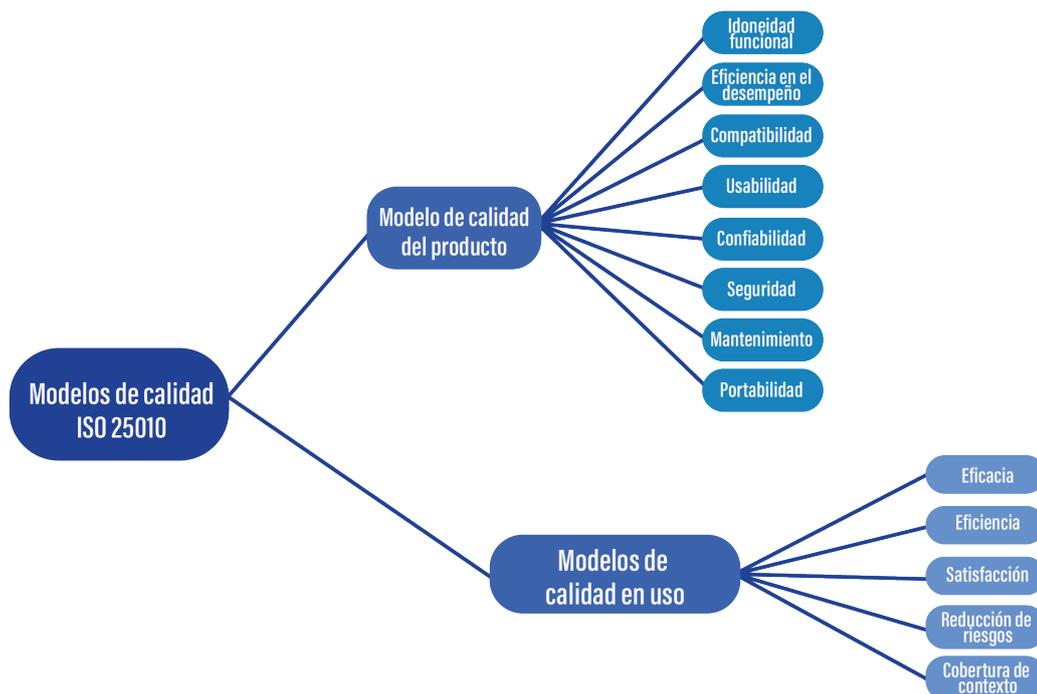


Figura 3.5: Estructura de los modelos de calidad en la norma ISO 25010. Adaptado de “Modelos de Calidad de la Norma ISO 25010” (p. 4). Fuente [114].

- **Evaluación heurística:** Nielsen y Molich utilizan principios heurísticos o heurística de usabilidad para evaluar la usabilidad en el análisis de la interfaz de los SI [11, 115].
- **Recorrido cognitivo (*Cognitive Walkthrough*):** Se centra en la evaluación de la facilidad de aprendizaje de un SI. Identifica el pensamiento y el comportamiento del usuario cuando utiliza un SI por primera vez. Sin embargo, la fidelidad de los resultados puede ser baja, además, esta técnica se utiliza en las primeras fases del diseño de los SI para obtener una lluvia de ideas de los usuarios [116, 117, 118].
- **Clasificación de técnicas de inspección:** Se denomina inspecciones formales de usabilidad, dentro de estas se pueden considerar las siguientes categorías [118, 119]:
 1. Inspección de estándares, se centra definir los estándares a evaluar y examina las acciones para ejecutar y realizar una tarea, así como crear una historia del por qué un usuario hará una acción.

2. Inspección formal o de código de usabilidad también conocida como inspección del software de inicio en IBM. Consiste en examinar el software para identificar los defectos del mismo. Esta técnica comprueba y documenta los defectos encontrados.

- **Checklist.** Es un conjunto predefinido de tareas, características y/o elementos de las interfaces de los Sistemas de la Información que deben ser evaluados por diferentes perfiles de usuarios para detectar defectos de inspección.
- **Observación de campo.** Comprende la interacción entre los usuarios y los SI a través de la observación del uso del sistema mientras se realizan las tareas diarias [115]. Tiene mayor eficacia en las primeras fases de desarrollo del software.
- **Grupos focales.** Esta técnica recoge datos con grupos de personas mediante el uso de preguntas sencillas que llevarán la sesión hacia la discusión y el diálogo, permite recoger opiniones e ideas espontáneas de los usuarios [15]. El nivel de satisfacción puede obtenerse al final de la fase de desarrollo.
- **Entrevistas.** Esta técnica exploratoria mide el grado de satisfacción de los usuarios en el uso del SI. Además, se considera una herramienta de recogida de datos cualitativa.
- **Inicio de sesión (logging) o análisis de logs.** Se basa en recoger los registros detalladamente del sistema todas las acciones cuando el usuario realiza actividades. Tiene resultados positivos cuando las personas que evalúan se encuentran en espacios físicos diferentes o hay suficientes personas para realizar el test. Como desventaja, no obtiene información cualitativa.

■ Indagación

Una estrategia que afecta directamente al análisis de usabilidad es el llamado método de indagación. Se centra en la atenta observación a los usuarios que hacen uso del sistema en las actividades cotidianas. Es habitual que los usuarios se planteen algunas preguntas como parte de la interacción con el sistema. Las técnicas más importantes del método de indagación mediante preguntas son los cuestionarios estandarizados. Se proporciona mas detalle sobre los cuestionarios estandarizados en la sección 3.3.3.2.

■ Test

Este método se centra en la obtención de la información a partir de pruebas realizadas a usuarios reales [120]. El principal objetivo es corregir errores que pudieron ser omitidos por expertos en usabilidad. Esas pruebas se hacen posteriores a otras técnicas y al ser en tiempo real con usuarios, tienen un alto grado de fiabilidad. Los test más utilizados son:

- **Pensando en voz alta** [121]: Se pide a los usuarios que utilicen un sistema en voz alta para describir sus pensamientos, impresiones, sentimientos y acciones en tiempo real. La verbalización brinda información sobre su experiencia sobre el sistema.
- **Protocolo rápido** [122]: Esta técnica es un prototipo de baja fidelidad. Utiliza prototipos de materiales como papel, cartón, espuma y arcilla e incluyen componentes básicos de diseño como botones, menús y pantallas. La información se utiliza para refinar el prototipo. El proceso se puede repetir varias veces hasta lograr el diseño óptimo en términos de usabilidad.
- **Registro visual** [123, 124]: Registra la interacción del usuario con la pantalla mientras usa un producto o sistema. Brinda información en tiempo real sobre el comportamiento del usuario para analizar patrones de uso e identificar problemas de usabilidad. Es útil para sistemas con interacciones complejas y numerosos elementos de interfaz de usuario. Permite mejorar su desempeño y desarrollar soluciones innovadoras, lo que da como resultado una mejor experiencia de usuario.
- **Clasificación de tarjetas** [125]: Comprende cómo los usuarios organizan la información en las interfaces gráficas se les presentan una serie de fichas con información y se les solicita que las organicen según unos criterios determinados. Así, evalúan los patrones de pensamiento y las preferencias de sus usuarios para organizar la información.
- **Pruebas de usabilidad** [126]: Observa a los usuarios interactuando con un producto o sistema con el objetivo de identificar problemas de usabilidad y mejorar la experiencia del usuario. Se pide a los usuarios que realicen una serie de tareas específicas y se registran sus interacciones y respuestas. Luego, se

identifican problemas de usabilidad y desarrollan soluciones efectivas basadas en una evaluación objetiva y permiten mejoras en el diseño y la funcionalidad del producto o sistema.

Además de los métodos de inspección, indagación y tests, existen los cuestionarios estandarizados para evaluar la usabilidad.

3.3.3.2. Cuestionarios estandarizados

Un cuestionario estandarizado se refiere a un conjunto de preguntas diseñadas de forma específica con la finalidad de recopilar información de manera sistemática y objetiva. Los cuestionarios son diseñados para ser aplicados de forma igualitaria a todos los participantes, para lograr la comparabilidad y confiabilidad de los resultados [127]. Entre los cuestionarios estandarizados comunes para evaluar la usabilidad se encuentran los siguientes:

- **Cuestionario SUS**

System Usability Scale (SUS) fue desarrollada por John Brooke en 1996 [31, 128]. El SUS tiene 3 características principales: 1) se compone de sólo 10 ítems con una mezcla de ítems positivos y negativos que es simple y rápido de responder por los participantes (ver tabla 3.2), 2) es un cuestionario de distribución libre y de código abierto y, 3) es aplicable para la evaluación de la usabilidad de diferentes tipos de GUI como sistemas web, móviles o TV's.

Para obtener la puntuación SUS, cada ítem se evalúa en un rango de 0 a 4. Para los ítems 1,3,5,7,9 se resta 1 a la posición de cada ítem. Para los ítems 2,4,6,8,10, se resta el valor 5 de la posición evaluada de los ítems. Posteriormente, se suma el resultado de las dos operaciones anteriores y se multiplica por 2,5. El resultado final se calcula por la media de todos los juicios de los expertos y se pondera en un rango 0 - 100.

- **Cuestionario NPS**

El Net Promoter Score (NPS) [99] se utiliza para evaluar los servicios de empresas o instituciones, utilizando una pregunta única que mide la fidelidad de los usuarios y calcula la probabilidad de que estos actúen como promotores del mismo sistema.

N°	Pregunta	1	2	3	4	5
1	Me gustaría utilizar este sistema más a menudo.					
2	Me parece que este sistema es más complicado de lo que debería ser.					
3	Creo que el sistema es sencillo y fácil de usar.					
4	Necesito apoyo técnico para utilizar este sistema.					
5	Creo que el sistema funciona bien y está bien integrado					
6	Creo que hay muchas irregularidades en el sistema.					
7	Creo que la mayoría de la gente puede aprender este sistema rápidamente.					
8	Creo que este sistema requiere mucho tiempo.					
9	Me siento seguro al utilizar este sistema.					
10	Creo que hay muchas cosas que aprender antes de poder empezar a utilizar este sistema.					

Tabla 3.2: Cuestionario SUS en español.

En este sentido, los usuarios se clasifican de la siguiente manera: 1) Promotores: aquellos que definitivamente recomendarían el servicio y evalúan el servicio entre ocho y diez; 2) Pasivos: son aquellos que están satisfechos pero no recomendarían el servicio a otra persona, su evaluación está entre cinco y ocho; y 3) Detractores: son aquellos que están insatisfechos con el servicio y desanimarían a otros de hacer uso del servicio [99, 129].

Diferentes estudios hacen uso de este test debido a la facilidad de aplicación [130, 131]. La puntuación es un número porcentual que se calcula de la siguiente manera: $N = P - D$.

■ Cuestionario QUIS

El *Questionnaire for User Interface Satisfaction* (QUIS), compuesto por 89 ítems, está diseñado para medir la satisfacción subjetiva del usuario con la interfaz del ordenador [132]. QUIS 7.0 tiene como objetivo medir 4 aspectos principales de las interfaces de usuario: 1) factores de la pantalla, 2) terminología y retroalimentación del sistema, 3) factores de aprendizaje y 4) capacidades del sistema. Algunos trabajos de investigación han utilizado este cuestionario para evaluar las interfaces del sistema [133, 134]. Por lo tanto, la usabilidad, como atributo, brinda un mejor acceso a los usuarios finales, provocando un impacto positivo en el uso de los sistemas, en particular, en los ámbitos educativos.

3.4. Resumen de capítulo

En este capítulo hemos visto que:

- DT Permite diseñar soluciones que respondan a las necesidades reales de los usuarios para garantizar la satisfacción del usuario y su experiencia óptima.
- DT puede reducir los costos de desarrollo y mantenimiento debido a la detección temprana de las áreas de mejora y acercamiento al usuario final al prototipo final al facilitar la comunicación entre el equipo de desarrollo y las personas.
- Una metodología centrada en el usuario ayuda a identificar los puntos débiles clave en un producto o servicio para que se puedan realizar ajustes para mejorar su usabilidad.
- UCD fomenta la innovación y la creatividad, ya que los usuarios participan en el proceso de diseño y los comentarios de los usuarios se tienen en cuenta para mejorar y optimizar los productos y servicios. Por tanto, asegura una experiencia de usuario satisfactoria y aumenta la probabilidad de éxito de nuestros productos y servicios.
- La accesibilidad web se refiere a la capacidad que tienen los sitios web para que las personas con limitaciones físicas o cognitivas accedan a los contenidos y servicios en línea. Es importante evaluar la accesibilidad web para garantizar que los sitios de Internet sean visitados y manipulados satisfactoriamente por la mayoría de las personas, incluyendo aquellas con limitaciones físicas o cognitivas.
- La evaluación de la accesibilidad web permite identificar obstáculos que impiden el acceso a estos contenidos y servicios, lo que permite a los desarrolladores corregir estos problemas y mejorar la experiencia de usuario para todos.
- Evaluar la usabilidad web ayuda a identificar dificultades como contenido confuso o poco claro, problemas de carga, problemas de interacción y respuesta, lo que ocasiona que los usuarios se sientan desesperados o descontentos con su experiencia.
- El cuestionario SUS ha demostrado ser altamente confiable y sensible a los cambios en la usabilidad. Asimismo, proporciona una evaluación rápida y confiable de la usabilidad de un sistema, permitiendo recopilar datos cuantitativos sobre la percepción del usuario en términos de facilidad de uso y satisfacción.

- A través NPS se recopila información de la percepción del usuario, incluida la satisfacción del usuario, la probabilidad de recomendación del producto y las percepciones del diseño y la calidad de la experiencia del usuario en general.
- Al diseñar contenidos de páginas web teniendo en cuenta los aspectos mencionados en este capítulo, se puede mejorar la experiencia del usuario y aumentar la frecuencia con la que los usuarios regresan al sitio.

Capítulo 4

Diseño de un cuestionario para la evaluación de la usabilidad

La validación de un cuestionario es un paso crucial en cualquier investigación, ya que garantiza la fiabilidad y la validez de los datos recopilados. Al validar un cuestionario, se busca confirmar que las preguntas sean claras, pertinentes y que midan con precisión las variables de interés. Una validación adecuada asegura que los resultados obtenidos reflejen de manera precisa la realidad que se está estudiando, lo que aumenta la credibilidad y la robustez de los hallazgos.

Este capítulo se divide en las siguientes secciones: La sección 4.1, describe las características válidas que debe contener un cuestionario de investigación. La sección 4.2, detalla los pasos para construir un cuestionario. La sección 4.3, expone los pasos para realizar la validación de contenido mediante un modelo de toma de decisiones con enfoque lingüístico. La sección 4.4, describe la validación de confiabilidad y la versión final del cuestionario para la evaluación de la usabilidad. Finalmente, la sección 4.5 presenta los puntos principales de este capítulo.

4.1. Aspectos deseables para la calidad de un cuestionario

Existen diversos instrumentos para la recolección de datos en el área de investigación, los cuales son la base del proyecto.

Para diseñar un instrumento de investigación adecuado para la recolección de datos es necesario definir el propósito, identificar a la población de interés, diseñar las preguntas o tareas de forma lógica, formular preguntas de control, realizar pruebas piloto, entre otras, con la finalidad de ajustar dicho instrumento.

Además, dado que en la investigación es relevante la calidad de los instrumentos de recolección de datos, se busca que la realización del mismo sea metódica, garantizando su confiabilidad, objetividad y validez.

1. La confiabilidad tiene como finalidad que los datos obtenidos sean fiables, exactos y consistentes. Se utiliza el alfa de Cronbach para realizar la estimación estadística [135], al igual que el coeficiente de correlación de Pearson [136].
2. La objetividad evalúa si la medición está libre de sesgos, los cuales se pueden deber a errores en la administración, calificación e interpretación del instrumento de medición [25].
3. La validez, conocida como la suficiencia de un instrumento para medir lo que desea medir, se divide en tres aspectos:
 - a) Validez de criterio, basada en la capacidad del instrumento de predecir los resultados, está relacionada con un resultado esperado o conocido [137].
 - b) Validez de constructo, enfocada a que las preguntas, o tareas del instrumento, se fundamenten en los conceptos teóricos o dimensiones examinadas mediante la prueba KMO [138] y la prueba de esfericidad de Barlett [139].
 - c) Validez de contenido, la cual se realiza con la participación de jueces y evaluación estadística referente a la comprensión de las preguntas, o tareas, y la modificación de las mismas [140, 141].

Existen diversos pasos para la validación de contenido de los instrumentos de medición en el área de investigación, los cuales se basan en la validación por parte de jueces, los que son identificados por expertos en el tema, con la finalidad de llegar a un consenso con respecto a las características del instrumento. A continuación, se mencionan los métodos mayormente utilizados para la validación [142]: Método de Agregados individuales [143], Técnica de Grupo Nominal [144], Método del Consenso Grupal [144], Validez de Contenido (IVC) [145], y Método Delphi [146], éste último reconocido como uno de los métodos de validación más utilizados en el área de investigación.

Por lo tanto, todo instrumento, debe pasar por un proceso de validación, utilizando diversas técnicas y herramientas de soporte de decisión actualizadas.

4.2. Procedimiento para el diseño de un cuestionario

El diseño de un test válido es un proceso crucial para garantizar que las mediciones realizadas sean confiables y significativas. Los pasos generales, propuestos por diversos autores [144], para diseñar un test válido son:

- **Definir el objetivo del test:** Se debe tener claro qué se quiere medir y con qué propósito, para guiar correcta realización de los ítems.
- **Revisión de literatura:** Es fundamental realizar una revisión de la literatura existente sobre el tema a evaluar, para conocer las dimensiones a incluir.
- **Desarrollo de los ítems:** Con base en la literatura, se elabora los ítems del cuestionario. Estos deben ser claros, relevantes para el objetivo del test y estar redactados de manera que sean comprensibles para los participantes.
- **Consulta de expertos y validación de contenido:** Se debe consultar a expertos en el área para comprender qué elementos y conceptos se deben incluir en el cuestionario. El proceso implica asegurarse de que los ítems representan de manera precisa el dominio que se está evaluando. El método Delphi [147] es de los más utilizados para lograr el consenso entre expertos.
- **Prueba piloto:** Se realiza una prueba piloto del test con una muestra pequeña

de participantes. Esto permite identificar posibles problemas con los ítems, como ambigüedad o dificultad de comprensión.

- **Análisis de la confiabilidad:** Se utilizan técnicas estadísticas para analizar la confiabilidad del test, es decir, su consistencia interna. En este caso, se trabaja mediante el alfa de Cronbach.
- **Revisión final:** Con base en los resultados, se revisa y ajusta el cuestionario según sea necesario. Esto puede implicar eliminar ítems, modificar la redacción de preguntas o agregar nuevos ítems.

Así, todos estos pasos ayudan a garantizar que se cumplan los criterios de confiabilidad, validez y objetividad del instrumento.

4.3. Solución Software para la validación de constructo: el modelo lingüístico 2-Tuplas Fuzzy Delphi

La validación de un cuestionario se relaciona con el modelo de Toma de Decisiones Lingüística, dado que tienen como finalidad garantizar la claridad semántica, adaptación lingüística, comprensión del contexto, relevancia cultural y extracción de información. Por ello, utilizar el modelo de Toma de Decisiones Lingüística mejora la comprensión y calidad del lenguaje utilizado en el cuestionario, aportando a su validez y utilidad en la toma de decisiones.

El modelo lingüístico 2-Tuplas Fuzzy Delphi (2TLFD) [25], un sistema de Toma de Decisiones Lingüísticas Multi-Criterio Multi-Experto (MEMCDM, por sus siglas en inglés) para agilizar y facilitar el proceso de consenso entre expertos. Además se acompaña de un *Decision Support System (DSS)* gratuito para facilitar la obtención de los resultados de consenso de cada ítem y una valoración lingüística altamente interpretable. La figura 4.1 detalla el procedimiento para el consenso en 2TLFD.

El método 2TLFD [25] tiene las siguientes características:

- **Definición del cuestionario:** Un cuestionario Q es una sucesión de ítem $I = \{I_1, \dots, I_r, \dots, I_n\}$, $r \in \{1, \dots, n\}$ agrupados por l dimensiones D_i , $i \in \{1, \dots, l\}$.

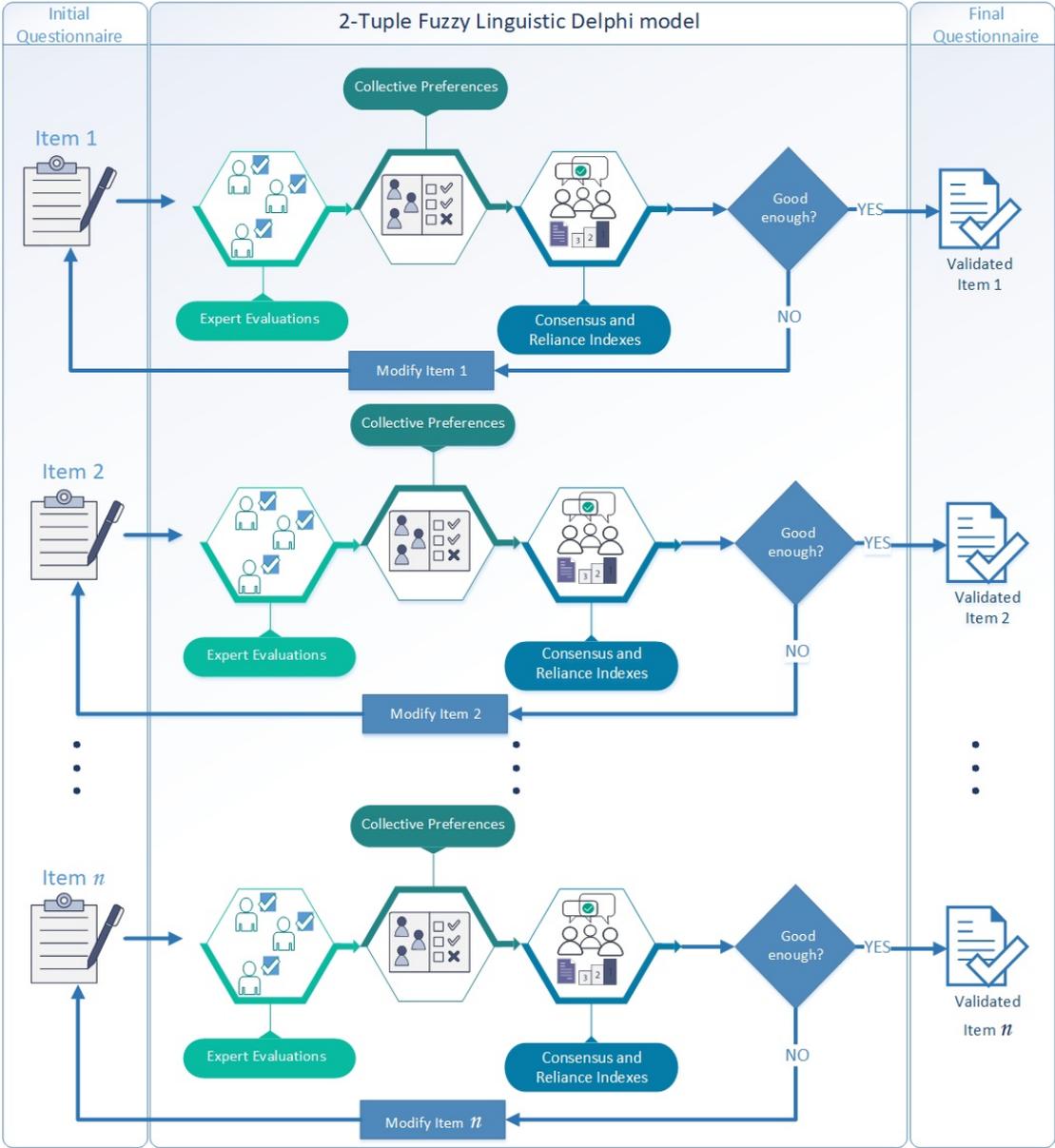


Figura 4.1: Diagrama del método Delphi Lingüístico Difuso de 2-Tuplas. Fuente [25].

- **Criterios:** Cada ítem I_r se evalúa a través de un conjunto de criterios C , donde $C = \{C_1, \dots, C_q\}$.
- **Conjunto de expertos:** 2TFLD pretende llevar al consenso al panel de expertos J , tal que J_i , donde $i = (1, \dots, p)$ es el experto que evalúa el cuestionario Q sobre cada criterio C_q .
- **Importancia en las opiniones de los expertos:** Debido a que cada experto puede tener un grado de experiencia que diste de los demás, el modelo 2TFLD puede asignar diferente importancia a la opinión elegida por cada J_i , siendo la importancia un vector de pesos W_{D_m} tal que $W_{D_m} = w_{1D_m}, \dots, w_{pD_m}$ es un vector normalizado de tamaño p .
- **Conjunto de términos lingüísticos:** Cada experto puede elegir entre los distintos conjuntos de términos lingüísticos $S^{n(t)} \in \{S^3 \cup S^5 \cup S^7\}$ (ver sección 2.2.1) definidos previamente por el grupo o persona que desea validar el consenso. EL modelo 2TFLD asume para que el experto J_i que tiene un alto conocimiento en un campo concreto, es mejor tener un conjunto más rico de términos.
- **Índice de consenso:** Sea el Índice de Consenso $CI_r \in [0, 1]$ el consenso entre expertos del conjunto J respecto al ítem I_r considerando que la opinión de los jueces tiene una ponderación por el vector de pesos normalizados $\{W_{D_1}, \dots, W_{D_i}\}$. Entonces, CS_r es un valor booleano que toma *verdadero* si hay consenso, es decir cuando $CI_r \geq 0,5$ o *falso* en caso contrario.
- **Nivel de confianza:** Sea el nivel de confianza RI , un valor $\epsilon \in [0, 1]$ configurable por el moderador que determina el grado de cercanía entre las opiniones de los expertos J . Sea RI el valor determinado por el moderador para saber si la solución es aceptable. Entonces, RS es un valor booleano que toma *verdadero* si $RI \leq \epsilon$ falso en caso contrario.

El modelo descrito anteriormente se presenta en un *Decision Support System*¹ y presenta las siguientes características:

- **Carga de archivos:** 2TLFD-DSS importa archivos en formato *.cvs, con delimitador ';', cada uno debe contener las siguientes características:

¹<https://sci2s.ugr.es/2tuple-fuzzy-delphi/results.php>

1. **Items Dimension:** En él, se asocia la pericia de cada juez con las dimensiones del cuestionario así como los rangos de los ítems con las dimensiones. Esta importación no es obligatoria, ya que se utilizarían pesos uniformes en caso de ausencia (ver tabla 5 en [25]). Este archivo no es obligatorio para obtener solución del cuestionario Q_i .
2. **Items Responses:** En este archivo, cada fila corresponde a las valoraciones hechas por cada experto J_i , sobre el ítem I_r . Cada valoración realizada para el conjunto de criterios, esta hecha con un término lingüístico en la granularidad elegida en el LTS para el experto J_i . Por ejemplo, $J_1; 7; 6, 6, 6, 6, 1$, el experto 1 eligió evaluar con S^7 el ítem I_1 , dando valores de $s_6^7, s_6^7, s_6^7, s_6^7$ para cada criterio. Finalmente, asigna un valor entre $[0, 1]$ donde determina la importancia que el experto J_1 da al ítem I_1 del cuestionario Q_0 , siendo 0 nada relevante y 1 totalmente relevante (ver tabla 6 en [25]). Este archivo es obligatorio para obtener solución del cuestionario Q_i .
3. **Items Description:** Contiene el texto de descripción por cada pregunta. Esta importación no es obligatoria, ya que se utilizaría un texto genérico en caso de ausencia. La primera fila podría ser el nombre de la cabecera, en cuyo caso n es el número de líneas leídas menos 1 (ver tabla 7 en [25]). Este archivo no es obligatorio para obtener solución del cuestionario Q_i .

Una vez que la carga de archivos se realizo de forma correcta, se muestra una pantalla con las valoraciones obtenidas (ver figura 4.2).

2TFLD-DSS, cuenta con las siguientes herramientas que facilita al moderador la toma de decisiones sobre el cuestionario Q_i .

- **Filtering:** El usuario puede visualizar los datos, incluso ocultar determinadas columnas, seleccionando distintas opciones de filtrado como: toda la información, claridad colectiva, redacción colectiva, presencia colectiva, escala de respuesta colectiva, relevancia media y consenso.
- **Trimming:** El panel de expertos puede aconsejar una reducción del número de elementos. El DSS proporciona una herramienta de recorte que ayuda a resolver la pregunta *¿Qué elementos deben eliminarse para alcanzar el número de elementos*

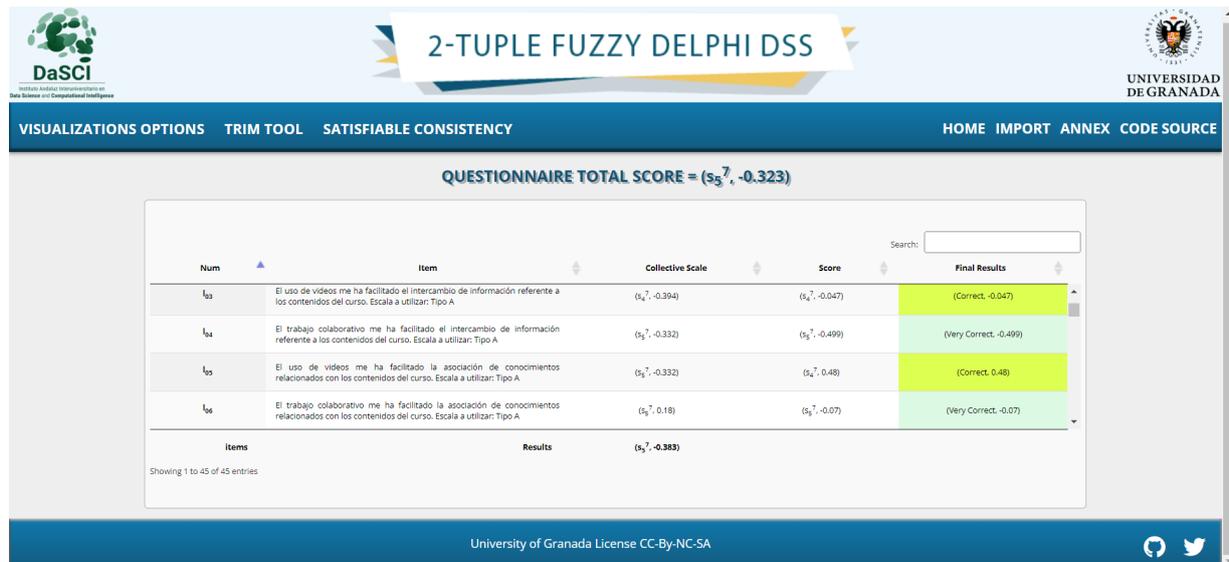


Figura 4.2: Ejemplo de una tabla de respuestas para un cuestionario en 2TLFD-DSS. Fuente: Elaboración propia.

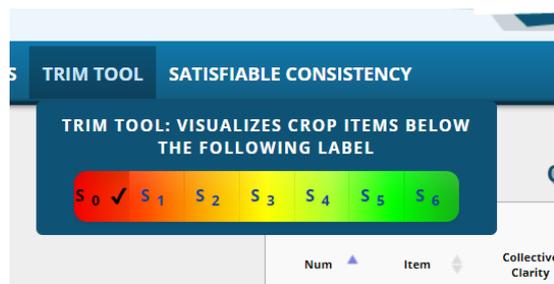


Figura 4.3: Opción de filtrado en 2TLFD-DSS. Fuente: Elaboración propia.

deseado? La herramienta de recorte considera el conjunto de términos lingüísticos S^7 como escala de entrada. Por defecto, se establece en s_0 , lo que significa cero elementos recortados. Al subir el término lingüístico, se ocultan los elementos por debajo de la etiqueta lingüística seleccionada, mostrando las etiquetas iguales o mayores.

- **Data simulation:** El usuario puede ajustar, en la barra de navegación de *satisfiable consistency*, la solución del modelo mediante un parámetro deslizante que permite establecer distintos niveles aceptables de consenso.

4.3.1. Definición del cuestionario

Uno de los elementos fundamentales para el desarrollo y diseño de sistemas web es la usabilidad, esto con la finalidad de garantizar la satisfacción del usuario. Debido a la diversificación de los usuarios, y las mejoras tecnológicas, han aumentado las demandas

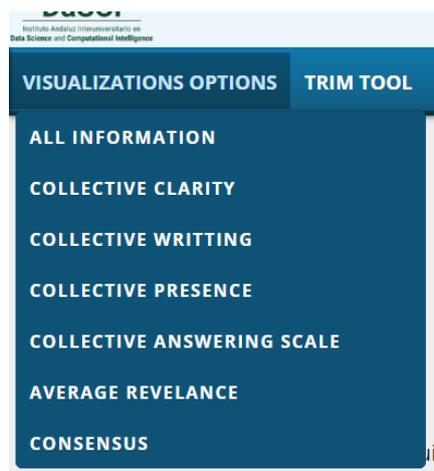


Figura 4.4: Opción de recorte en 2TFLD-DSS. Fuente: Elaboración propia.



Figura 4.5: Herramienta para ocultar items según el nivel de consenso del panel de expertos en 2TFLD-DSS. Fuente: Elaboración propia.

de evaluaciones más exhaustivas. Como se mencionó previamente, la Norma ISO 25010 surgió en el marco de referencia para evaluar la calidad del software, y su usabilidad.

El cuestionario creado, llamado *Emotional Linguistic Usability Test (ELUT)*, se basa en las dimensiones previamente mencionadas en la sección 3.3.2, con la finalidad de abarcar la complejidad de la usabilidad desde el enfoque estandarizado de la Norma ISO 25010, realizando una evaluación minuciosa del sistema, desde la perspectiva del usuario. La importancia de este cuestionario reside en su capacidad para brindar información detalla, sobre la usabilidad del sistema, desde la visión del usuario final, para facilitar el trabajo de diseñadores y desarrolladores, en conclusión, se contribuye al avance de la *HCI*. Finalmente, en la tabla 4.1 se presentan las preguntas utilizadas en el cuestionario.

La tabla 4.2 muestra los perfiles de los jueces que conforman el panel de expertos.

4.3.2. Aplicación del método 2TFLD: Primera ronda

Una vez definido el problema MEMCDM para validar el consenso del cuestionario Q_0 para la evaluación de la usabilidad. Se aplicaron las fases descritas a continuación:

Q_0		
Dimensión D_l	Ítem I_r	Pregunta
D_1	I_1	¿Qué sensación tienes al utilizar el sistema web para finalizar con éxito la tarea?
	I_2	¿Qué sensación tienes respecto a la integración de las diferentes funciones del sistema?
D_2	I_3	¿Cuál es la sensación que tienes respecto a la facilidad de aprender el uso del sistema web?
	I_4	¿Cuál es la sensación que tienes respecto a la facilidad de volver a realizar una tarea en el sistema web?
D_3	I_5	¿Cuál es la sensación que tienes al acceder a las diferentes secciones del sistema web?
	I_6	¿Cuál es la sensación que tienes al buscar información en el sistema?
D_4	I_7	¿Cuál es la sensación que tienes cuando el sistema web muestra un error?
	I_8	¿Cuál es la sensación que tienes al leer el mensaje de error en el sistema web?
D_5	I_9	¿Cuál es la sensación que tienes en la distribución de los componentes en el sistema web?
	I_{10}	¿Cuál es la sensación que tienes con el uso de los elementos del sistema web?
D_6	I_{11}	¿Cuál es la sensación que tienes al operar el sistema con tecnologías de asistencia web?
	I_{12}	¿Cuál es la sensación que tienes en la legibilidad de la información en el sistema web?

Tabla 4.1: Conjunto de preguntas para la primera ronda sobre Q_0 de ELUT.

- **Fase preliminar:** Esta fase se encargó de la preparación del cuestionario y el envío del formulario al conjunto de expertos.

Envío de Google forms: El formulario² se envió a 13 expertos, por medio de correo electrónico (ver figura 4.6), en donde se les solicitaba su participación en la validación del cuestionario.

Dichos expertos tenían diversas formaciones académicas como se muestra en la tabla 4.2. Este formulario se compone de la siguiente manera:

- **Parte superior:** En la sección 1, se encuentra una breve explicación sobre la norma ISO 25010 y su aportación en la evaluación de la usabilidad.
- **Conjunto de términos lingüísticos:** En la sección 2, están las opciones de términos lingüísticos a elegir. Sean $S^3 = \{s_0 = \text{Incorrecto}, s_1 = \text{Neutral}$

²Cuestionario ELUT, ronda 1: <https://forms.gle/jMTkejYRQmikh99c8>

Caracterización de los expertos		
J_i	Área de experiencia	Importancia W_{D_m}
J_1	Licenciado en Administración de empresas con posgrado en Tecnologías para el Aprendizaje	0.084
J_2	Licenciado en Cultura con postrado en Tecnologías para el Aprendizaje	0.050
J_3	Licenciado en Psicología con Doctorado en Ciencias de la Salud Pública	0.067
J_4	Licenciado en Informática con posgrado en Ingeniería de Software	0.084
J_5	Licenciado en Educación con posgrado en Tecnología para el Aprendizaje	0.076
J_6	Ingeniero en Computación con posgrado en Tecnologías de la Información.	0.084
J_7	Ingeniero en Energía con posgrado en Energía	0.076
J_8	Licenciado en Informática con posgrado en Tecnologías para el Aprendizaje	0.084
J_9	Licenciado en Informática con posgrado en Tecnologías para el Aprendizaje	0.076
J_{10}	Licenciado en Informática con Doctorado en Educación	0.084
J_{11}	Ingeniero en Telemática con posgrado en Tecnologías para el Aprendizaje	0.076
J_{12}	Ingeniero en Telemática con doctorado en Educación	0.084
J_{13}	Licenciado en Educación con doctorado en Psicología	0.076

Tabla 4.2: Panel de expertos.

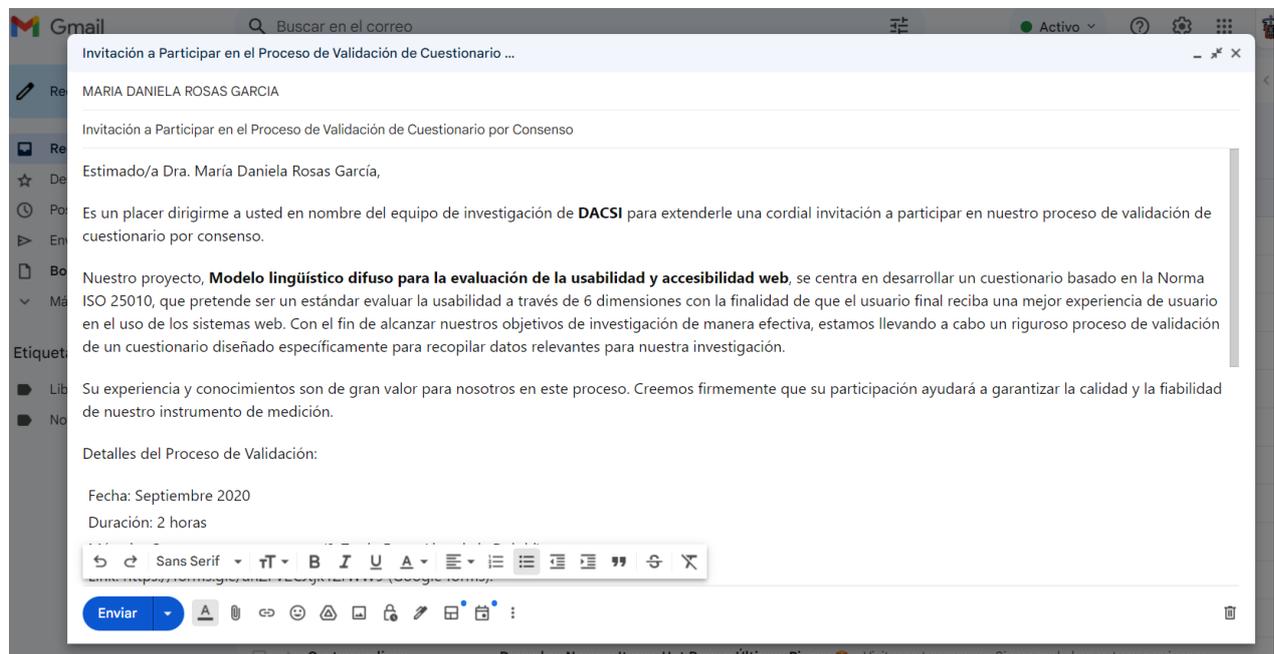


Figura 4.6: Correo electrónico enviado a los jueces que evaluaron el cuestionario. Fuente: Elaboración propia.

, $s_2 = \text{Correcto}$ }, $S^5 = \{s_0 = \text{Muy Incorrecto}, s_1 = \text{Incorrectos}, s_2 = \text{Neutral}, s_3 = \text{Correcto}, s_4 = \text{Muy correcto}\}$ y $S^7 = \{s_0 = \text{Muy Incorrecto}, s_1 = \text{Poco incorrecto}, s_2 = \text{Incorrecto}, s_3 = \text{Neutral}, s_4 = \text{Poco correcto}, s_5 = \text{Correcto}, s_6 = \text{Muy correcto}\}$ los LTS para que cada experto J_i emita su opinión a cada ítem I_r sobre el conjunto de criterios $C = \{c_1 = \text{claridad}, c_2 = \text{redacción}, c_3 = \text{pertenencia}, c_4 = \text{escala evaluativa}\}$.

- **Cuerpo del cuestionario:** En la sección 3, se presenta la descripción de cada una de las seis dimensiones (constructos) que el cuestionario Q_0 pretende evaluar, así como los ítems I_r asociados a cada dimensión D^l . En este espacio,

cada experto J_i emite su opinión sobre cada criterio del conjunto C . Finalmente, debajo de cada ítem I_r , esta un espacio de texto abierto para que cada experto J_i agregue un comentario.

- Fase 1: Recolección de la información:** Una vez finalizado el tiempo de respuesta para la evaluación de los ítems del cuestionario Q_0 , se recolecta la información de las valoraciones de todos los jueces en un archivo de texto ***.cvs**, considerando la fila de encabezados que debe contener. Se escribió como segunda columna la granularidad del conjunto de términos que eligió el experto J_i , tal que $S^{n(t)} \in S^{n(t^*)}$, seguido de las valoraciones de cada criterio, así como la importancia del ítem I_r para el cuestionario Q_0 . Los jueces del uno al tres eligieron la escala de tres términos (S^3), por su parte, los jueces del cuatro al nueve eligieron responder en escala de cinco términos (S^5), y finalmente los jueces del diez al 13 se sintieron cómodos al elegir escala de siete términos (S^7). La tabla 4.3 muestra las evaluaciones originales de la matriz $(X_4)_{13 \times 4}$.

Primera ronda para el ítem I_4					
J_i	Claridad	Redacción	Presencia	E. Evaluativa	Pertinencia
J_1	s_2^3	s_2^3	s_2^3	s_1^3	1
J_2	s_2^3	s_2^3	s_2^3	s_2^3	1
J_3	s_2^3	s_2^3	s_2^3	s_1^3	1
J_4	s_3^5	s_3^5	s_3^5	s_2^5	1
J_5	s_1^5	s_1^5	s_1^5	s_4^5	1
J_6	s_0^5	s_0^5	s_0^5	s_0^5	1
J_7	s_4^5	s_4^5	s_4^5	s_4^5	1
J_8	s_4^5	s_4^5	s_4^5	s_2^5	1
J_9	s_4^5	s_4^5	s_4^5	s_4^5	1
J_{10}	s_6^7	s_6^7	s_6^7	s_6^7	1
J_{11}	s_2^7	s_2^7	s_6^7	s_3^7	1
J_{12}	s_6^7	s_6^7	s_6^7	s_6^7	1
J_{13}	s_6^7	s_6^7	s_6^7	s_3^7	1

Tabla 4.3: Valoraciones en escala original de los jueces para el ítem I_4 .

En la carpeta *use-cases*³ se describe la conformación de los archivo de respuestas para la ronda 0 del cuestionario ELUT, asimismo archivo para los distintos pesos de los jueces evaluadores y la descripción de cada ítem.

- Fase de transformación a 2-Tuplas.** En este paso, se crea la matriz $(X'_4)_{13 \times 4}$ de los valores lingüísticos de 2-Tuplas.

³<https://github.com/ari-dasci/S-USE-AB-Tool/tree/main/use-cases>

- **Fase de unificación:** La tabla 4.4 muestra las evaluaciones unificadas de $(X''_4)_{13 \times 4}$ a S^{13} .

Primer ronda para I_4				
J_i	Claridad	Redacción	Presencia	E. Evaluativa
J_1	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_6^{13}, 0)$
J_2	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$
J_3	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_6^{13}, 0)$
J_4	$(s_9^{13}, 0)$	$(s_9^{13}, 0)$	$(s_9^{13}, 0)$	$(s_6^{13}, 0)$
J_5	$(s_3^{13}, 0)$	$(s_3^{13}, 0)$	$(s_3^{13}, 0)$	$(s_{12}^{13}, 0)$
J_6	$(s_0^{13}, 0)$	$(s_0^{13}, 0)$	$(s_0^{13}, 0)$	$(s_0^{13}, 0)$
J_7	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$
J_8	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_6^{13}, 0)$
J_9	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$
J_{10}	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$
J_{11}	$(s_4^{13}, 0)$	$(s_4^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_6^{13}, 0)$
J_{12}	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$
J_{13}	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_6^{13}, 0)$

Tabla 4.4: Valoraciones de los 13 jueces unificada a S^{13} .

- **Fase de doble agregación:** Por medio de los operadores presentados en Montes [25], se agregan las opiniones de los expertos para cada criterio. Se obtiene que la relevancia del ítem $W^4 = 1$ y el resultado para cada criterio en un vector Y_r que contiene la evaluación agregada de los 13 expertos para cada uno de los criterios.

$$Y_4 = \{(s_9^{13}, 0.387), (s_9^{13}, 0.387), (s_{10}^{13}, 0.059), (s_8^{13}, 0.178)\}$$

- **Fase de explotación:** Se re-transforma el resultado final del ítem I_4 al escala lingüísticas en S^7 dando como resultado $(s_5^7, -0.06)$, donde, el consenso $CS = falso$ y la confiabilidad $RS = verdadero$.

Para los demás ítems del cuestionario, el panel de expertos dio sugerencias variadas, por ejemplo, la redacción de algunas de las preguntas es muy similar a alguna anterior, la escala evaluativa habla emociones y las preguntas de sensaciones. Entre las evaluaciones recibidas por cuestiones de redacción (caso de I_2, I_3, I_4, I_{12}). Por otro lado, hubo ítems peor evaluados negativamente en la dimensión de pertinencia (caso de I_3, I_4, I_{11}). Asimismo se mencionó que algunas de las preguntas no se comprende en primera instancia que pretenden evaluar (caso de I_3, I_4, I_5). Finalmente, la dimensión de escala evaluativa fue la peor evaluada por el panel de expertos, debido a que en las respuesta se proponen emociones y las preguntas

QUESTIONNAIRE TOTAL SCORE = (Very Correct, -0.222)

Num	Item	Collective Clarity	Collective Writing	Collective Presence	Collective Scale	Collective Relevance	Score	Consensus	Final Results
I ₁	¿Qué sensación tienes al utilizar el sistema web para finalizar con éxito la tarea?	(Very Correct, 0.064)	(Very Correct, 0.383)	(Very Correct, 0.467)	(Very Correct, 0.421)	1.000	(Very Correct, 0.334)	<input checked="" type="checkbox"/>	(Very Correct, 0.334)
I ₃	¿Cuál es la sensación que tienes respecto a la facilidad de aprender el uso del sistema web?	(Very Correct, -0.055)	(Very Correct, -0.307)	(Very Correct, 0.029)	(Correct, 0.43)	1.000	(Very Correct, -0.226)	<input checked="" type="checkbox"/>	(Very Correct, -0.226)
I ₄	¿Cuál es la sensación que tienes respecto a la facilidad de volver a realizar una tarea en el sistema web?	(Very Correct, -0.307)	(Very Correct, -0.307)	(Very Correct, 0.029)	(Correct, 0.089)	1.000	(Very Correct, -0.374)	<input type="checkbox"/>	(Very Correct, -0.374)
Items Results		(Very Correct, -0.411)	(Very Correct, -0.400)	(Excellent, 0.478)	(Very Correct, 0.448)				

Showing 10 entries (filtered from 12 total entries)

As Linguistic Labels

University of Granada License CC-BY-NC-SA

Figura 4.7: Resultados finales del 2TLFD-DSS para Q_0 . Fuente: Elaboración propia.

hablan de sensaciones, motivo por el cual se recomienda redactar acorde a los datos que se pretenden recolectar.

La tabla 4.7 muestra los resultados obtenidos mediante 2TLFD-DSS para la ronda 0 del cuestionario ELUT. En la figura 4.7, se nota que no aparece el ítem I_2 , debido a que el RS es falso, al igual que el ítem I_{12} . Además, en el ítem I_4 tiene $CS = falso$, dado por un nivel de consenso < 0.5 .

4.3.3. Aplicación del método 2TLFD: Segunda ronda

Una vez recibidas las respuestas y analizar la evaluación que tiene cada ítem I_r sobre cada criterio, además de comprender las evaluaciones textuales añadidas por cada juez J_i en cada ítem I_r , se realizaron los ajustes y cambios pertinentes en el cuestionario Q_0 . Posteriormente, una nueva versión Q_1 con valores obtenidos de *Google Forms*⁴ es enviada a los 13 jueces para una re-evaluación de cada ítem I_r , así como la valoración obtenida en la ronda anterior.

A continuación, se presentan los pasos necesarios para la ronda 2 sobre el cuestionario Q_1 :

- **Fase 1: Recolección de la información:** En esta ronda, los jueces eligieron la misma granularidad de términos lingüísticos que la ronda anterior Q_1 . Una vez

⁴Cuestionario ELUT, ronda 1: <https://forms.gle/anZFV2CJtjk1ZrWW9>

finalizado el tiempo de respuesta para la evaluación de los ítems del cuestionario Q_1 . Primeramente, se recolectó la información de las valoraciones de cada J_i en un archivo de texto ***.cvs**, en la carpeta en línea *use-cases*⁵, se encuentran los archivos necesarios que se ingresaron para la ronda 2 en 2TLFD-DSS. La tabla 4.5 muestra las evaluaciones originales de la matriz $(X_4)_{13 \times 4}$.

Primer ronda para el ítem I_4					
J_i	Claridad	Redacción	Presencia	E. Evaluativa	Pertinencia
J_1	s_2^3	s_2^3	s_2^3	s_2^3	1
J_2	s_2^3	s_2^3	s_2^3	s_1^3	1
J_3	s_4^5	s_4^5	s_4^5	s_4^5	1
J_4	s_4^5	s_4^5	s_4^5	s_4^5	1
J_5	s_4^5	s_4^5	s_4^5	s_4^5	1
J_6	s_2^3	s_2^3	s_2^3	s_2^3	1
J_7	s_2^3	s_2^3	s_2^3	s_2^3	1
J_8	s_4^5	s_4^5	s_4^5	s_4^5	1
J_9	s_4^5	s_4^5	s_4^5	s_4^5	1
J_{10}	s_6^7	s_6^7	s_6^7	s_6^7	1
J_{11}	s_6^7	s_5^7	s_6^7	s_5^7	1
J_{12}	s_5^7	s_5^7	s_5^7	s_5^7	1
J_{13}	s_4^5	s_4^5	s_4^5	s_4^5	1

Tabla 4.5: Valoraciones en escala original de los jueces para el ítem I_4 para la segunda ronda sobre Q_1 de ELUT.

En la carpeta *rounds*⁶ se describe la conformación de los archivo de respuestas para la ronda 0 del cuestionario ELUT, asimismo archivo para los distintos pesos de los jueces evaluadores y la descripción de cada ítem.

- **Fase de transformación a 2-Tuplas.** En este paso, se crea la matriz $(X'_4)_{13 \times 4}$ de los valores lingüísticos de 2-Tuplas.
- **Fase de unificación:** La tabla 4.6 muestra las evaluaciones unificadas de $(X''_4)_{13 \times 4}$ a S^{13} .
- **Fase de doble agregación:** Por medio de los operadores presentados en Montes [25], se agregan las opiniones de los expertos para cada criterio. Se obtiene que la relevancia del ítem $W^4 = 1$ y el resultado para cada criterio en un vector Y_r que contiene la evaluación agregada de los 13 expertos para cada uno de los criterios.

$$Y_4 = \{(s_{12}^{13}, -0.152), (s_{11}^{13}, -0.004), (s_{12}^{13}, -0.152), (s_{11}^{13}, 0.224)\}$$

⁵<https://github.com/ari-dasci/S-USE-AB-Tool/tree/main/use-cases>

⁶<https://github.com/ari-dasci/S-USE-AB-Tool/tree/main/use-cases/rounds>

Primer ronda para I_4			
Claridad	Redacción	Presencia	E. Evaluativa
$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$
$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_6^{13}, 0)$
$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$
$(s_{12}^{13}, 0)$	$(s_3^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$
$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$
$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$
$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$
$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$
$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$
$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$
$(s_{10}^{13}, 0)$	$(s_{10}^{13}, 0)$	$(s_{10}^{13}, 0)$	$(s_{10}^{13}, 0)$
$(s_{10}^{13}, 0)$	$(s_{10}^{13}, 0)$	$(s_{10}^{13}, 0)$	$(s_{10}^{13}, 0)$
$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$	$(s_{12}^{13}, 0)$

Tabla 4.6: Valoraciones de los 13 jueces unificada a S^{13} para el ítem I_4 en la ronda 2.

- Fase de explotación:** Se re-transforma el resultado final del ítem I_4 al escala lingüísticas en S^7 . Finalmente, resultado lingüístico para el ítem $I_4 = (s_6^7, -0.261)$, donde, el consenso $CS = verdadero$ y la confiabilidad $RS = verdadero$. La figura 4.8 muestra los resultados del cuestionario Q_1 obtenidos de 2TFLD-DSS.

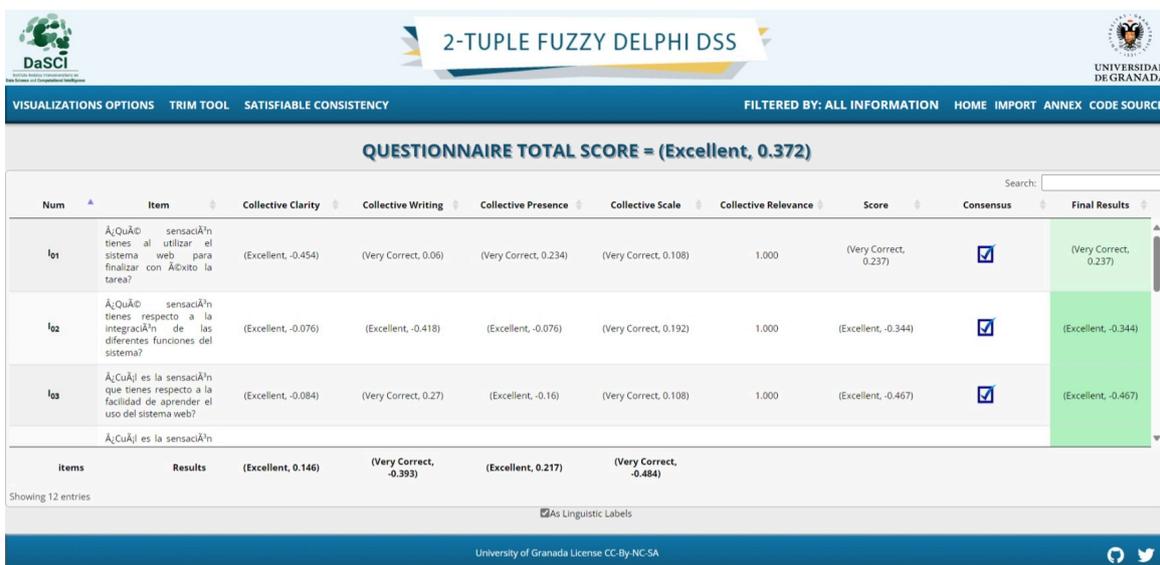


Figura 4.8: Resultados finales del 2TFLD-DSS para Q_1 . Fuente: Elaboración propia.

Al finalizar las dos rondas de validación de contenido por consenso, se obtuvo una valoración final para el $Q_1 = (s_6^7, -0.335)$ equivalente a *Label*. Los resultados de este proceso no solo fortalecen la validez y confiabilidad del instrumento, sino que también contribuyen significativamente a la calidad y robustez de los datos recopilados durante

el estudio.

La tabla 4.7 muestra los resultados obtenidos mediante 2TLFD-DSS para la ronda 0 del cuestionario ELUT. En la figura 4.7, se nota que no aparece el ítem I_2 , debido a que el RS es falso, al igual que el ítem I_{12} . Además, en el ítem I_4 tiene $CS = falso$, dado por un nivel de consenso < 0.5 .

4.4. Cuestionario ELUT

Como se observó, en la ronda dos se alcanzaron los requisitos necesarios para tener la validación por consenso del cuestionario, mismo que a continuación se presentan las preguntas asociadas al cuestionario ELUT.

Como se vio en la subsección anterior, luego de una segunda ronda se llegó a un consenso adecuado por parte del panel de expertos, asimismo, el nivel de confianza por encima de 0,75 también se logro con los ajustes y recomendaciones obtenidas en la ronda 1. La tabla 4.7 muestra el puntaje final para cada ítem I_r .

Item	Q_0					Q_1				
	IS	CI	CI	RI	RS	IS	CI	CS	RI	RS
I_1	$(s_5^7, 0.334)$	0.672	Verdadero	1.00	Verdadero	$(s_5^7, 0.275)$	0.616	Verdadero	1.00	Verdadero
I_2	$(s_4^7, 0.5)$	0.506	Verdadero	0.50	Falso	$(s_6^7, -0.316)$	0.745	Verdadero	1.00	Verdadero
I_3	$(s_5^7, -0.226)$	0.462	Falso	0.75	Verdadero	$(s_6^7, -0.467)$	0.693	Verdadero	1.00	Verdadero
I_4	$(s_5^7, -0.374)$	0.426	Falso	0.75	Verdadero	$(s_6^7, -0.261)$	0.811	Verdadero	1.00	Verdadero
I_5	$(s_6^7, -0.436)$	0.749	Verdadero	1.00	Verdadero	$(s_6^7, -0.378)$	0.750	Verdadero	1.00	Verdadero
I_6	$(s_6^7, -0.399)$	0.743	Verdadero	1.00	Verdadero	$(s_6^7, -0.498)$	0.652	Verdadero	1.00	Verdadero
I_7	$(s_6^7, -0.365)$	0.754	Verdadero	1.00	Verdadero	$(s_6^7, 0.247)$	0.581	Verdadero	1.00	Verdadero
I_8	$(s_6^7, -0.375)$	0.752	Verdadero	1.00	Verdadero	$(s_6^7, -0.224)$	0.795	Verdadero	1.00	Verdadero
I_9	$(s_5^7, 0.462)$	0.720	Verdadero	1.00	Verdadero	$(s_6^7, -0.029)$	0.965	Verdadero	1.00	Verdadero
I_{10}	$(s_5^7, 0.38)$	0.681	Verdadero	1.00	Verdadero	$(s_6^7, -0.029)$	0.965	Verdadero	1.00	Verdadero
I_{11}	$(s_4^7, 0.253)$	0.644	Verdadero	0.75	Verdadero	$(s_6^7, -0.17)$	0.876	Verdadero	1.00	Verdadero
I_{12}	$(s_5^7, 0.085)$	0.630	Verdadero	0.75	Verdadero	$(s_6^7, -0.17)$	0.876	Verdadero	1.00	Verdadero
	1 ronda					2 ronda				
Q	CC	CW	CP	CAS	QS	CC	CW	CP	CAS	QS
	$(s_5^7, 0.256)$	$(s_5^7, 0.195)$	$(s_6^7, -0.461)$	$(s_5^7, -0.377)$	$(s_5^7, 0.153)$	$(s_6^7, -0.147)$	$(s_6^7, -0.48)$	$(s_6^7, -0.217)$	$(s_6^7, -0.495)$	$(s_6^7, -0.335)$

Tabla 4.7: Tabla comparativa de resultados lingüísticos S^7 , consenso CS y confianza RS para cada ítem I_r en ronda 1 y 2.

Sin embargo, para la recolección de los datos, además de validez de contenido es necesario validar la confiabilidad de un cuestionario para garantizar características como la consistencia de las mediciones, precisión en las respuestas y credibilidad del estudio.

4.4.1. Validación de la confiabilidad del cuestionario ELUT

La validación de contenido en un instrumento es esencial para garantizar que cada ítem contribuya de manera significativa y coherente al constructo que se está midiendo. La inclusión de ítems relevantes y representativos dentro del instrumento mejora la validez y confiabilidad de los resultados obtenidos. Al utilizar el coeficiente Alfa de Cronbach como medida de consistencia interna, se busca evaluar la homogeneidad de los ítems y su capacidad para medir la misma variable subyacente. Establecer un umbral deseable por encima de 0.70 para el coeficiente Alfa de Cronbach es crucial, ya que indica el nivel de fiabilidad del instrumento. Un Alfa de Cronbach más alto sugiere una mayor consistencia interna entre los ítems, lo que implica una medición más confiable y precisa del constructo.

El Coeficiente Alfa de Cronbach, al ser aplicado una única vez al instrumento de medición, genera valores comprendidos entre 0 y 1. Su beneficio radica en la eliminación de la necesidad de dividir los ítems del instrumento en dos partes, ya que solo se requiere una administración única para realizar la medición y calcular el coeficiente [148].

Para realizar la validación de la confiabilidad del Test de Usabilidad Lingüístico Emocional (ELUT, por sus siglas en Inglés), el proceso se dividió en cuatro fases. En esta validación se realizó una prueba piloto con 62 estudiantes del Centro Universitario de Tonalá, de la carrera de Ingeniería en Ciencias Computacionales, con un rango de edad de 18 a 25 años, siendo la mayoría hombres. Los estudiantes se encargaron de responder el cuestionario ELUT basado en su interacción con un sitio web al final de la Prueba de Usabilidad.

Fase 1: Definición de la Prueba de Usabilidad.

La Prueba de Usabilidad (PU) se realizó siguiendo los pasos que se detallan a continuación:

1. **Desarrollo de Prueba de Usabilidad:** El primer paso fue la creación de la Prueba de Usabilidad, la que se conformó por diez de tareas que tenían como finalidad conocer y evaluar el uso de una web.
2. **Conformación de grupo de usuarios:** Posteriormente, se seleccionó a un grupo de usuarios, quienes realizaran la Prueba de Usabilidad. El proceso de selección de usuarios se basó en las características de los mismos, en este caso, fueron seleccionados

estudiantes.

3. **Recolección de respuestas:** Finalmente, se abrió una clase en la plataforma Classroom, en donde se inscribieron a los estudiantes y se publicaron diversas tareas que tenían que realizar como parte de la PU.

Fase 2: Aplicación de la Prueba de Usabilidad.

La forma de aplicación de la PU fue la siguiente, se les pidió a los estudiantes que leyeran las instrucciones sobre una serie de diez tareas que debían realizar utilizando el lector de pantalla de Windows simulando una discapacidad visual. Se les dejó elegir una red social para realizar las tareas que se describen a continuación (ver tabla 4.8):

Nº	Instrucción
1	Iniciar sesión en la red social con la cuenta personal.
2	Actualizar perfil, por medio de un cambio en la foto de perfil y actualizar su información personal.
3	Publicar un estado o actualización en su perfil con un mensaje y una imagen.
4	Buscar amigos o encontrar a un compañero de clase y agregarlo como amigo.
5	Enviar un mensaje privado a un amigo preguntándole cómo está.
6	Configurar la privacidad de su perfil para que solo sus amigos puedan ver sus publicaciones.
7	Participar en un grupo al que debe unirse, que esté relacionado con sus intereses académicos.
8	Publicar en el grupo al que se unió, compartiendo información relevante.
9	Explorar el contenido de la red social, para encontrar una publicación interesante y poner Me gusta o dejar un comentario.
10	Cerrar sesión en su cuenta.

Tabla 4.8: Prueba de usabilidad para redes sociales.

Fase 3: Aplicación de ELUT.

Finalmente, se les solicitó que contestaran cuidadosamente y con suma consciencia el cuestionario de 12 preguntas ELUT.

Cuestionario	Test de Evaluación Lingüística de la Usabilidad ELUT
Objetivo	Comprender profundamente como las interacciones en un sistema web afectan emocionalmente a los usuarios. Este cuestionario busca capturar no solo la eficacia técnica del sitio web sino también la calidad de la experiencia emocional de los usuarios durante su navegación.
Muestra	62 alumnos entre 18 y 25 años

Tabla 4.9: Test de Usabilidad Lingüístico Emocional.

Fase 4: Análisis de resultados.

A continuación, se presentan los resultados estadísticos, obtenidos mediante el programa SPSS V25, con relación a los datos informativos de los usuarios que respondieron el cuestionario y los valores del alfa de Cronbach para Q_2 . La tabla 4.10 muestra que los valores del alfa de Cronbach, con un valor global de .833, lo que, según la literatura, es un nivel alto de consistencia interna [148].

Alfa de Cronbach	N de elementos
0.833	12

Tabla 4.10: Estadísticas de fiabilidad.

Además, se presentan los resultados por cada una de las 12 preguntas realizadas en el cuestionario, correspondientes a las seis dimensiones del cuestionario. En la tabla 4.11 se presenta la medida si el elemento se ha suprimido, la varianza de escala si los elementos se han suprimido, la correlación corregida, para finalmente, mostrar el Alfa de Cronbach para cada uno de los ítems, reafirmando una consistencia interna buena, ya que todos los ítems se encuentran por encima del .80 en su resultado final.

	Media de escala si el elemento se ha suprimido	Varianza de escala si el elemento se ha suprimido	Correlación total de elementos corregida	Alfa de Cronbach si el elemento se ha suprimido
I_1	32.18	38.017	.636	.808
I_2	32.16	38.793	.563	.814
I_3	32.29	39.160	.571	.814
I_4	31.69	39.200	.470	.823
I_5	33.23	43.194	.349	.830
I_6	33.00	45.016	.122	.845
I_7	32.34	39.670	.624	.811
I_8	32.29	37.882	.687	.804
I_9	32.48	40.549	.455	.823
I_{10}	32.15	39.897	.486	.821
I_{11}	31.19	42.716	.281	.836
I_{12}	31.95	38.047	.641	.807

Tabla 4.11: Estadísticas de total de elemento.

También, se consideraron la distintas formas, contextos y aplicaciones en las que los usuarios pueden interactuar con el sistema. Los expertos deben poseer habilidades para analizar y comprender los principios teóricos de aquello que están evaluando [149], en este caso, elementos referentes a la usabilidad y sus dimensiones según la Norma ISO 25010.

4.4.2. Cuestionario final Q_1

A continuación, en la tabla 4.12 se presenta el cuestionario final para ELUT validado por consenso por expertos así como también por metodologías estadísticas que confirman su confiabilidad y fiabilidad en los resultados.

		Q_1
D_i	I_r	Pregunta
D_1	I_1	¿Qué emoción genera en ti la facilidad de aprender a utilizar el sistema web?
	I_2	¿Qué emoción emerge en ti al utilizar de nuevo el sistema web?
D_2	I_3	¿Qué emoción te genera el uso del sistema web?
	I_4	¿Qué emoción te genera realizar con facilidad una tarea en el sistema web?
D_3	I_5	¿Qué emoción emerge en ti cuando el sistema web muestra un error?
	I_6	¿Qué emoción genera en ti el texto en los mensajes de error que muestra el sistema web?
D_4	I_7	¿Qué emoción emerge en ti al utilizar los distintos componentes tales como botones, links, formularios, entre otros, del sistema web?
	I_8	¿Qué emoción generan en ti la distribución de distintos componentes tales como botones, links, formularios, entre otros, del sistema web?
D_5	I_9	¿Qué emoción emerge en ti al utilizar tecnologías de asistencia como lectores de pantalla en el sistema web?
	I_{10}	¿Qué emoción genera en ti la legibilidad de los textos, imágenes y componentes en el sistema web?
D_6	I_{11}	Al utilizar el sistema web, ¿Qué emoción emerge en ti al finalizar con éxito una tarea?
	I_{12}	¿Qué emoción emerge en ti al utilizar las diferentes secciones del sistema web?
Escala evaluativa: a) Enojo b) Frustración c) Indiferencia d) Alegría e) Felicidad		

Tabla 4.12: Conjunto de preguntas para el cuestionario ELUT tras la ronda dos.

El cuestionario ELUT, sirve como entrada de datos para el modelo propuesto en la presente investigación.

4.5. Resumen de capítulo

En este capítulo hemos visto que:

- Para realizar una investigación, es importante tener instrumentos que sean válidos, confiables y objetivos, siguiendo principios básicos metodológicos. Cuando se elige un instrumento de medición se debe garantizar que mida lo que dice medir.
 - Existen pasos para construir un cuestionario o test válidos, que son: definir el objetivo del test, revisión de literatura, desarrollo de los items, consulta de expertos y validación de contenido, prueba piloto, análisis de la confiabilidad y revisión final.
 - Se diseña un instrumento llamado ELUT, siguiendo la propuesta de las dimensiones de la Norma ISO 25010, con la finalidad medir la usabilidad de sitios web.
 - Para realizar la validación de contenido del ELUT, se utilizó el Método Delphi Linguístico Difuso de 2-Tuplas como una innovación del modelo Delphi clásico. Se realizó el proceso de validación de contenido, en el que se utilizaron las opiniones de 13 expertos, de diversas formaciones académicas y nivel de pericia sobre el tema. El proceso se llevó a cabo mediante un Modelo de Toma de Decisiones con enfoque lingüístico.
 - Se alcanzó el consenso entre los expertos con respecto al contenido de los items del ELUT posterior de dos rondas de evaluación.
 - Posterior a la validación de contenido del ELUT, se trabajó con respecto a la fiabilidad por medio de un pilotaje del test. Ese pilotaje se llevó a cabo con 62 estudiantes del Centro Universitario de Tonalá, de la carrera de Ingeniería en Ciencias Computacionales. Los resultados del análisis estadístico reflejaron un Alfa de Cronbach de .833, siendo un nivel alto de consistencia interna.
 - La versión final del cuestionario para la evaluación de la usabilidad ELUT, se propone como un test complementario a otros instrumentos estandarizados para la evaluación de la usabilidad.
-

Capítulo 5

Un modelo de toma de decisiones lingüísticas multiexperto y multicriterio para la evaluación de la usabilidad

LDM4WUE es una metodología centrada en el usuario para la evaluación de la usabilidad a través de un problema Lingüístico de Toma de Decisiones Multi Experto Multi Criteria (MEMCLDM por sus siglas en inglés). LDM4WUE combina múltiples fuentes de datos para generar información integrada, desde el uso de pruebas estandarizadas (ver sección 3.3.3.2), hasta pruebas de usabilidad (PU). Además, tiene en cuenta la aproximación al razonamiento humano a través de la transformación lingüística de las evaluaciones/opiniones de los usuarios sobre los sitios web. LDM4WUE mejora herramientas de A/B testing, teniendo como ventaja la incorporación de los roles de usuarios así como mantener el mismo grado de complejidad cuando se evalúan mas de dos alternativas.

Este capítulo se conforma de la siguiente manera: La sección 5.1, presenta el flujo de trabajo del modelo de toma de decisiones lingüístico para la evaluación de la usabilidad. La sección 5.2, describe las características que requiere el modelo lingüístico de toma de decisiones. La sección 5.4, describe los criterios y la forma en que se recolecta la información para la evaluación de las alternativas. La sección 5.5, detalla los procedimientos para agregar las evaluaciones del conjunto de usuarios sobre cada alternativa. La sección 5.6,

76.1. Descripción general y diagrama de flujo de la metodología LDM4WUE.

describe el procedimiento para obtener los rankings de cada rol sobre una alternativa así como el ranking global de las alternativas evaluadas. Finalmente, la sección 5.7 presenta los puntos mas importantes de la metodología LDM4WUE.

5.1. Descripción general y diagrama de flujo de la metodología LDM4WUE.

La evaluación de la usabilidad web desde una perspectiva centrada en el usuario se enfoca principalmente en la satisfacción con el uso del sitio, así como en la información recopilada desde la Experiencia de usuario (*User eXperience*, por sus siglas en inglés) del experto o usuario final.

Dado el amplio uso de la escala SUS Adjetivo [45], nos basamos en la información lingüística y en el paradigma de Toma de Decisiones Lingüísticas [33]. Por ello, se adaptan los siguientes pasos para disponer de un esquema resolutivo.

1. **Estructuración del problema.** Se establece un moderador que contribuye a definir el conjunto de alternativas como sitios web a comparar, sus criterios asociados como pruebas para la evaluación de la usabilidad web y el conjunto de usuarios que evalúa los sitios web alternativos en función de los criterios, así como la configuración de usuarios, roles y su importancia relativa.
2. **Empatizar y role-playing.** Antes de realizar la prueba de usabilidad o cualquier otra prueba, el moderador define una serie de papeles que los usuarios pueden desempeñar al evaluar. Esta fase consiste en explicar los objetivos de la evaluación de usabilidad y la técnica del juego de roles. El moderador de la prueba, además, da un tiempo extra para que los participantes elijan un papel. Normalmente se lanza un dado para que este momento sea más dinámico.
3. **Elicitación de la información de los usuarios.** Los usuarios, que desempeñan determinados papeles, evalúan individualmente las alternativas con base en un conjunto de criterios. Se recopilan las evaluaciones de cada prueba y, a continuación, se construye una matriz de decisión individual para cada usuario que desempeña un papel. Posteriormente, se integran computacionalmente las evaluaciones de cada

prueba dando lugar a términos lingüísticos. Se define S^g como un conjunto de términos lingüísticos de g elementos lingüísticos $\{s_0, \dots, s_{g-1}\}$. Una simple opinión lingüística se denota como $s_i \in S^g$. Bajo la representación lingüística de 2-Tuplas, el mismo término se denota como $(s_i, 0)$.

4. **Agregación colectiva.** Esta fase agrega las evaluaciones individuales de los usuarios para obtener la agregación colectiva. Las evaluaciones de términos lingüísticos de las pruebas de criterios pertenecen a diferentes conjuntos de términos lingüísticos, que nosotros representamos como una jerarquía. Para agregarlos, primero se unifican los distintos términos lingüísticos de modo que todos pertenezcan al nivel más profundo de la jerarquía, que en este caso es el conjunto de términos S^g . A continuación, se agregan las evaluaciones individuales por roles obteniendo una matriz de decisión colectiva unificada para cada rol. Por último, se agregan las matrices anteriores en la decisión colectiva unificada que compila las evaluaciones de todos los usuarios en una única matriz. Las agregaciones se calculan mediante el operador de media ponderada de 2 matrices (*2TWA*).
5. **Explotación.** Para rankear las alternativas de la mejor a la peor valorada, aplicamos el algoritmo TOPSIS [150]. Para mayor comodidad, además de disponer de una clasificación, es posible presentar información lingüística de salida en una escala específica. En este caso, se realiza este paso antes de generar el informe utilizando la escala SUS Adjetivo.

En la figura 5.1, se describe el flujo de trabajo que propone el modelo LDM4WUE para la obtención de las evaluaciones de la usabilidad.

5.2. Fase 1: Descripción del Problema

La configuración del modelo lingüístico con el que se basa la metodología LDMWUE establece los elementos fundamentales requeridos para realizar la evaluación de la usabilidad web. En particular, se define una entidad moderadora que engloba los siguientes aspectos: (1) alternativas, que se refieren a los sitios web o plataformas en línea que se van a evaluar, (2) criterios, que abarcan cuestionarios estandarizados, herramientas de accesibilidad y pruebas de usabilidad, (3) evaluadores, denominados usuarios, que comprenden tanto a



Figura 5.1: Diagrama de flujo que representa las fases de la aplicación. Fuente: Elaboración propia.

los usuarios finales como a los expertos en alguna disciplina del *HCI*, los cuales evalúan las alternativas según los criterios, (4) roles, donde el moderador puede elegir todas o una parte de los roles de usuario descritos en la Figura 5.2, y (5) ponderaciones, el modelo puede incluir ponderaciones de estándares, ponderaciones de usuarios y ponderaciones de roles.

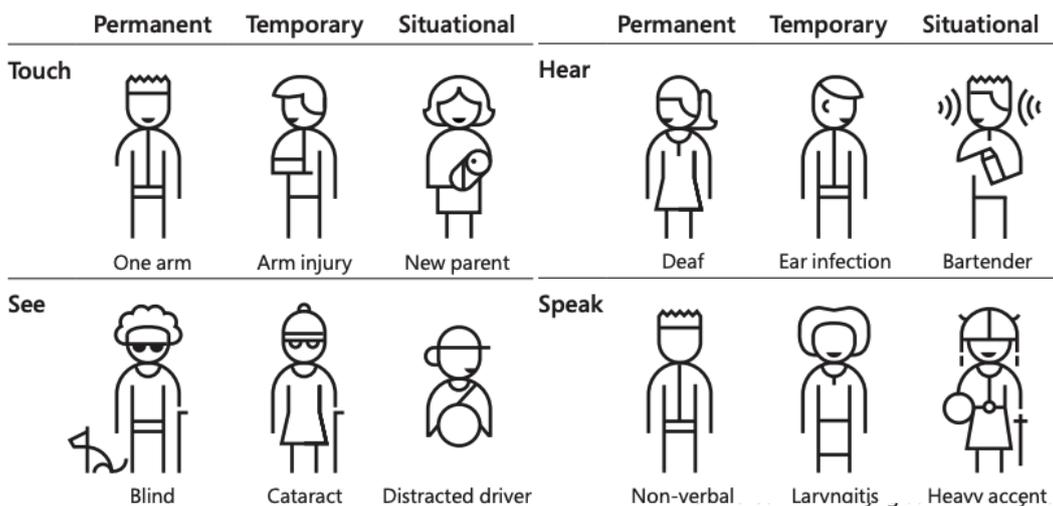


Figura 5.2: *Roleplaying* que permite empatizar con usuarios. Fuente [151].

Paso 1. Definición del conjunto de alternativas.

Se define al conjunto de alternativas $A = \{A_1, \dots, A_n\}$ como sitios web, versiones de sitios web, páginas web o plataformas en línea que tienen objetivos similares. Se refiere a cada alternativa como $A_i, (i = 1, \dots, n)$.

Paso 2. Definición del conjunto de criterios.

Se define un conjunto de criterios $C = \{C_1, \dots, C_m\}$ a ser evaluados, donde cada uno es referenciado por $C_j, (j = 1, \dots, m)$. En esta propuesta, con el *Decision Support Sytem*, el *moderador* (normalmente el desarrollador o diseñador del sitio) puede crear un A/B testing con m cantidad de pruebas. Dependiendo del LDM, LDM4WUE, se pueden considerar diferente importancia a cada criterio. Sea el vector de importancia para el conjunto de criterios C , el vector $WC' = \{WC'_1, \dots, WC'_m\}$ y WC el vector normalizado de importancia de criterios, tal que $WC = \{WC_1, \dots, WC_m\}$, donde $\sum_{j=1}^m WC_j = 1$. Las pruebas que define LDM4WUE son las siguientes:

- $C_1 \cong SUS$ se completa con 10 ítem en escala Likert (ver subsección 3.3.3.2).
- $C_2 \cong NPS$ se completa con un simple numero entre 0 y 10 (ver subsección 3.3.3.2).
- $C_3 \cong PU$ se completa por cada tarea perteneciente a esta prueba, tal que cada tarea tiene un valor booleano *task-done*, *task_time* en segundos, y un término lingüístico para la satisfacción al realizar la tarea. Este término está dentro de $S^5 = \{\text{Insatisfactorio, Un poco satisfactorio, Indiferente, Satisfactorio, Muy satisfactorio}\}$
- $C_4 \cong ACC$ se rellena con el valor del informe para Accesibilidad. Este término está en $S^3 = \{A, AA, AAA\}$ (ver subsección 3.2.2.2).
- $C_5 \cong ELUT$. Se completa con 12 ítems con respuestas en términos lingüísticos (ver subsección 4.4.2).

Para determinar con precisión la ponderación de los criterios, se sugiere aplicar el método del Proceso Analítico Jerárquico Difuso Extendido (FAHP) como se detalla en la sección 2.3.1. Este método facilita la organización de los criterios en una jerarquía y permite detectar incoherencias mediante un análisis de coherencia entre las opiniones del moderador. El resultado del uso de esta técnica es un vector normalizado de pesos WC para cada criterio $C_j, (j = 1, \dots, m)$.

Paso 3. Definición del conjunto de usuarios.

Se consideran dos grupos de usuarios, denominados expertos y usuarios finales. Sea $E = \{E_1, \dots, E_p\}$ el conjunto de expertos con conocimientos en algún área de tecnología, interfaces o experiencia de usuario, donde p es el número total de expertos. Sea $D = \{D_1, \dots, D_q\}$ el conjunto de usuarios finales, donde q es el número total de usuarios no expertos. Por lo tanto, el conjunto de usuarios $U = E \cup D$ es la unión de expertos y usuarios finales, y cada usuario es referenciado por U_k , ($k = 1, \dots, u$) donde $u = p + q$. Los usuarios tienen asociado un peso correspondiente a su grupo: $WE \in [0, 1]$ si son expertos, o $WD \in [0, 1]$ para los usuarios no expertos. Ambos valores los define directamente el moderador. Finalmente, el vector $WU = \{WU_1, \dots, WU_u\}$ se rellena con valores según la pertenencia del decisor a uno de los dos grupos. Por ejemplo, si tenemos un experto y dos usuarios finales, este vector es $WU = \{WE, WD, WD\}$. Se puede notar que no se impone que $WE + WD = 1$, sino que tenemos que calcular la normalización de WU .

Paso 4. Definición del conjunto de roles .

La metodología LDM4WUE pretende centrarse en el usuario final, por lo que se apoya en el *Design Thinking* con la técnica de role-playing (ver sección 3.1) con la técnica de role-playing (ver sección 5.2) para retener las necesidades del usuario final. Sea $R = \{R_1, \dots, R_r\}$ el conjunto de roles definidos por el moderador donde los roles posibles son R_l , ($l = 1, \dots, r$). Cada usuario asume un rol en el que evalúa todas las alternativas, y la importancia de cada rol varía según los requisitos específicos del problema. Sea $WR' = \{WR'_1, \dots, WR'_r\}$ el vector de importancia asociados a los roles establecidos directamente por la figura del moderador. Este vector se normaliza obteniendo el vector de pesos de roles $WR = \{WR_1, \dots, WR_r\}$, donde $\sum_{l=1}^r WR_l = 1$.

5.3. Fase 2. Empatía y role-playing

Design Thinking (DT) se define como “un proceso de innovación centrado en el ser humano que hace hincapié en la observación, la colaboración, el aprendizaje rápido, la visualización de ideas, la creación rápida de prototipos y el análisis simultáneo del negocio”

[152]. Se trata de una metodología con gran capacidad para comprender las necesidades de las personas mediante el establecimiento de fases bien definidas y la aplicación de numerosas herramientas para el análisis conceptual de las necesidades de los usuarios y la identificación de requisitos para el desarrollo de software. Dada su naturaleza centrada en el usuario, este tipo de metodología se centra en la recopilación de las características y necesidades de los usuarios.

Se aplica *role-playing* en la metodología para permitir que las personas expresen alguna situación temporal de su vida, como un brazo roto o incluso estados de ánimo. Al vincular un rol a la evaluación, el experto en UX que realiza las pruebas A/B también puede aplicar varias evaluaciones para cada una de las *personas* definidas (un arquetipo de usuario que ayuda a diseñadores y desarrolladores a empatizar con personas con necesidades especiales [153]). Por ejemplo, pueden representar el papel de un estudiante extranjero que visita el sitio web de la universidad o empatizar con una persona con discapacidad visual. Existen numerosas posibilidades para definir estos roles, como se ilustra en la figura 5.2 [151], que ofrece algunos ejemplos.

Paso 5. Sesión informativa de prueba. Hay varios enfoques para medir la usabilidad de un sistema (como hemos descrito anteriormente en la Sección 3.3), y una de las mejores maneras de hacerlo es observar a los usuarios cómo se desempeñan en las tareas más simples o más complejas. Esto se consigue con una Prueba de Usabilidad. Existen guiones del libro 'No me hagas pensar' [154] que ayudan a realizar una prueba de usabilidad, como instrucciones para la persona que va a observar a los usuarios, o una lista de cosas neutras que el facilitador puede decir mientras el participante realiza las tareas. Sin embargo, lo mejor que se puede hacer para garantizar el éxito de la UT es un briefing a los participantes para que sepan de antemano lo que van a hacer.

Paso 6. Elegir rol o dados. Para empatizar con las personas y sus múltiples necesidades, utilizamos el juego de roles como técnica que nos permitirá simular condiciones variadas. Una forma gamificada de utilizarlo es mediante el uso de dados. Supongamos tres dados: el primer dado determina la edad (adolescente, anciano,..); un segundo dado simula una condición física concreta (ver, tocar,..); y un tercer dado establece una condición anímica concreta (deprimido, estresado, cansado,..). De este modo, podríamos jugar a empatizar con un anciano motivado y con problemas de visión.

5.4. Fase 3: Recolección de la información de los usuarios

Todos los usuarios evalúan los criterios C_1, C_2, C_3 (comunes para U) y C_4 (sólo para E) de forma particular según la prueba a la que se refiera. Y cada evaluación individual está vinculada al rol seleccionado. Cada usuario U_k , que juega un rol R_l , proporciona su opinión sobre cada alternativa A_i y según los criterios C_j . En conjunto, configura la matriz de decisión individual $ID^{k,l} = (ID_{ij}^{k,l})_{n \times m}$. A continuación, se explica en detalle los pasos 5 y 6 para cada instrumento o criterio C_j con el fin de ofrecer un procedimiento claro para construir las matrices de decisión individual, respectivamente.

Paso 7. Recopilación de evaluaciones de usuarios

El usuario U_k que juega un rol R_l evalúa una alternativa A_i realizando una o más pruebas C para evaluar la usabilidad de A_i . En seguida se muestra la información proporcionada por cada prueba C_j posible:

- $C_1 \cong$ **System Usability Scale (SUS)**.

El usuario responde a diez preguntas siguiendo una escala Likert con cinco opciones de respuesta para cada una. Las preguntas impares tienen una connotación positiva, mientras que las pares tienen un tinte negativo.

Para cada alternativa A_i , se denominan las diez respuestas proporcionadas por el usuario U_k jugando el rol R_l como $x_h^{k,l,i}$, $h = 1, \dots, 10$. Se obtiene *SUS score* del usuario U_k jugando el rol R_l para cada alternativa A_i , ($i = 1, \dots, n$), por medio de la ecuación 5.1:

$$SUS_score_i^{k,l} = 2.5 \times \sum_{h=1}^5 [(x_{2h-1}^{k,l,i} - 1) + (5 - x_{2h}^{k,l,i})], \forall i = 1, \dots, n \quad (5.1)$$

Para cada usuario U_k jugando el rol R_l , los valores $SUS_score_i^{k,l} \in [0, 100]$ están disponibles para cada alternativa A_i , ($i = 1, \dots, n$). Esto implica que se tiene por cada criterio C_1 una 2-Tupla lingüística dentro de la escala Adjetiva de SUS (ver figura 5.3). Los cálculos para obtener esta 2-Tupla lingüística se detalla en el paso 9 de la sección 5.5.

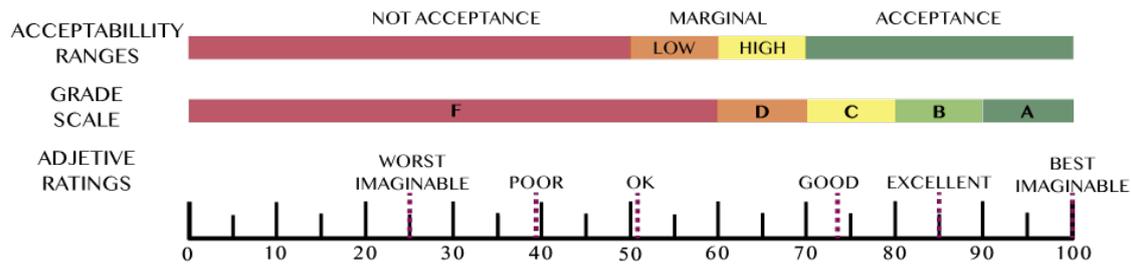


Figura 5.3: Equivalencias para escala SUS¹.

■ $C_2 \cong$ **Net Promoter Score (NPS)**.

El usuario se sitúa ante una única pregunta conocida como Probabilidad de Recomendar (LTR): *¿Qué tan dispuesto está a recomendar el sitio web representado por A_i ?* La respuesta debe ser un valor entero perteneciente al intervalo $[0, 10]$, de modo que los valores cercanos a 0 representan que no recomendaría el sitio web, mientras que los valores cercanos a 10 representan que sí lo recomendaría. Es frecuente obtener este cómputo con ayuda en línea².

En resumen, para cada usuario U_k que juega el rol R_l , existe un valor $NPS_LTR_score_i^{k,l} \in [0, 10]$ para cada alternativa A_i , ($i = 1, \dots, n$) que permite evaluar el criterio C_2 con una 2-Tupla lingüística en la escala Adjetiva de SUS. Los cálculos detallados se proporcionan en el paso 9 de la sección 5.5.

■ $C_3 \cong$ **Prueba de Usabilidad (PU)**. En este test, el usuario U_k debe responder tantas preguntas/actividades como el moderador estableció la PU . La PU se compone de d tareas que deben ser realizadas. Esas d tareas definen la prueba de usabilidad como $UT = \{q_1, \dots, q_d\}$. El usuario U_k jugando el rol R_l para la alternativa A_i , ($i = 1, \dots, n$) respecto a cada tarea q_v , $v = 1, \dots, d$ (provista en el cuestionario PU), se conforma de las siguientes cuatro medidas:

1. **Eficiencia** ($Efficiency_i^{k,l}(q_v)$). Esta medida establece si el usuario ha conseguido realizar la tarea solicitada en un tiempo adecuado. El usuario controla el tiempo empleado ($time_i^{k,l}(q_v)$) y LDM4WUE-DSS lo compara con la estimación del moderador para el tiempo máximo ($MaxTime(q_v)$). Esta medida puede tomar dos valores: 1 si el usuario completa q_v por debajo del tiempo estimado, *i.e.*, si $time_i^{k,l}(q_v) \leq MaxTime(q_v)$ y 0 en caso contrario.

²NPS calculator <https://npscalculator.com/>

2. **Éxito** ($Success_i^{k,l}(q_v)$). El usuario indica si ha realizado la tarea con éxito o sin éxito. Esta medida puede tomar dos valores: 1 en caso de éxito y 0 en caso de fracaso.
3. **Satisfacción** ($Satisfaction_i^{k,l}(q_v)$). El usuario indica la sensación que ha experimentado al resolver la tarea. Se expresa mediante un adjetivo de los cinco posibles: *unsatisfying*, *a little satisfying*, *indifferent*, *satisfying*, y *very satisfying* que corresponden a los cinco términos lingüísticos de $S^5 = \{s_0, s_1, s_2, s_3, s_4\}$.
4. **Satisfacción_PU** ($Satisfaction_PU_i^{k,l}$). Este indicador se obtiene al aplicar la media aritmética al conjunto de respuestas expresadas por el usuario para cada $PU(q_v)$. El resultado es una 2-Tupla lingüística que posteriormente se inserta en la matriz de decisiones individuales.

Se calculan las medidas de éxito, eficiencia y satisfacción para cada pregunta utilizando las respuestas de todos los usuarios, con el propósito de identificar aquellas tareas que resultan particularmente difíciles. Asimismo, se evalúan estas tres métricas para cada posible rol desempeñado para determinar si existen tareas que resulten más complejas para un rol específico de usuario. A partir de WU , obtenemos r vectores ponderados WU^l que contienen los pesos normalizados de todos los usuarios que desempeñan el rol R_l (donde $l = 1, \dots, r$). Luego, se calcula el éxito, la eficiencia y la satisfacción asociados a la tarea q_v , ($v = 1, \dots, d$) y al rol dado R_l mediante las siguientes ecuaciones:

$$Efficiency_i^l(q_v) = 100 \times \frac{\sum_{k=1}^{u^l} Efficiency_i^{k,l}(q_v)}{u^l} \quad (5.2)$$

$$Success_i^l(q_v) = 100 \times \frac{\sum_{k=1}^{u^l} Success_i^{k,l}(q_v)}{u^l} \quad (5.3)$$

$$Satisfaction_i^l(q_v) = \Delta \left(\sum_{k=1}^{u^l} \Delta^{-1}(Satisfaction_i^{k,l}(q_v), 0) \times WU_k^l \right) \quad (5.4)$$

se agrega la siguiente ecuación, como la media aritmética de todas las preguntas de la PU

$$Satisfaction_PU_i^l = \Delta \left(\sum_{k=1}^{u^l} \Delta^{-1}(Satisfaction_i^{k,l}(q_v), \alpha) \right) \quad (5.5)$$

donde u^l es el numero de usuarios jugando el rol R_l .

Las métricas anteriores de éxito y eficiencia son porcentajes, mientras que la métrica de satisfacción es un valor lingüístico de 2-Tuplas. Toda esta información se recoge en el informe de usabilidad complementario a las soluciones de clasificación.

Entonces, para cada usuario U_k jugando el rol R_l y para cada tarea q_v , se tienen cuatro valores: $Success_score_i^{k,l}(q_v) \in [0, 1]$, $Efficiency_score_i^{k,l}(q_v) \in [0, 1]$, $Satisfaction_score_i^{k,l}(q_v) \in S^5$ y $Satisfaction_PU_i^{k,l} \in S^5$. Después de la agregación para cada alternativa $A_i, (i = 1, \dots, n)$, se evalúa el criterio C_3 basado en la satisfacción global (el éxito y la eficiencia se considerarán métricas secundarias para el informe), ya que puede ser expresada como una 2-Tupla lingüística en la Escala Adjetiva de SUS.

- $C_4 \cong$ **Accesibilidad (ACC)**. Esta prueba sólo se realiza por usuarios expertos. En primer lugar, los expertos prueban las alternativas a través de la herramienta en línea WAVE, que enumera los errores y advertencias sobre posibles áreas de mejora del sitio web o la alternativa A_i . A la vista de este informe, el usuario experto U_k que juega un rol R_l asigna una puntuación ($Acc_score_i^{k,l}$) a cada alternativa $A_i, (i = 1, \dots, n)$. Esta puntuación evalúa la accesibilidad de la opción A_i , lo que está en concordancia con los estándares de conformidad. A, AA y AAA con las actuales Pautas de Accesibilidad al Contenido en la Web³. La etiqueta A indica el nivel más bajo, mientras que la etiqueta AAA indica el nivel más alto y, por tanto, la mayor calidad (ver sección 3.2.2.2).

En conclusión, para cada usuario experto U_k que juega el rol R_l , existe un valor $Acc_score_i^{k,l} \in S^3 = \{A, AA, AAA\}$ para cada alternativa $A_i, i = 1, \dots, n$. Según la unificación (Paso 6. en la sección 5.4), el objetivo es evaluar el criterio C_4 con una 2-Tupla lingüística en la escala Adjetiva de SUS.

- $C_5 \cong$ **Emotional Linguistic Usability Test (ELUT)**.

El usuario indica la emoción que ha experimentado en el uso de la alternativa A_i . Se expresa mediante una emoción de las cinco posibles: *Angry*, *Frustrated*, *Indifferent*, *Joy*, y *Happy* que corresponden a los cinco términos lingüísticos de $S^5 = \{s_0, s_1, s_2, s_3, s_4\}$. Para cada alternativa A_i , sea $y_i^{k,l} = \{(f_{i,1}^{k,l}, 0), \dots, (f_{i,g}^{k,l}, 0)\}$ el conjunto de respuestas proporcionadas por el usuario U_k jugando el rol R_l , donde

³WCAG 2.2 <https://www.w3.org/TR/WCAG22>

$h = (1, \dots, 12)$. Se obtiene *ELUT score* del usuario U_k jugando el rol R_l para cada alternativa A_i , ($i = 1, \dots, n$), a través de la media aritmética por medio de la ecuación:

$$ELUT_score_i^{k,l} = \Delta \left(\frac{\sum_{h=1}^{g=12} \Delta^{-1}(f_{i,h}^{k,l}, 0)}{g} \right) = \Delta \left(\frac{1}{g} \sum_{h=1}^{12} \beta_{i,h}^{k,l} \right). \quad (5.6)$$

Paso 8. Construcción de las matrices de decisión individual.

Los resultados de las pruebas realizadas previamente por los usuarios presentan información heterogénea. Para solventar esto, los modelos lingüísticos aportan métodos para interpretar toda la información de forma conjunta. Para ello, en esta fase se construye una matriz para cada usuario U_k que juega un rol R_l que recolecta los resultados de todas las pruebas C_j realizadas, es decir, las valoraciones que proporcionan, representadas mediante una aproximación lingüística. Para cada usuario U_k que juega el rol R_l se construye la matriz de decisión individual $ID^{k,l} = (ID_{ij}^{k,l})_{n \times m}$. Cada elemento corresponde a la información lingüística que se calcula a partir de las evaluaciones originales de los usuarios, atendiendo a cada criterio C_j , $j = 1, \dots, m$ de la siguiente forma:

- $C_1 \cong SUS$. El $SUS_score_i^{k,l}$, ($i = 1, \dots, n$) puntuado por el usuario U_k jugando el rol R_l es disponible. Esas puntuaciones son valores desde 0 hasta 100. Si el usuario que juega el rol no evalúa la alternativa A_i para este criterio, entonces $SUS_score_i^{k,l} = \{\emptyset\}$.

Bangor [45] propone una clasificación en función de adjetivos dentro de la escala SUS. Concretamente, la figura 5.3 muestra la equivalencia de la puntuación SUS en el intervalo $[0, 100]$ a través de un conjunto de términos no balanceados. Enseguida se explica el proceso de equivalencia. Se define el adjetivo SUS como $\{s_0^{sus}, s_1^{sus}, s_2^{sus}, s_3^{sus}, s_4^{sus}, s_5^{sus}, s_6^{sus}\}$ como el conjunto de siete términos lingüísticos no balanceados que denominamos $S_{SUS} = \{None, Worst\ Imaginable, Poor, Ok, Good, Excellent, Best\ Imaginable\} = \{N, WI, P, O, G, E, BI\}$.

Para tratar este conjunto de términos consideramos de forma conjunta la representación de 2-Tuplas y las estructuras lingüísticas jerárquicas. Así, la escala S_{SUS} se construye mediante dos pasos:

- Definir una jerarquía lingüística $LH = \cup_t l(t, n(t))$ (ver sección 2.2.4). Se establece el nivel 1 como $l(1, 3)$ para repartir la escala desde el centro y generamos el siguiente nivel como $l(t + 1, 2n(t) - 1)$. Por lo tanto, el segundo nivel de la jerarquía es $l(2, 5)$. Se establece un tercer nivel $l(3, 9)$ para adaptar todos los términos de S_{SUS} . En otras palabras, en el nivel $t = 1$ de la jerarquía se tiene $n(1) = 3$, en el nivel $t = 2$ se tiene $n(2) = 5$, y en el nivel $t = 3$ se tiene $n(3) = 9$. Esto lleva a establecer la jerarquía $LH = S^3 \cup S^5 \cup S^9$.
- Representar los términos no balanceados de S_{SUS} en LH . Al aplicar el procedimiento planteado en [155], se obtiene que S_{SUS} se representa en LH utilizando las etiquetas lingüísticas de los niveles 2 y 3 de la jerarquía como se muestra en la figura 5.4. Los términos lingüísticos de S_{SUS} corresponden a diferentes niveles de la jerarquía LH . La representación semántica de dichos términos se muestra en la figura 5.5 y corresponde a $N \leftarrow s_0^5$; $WI \leftarrow \overline{s_1^5} \cup \underline{s_2^9}$; $P \leftarrow s_3^9$; $OK \leftarrow \overline{s_4^9} \cup \underline{s_5^9}$; $G \leftarrow \overline{s_6^9} \cup \underline{s_7^9}$; $E \leftarrow s_7^9$; $BI \leftarrow s_8^9$.

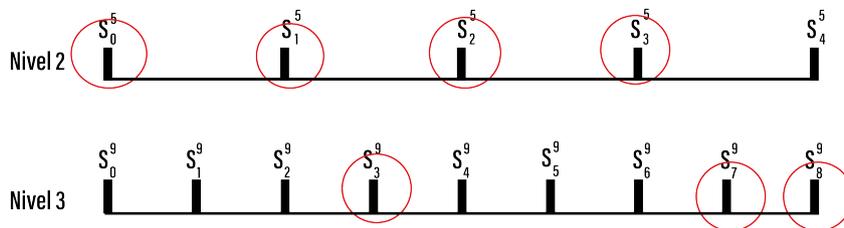


Figura 5.4: Conjunto de términos lingüísticos S_{SUS} de acuerdo a la jerarquía $LH = S^3 \cup S^5 \cup S^9$ marcada con círculos. Fuente: Elaboración propia.

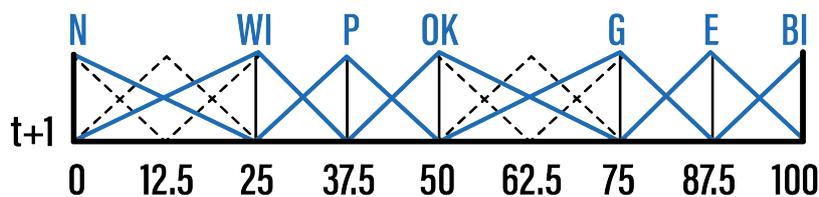


Figura 5.5: Representación semántica de escala S_{SUS} . Fuente: Elaboración propia.

Para operar cómodamente con esta información, se define el conjunto de Componentes de Término (TC) asociado a cada término lingüístico de S_{SUS} como el conjunto de términos que pertenecen a algún nivel de la jerarquía que conforman la semántica del término en cuestión de la siguiente manera:

- $TC(s_0^{sus}) = \{s_0^5\}$
- $TC(s_1^{sus}) = \{s_1^5, s_2^9\}$
- $TC(s_2^{sus}) = \{s_3^9\}$
- $TC(s_3^{sus}) = \{s_4^9, s_2^5\}$
- $TC(s_4^{sus}) = \{s_3^5, s_6^9\}$
- $TC(s_5^{sus}) = \{s_7^9\}$
- $TC(s_6^{sus}) = \{s_8^9\}$

Hasta aquí se ha establecido la estructura jerárquica asociada a S_{SUS} así como la semántica asociada a cada uno de sus términos. Es momento de transformar las puntuaciones $SUS_score_i^{k,l}, i = 1, \dots, n$ de los usuarios a la escala S_{SUS} . La definición 5.1 presenta la función que transforma los valores de puntuación SUS pertenecientes al intervalo $[0, 100]$ a la escala S_{SUS} :

Definición 5.1. *Sea $Score \in [0, 100]$ la puntuación obtenida tras realizar una prueba SUS (generalmente se dispone de calculadora online ⁴). Sea $S_{SUS} = \{s_i^{sus}; i = 0, \dots, 6\}$ la escala lingüística no equilibrada del SUS. Sea $LH = S^3 \cup S^5 \cup S^9$ la jerarquía lingüística asociada a S_{SUS} tal que $S^{n(t)} = \{s_0^{n(t)}, \dots, s_{n(t)-1}^{n(t)}\}, t = 1, 2, 3$. Se obtiene la 2-Tupla que pertenece al nivel t de LH asociada a la $Score$ como:*

$$(s', \alpha') = \Delta \left(\frac{(n(t) - 1)Score}{100} \right) \quad (5.7)$$

donde $t = 2$ si $Score \in [0, 25] \cup [50, 75]$ y $t = 3$ if $Score \in (25, 50) \cup (75, 100]$.

Se define una función de transformación para SUS (TF_{SUS}) que asocia la puntuación SUS con su respectiva 2-Tupla lingüística no balanceada como:

⁴SUS calculator <https://uiuxtrend.com/sus-calculator/>

$$TF_{SUS} : [0, 100] \rightarrow (S_{SUS} \times [-0.5, 0.5]) \quad (5.8)$$

$$TF_{SUS}(Score) = (s_i^{sus}, \alpha') \mid s' \in TC(s_i^{sus})$$

Ejemplo 1. Sea el valor calculado para C_1 , $Score = 65$. Corresponde al nivel $t = 2$ (desde $Score \in [50, 75]$) y, además, su valor 2-Tupla asociado con el nivel 2 de la jerarquía es $\Delta\left(\frac{4 \times 65}{100}\right) = (s_3^5, -0.40)$. En consecuencia de ello, la 2-Tupla lingüística no balanceada asociada es $TF_{SUS}(65) = (s_4^{sus}, -0.40)$, que se entiende como usabilidad = (OK, -0.40).

Se transforman los valores $SUS_score_i^{k,l}, i = 1, \dots, n$ proporcionados por el usuario U_k que juega el rol R_l , en 2-Tuplas lingüísticas asociadas a la escala S_{SUS} aplicando la Definición 5.1 y se agrupan en las matrices $ID^{k,l}$ como se muestra a continuación:

$$ID_{i1}^{k,l} = TF_{SUS}(SUS_score_i^{k,l}) \quad \forall i = 1, \dots, n \quad (5.9)$$

- $C2 \cong NPS$. Los valores NPS $NPS_LTR_score_i^{k,l}$ del usuario U_k jugando el rol R_l se encuentran disponibles. Estas puntuaciones son valores desde 0 hasta 10. Se considera $NPS_LTR_score_i^{k,l} = \{\emptyset\}$ cuando un usuario con el rol no evalúa la alternativa A_i con este test.

Para recoger estos valores en la matriz $ID^{k,l}$, se transforman las puntuaciones no vacías, es decir, los valores LTR obtenidos mediante la prueba NPS, a la escala S_{SUS} . Para ello, realizamos dos pasos:

1. Sauro *et al.* [156] establece una relación para transformar los valores NPS $NPS_LTR_score_i^{k,l}$ hacia valores $NPS_SUS_score_i^{k,l}$ proporcionados por los usuarios a la pregunta planteada en la prueba NPS (valores LTR). El primer enfoque propuesto proporciona la ecuación $LTR = SUS/10$ que permite predecir el valor LTR a través de un valor SUS de forma muy sencilla. Posteriormente, a través de un estudio con más de 2000 usuarios, los autores lograron obtener una mejor aproximación de la ecuación de regresión que relaciona ambos valores como $LTR = 1.33 + 0.08(SUS)$. Se utiliza esta última aproximación por ser más precisa. Por lo tanto,

se establece una correspondencia entre las puntuaciones obtenidas tras responder a la pregunta LTR del test NPS y un test SUS mediante:

$$NPS_SUS_score_i^{k,l} = \frac{NPS_LTR_score_i^{k,l} - 1.33}{0.08} \quad (5.10)$$

Si el valor $NPS_SUS_score_i^{k,l}$ es negativo, entonces se restablece a 0. Si el valor es mayor que 100, se restablece a 100.

2. La transformación de los valores NPST $NPS_SUS_score_i^{k,l}$ hacia la representación lingüística S_{SUS} se recolecta en las matrices $ID^{k,l}$. Se establece la correspondencia entre los valores obtenidos en el test NPS representado con puntuaciones del test SUS hacia la escala lingüística S_{SUS} por medio de la Definición 5.1. Se transforman las puntuaciones $NPS_SUS_score_i^{k,l}, i = 1, \dots, n$ del usuario U_k jugando el rol R_l en una 2-Tupla lingüística asociada con la escala S_{SUS} y se insertan en las matrices $ID^{k,l}$. Específicamente, para cada usuario U_k jugando el rol R_l se completa la segunda columna ($j = 2$) asociada a la matriz $ID^{k,l}$ como se muestra a continuación:

$$ID_{i,2}^{k,l} = TF_{SUS}(NPS_SUS_score_i^{k,l}) \quad \forall i = 1, \dots, n \quad (5.11)$$

- $C_3 \cong UT$. Un usuario U_k jugando el rol R_l y completando el conjunto de tareas, se recolectan los siguientes datos importantes: $Success_score_i^{k,l}(q_v)$ (puntuación de éxito de la tarea), $Efficiency_score_i^{k,l}(q_v)$ (puntuación de la eficiencia en el tiempo de realización de la tarea) y su evaluación de acuerdo a la experiencia al realizar la prueba de usabilidad PU , la valoración del nivel de satisfacción $Satisfaction_score_i^{k,l} \in S^5$. Las primeras dos puntuaciones son asignadas a la tarea q_v e incluidas como porcentaje en el reporte final, mientras que el nivel de satisfacción se considera como una variable lingüística para el test C_3 . Específicamente, para un usuario U_k jugando un rol R_l se completa la tercer columna ($j = 3$) en la matriz asociada $ID^{k,l}$ como se presenta a continuación:

$$ID_{i,3}^{k,l} = (Satisfaction_PU_i^{k,l}, \alpha) \quad \forall i = 1, \dots, n \quad (5.12)$$

- $C_4 \cong ACC$. La puntuación de este test para la alternativa A_i entre las etiquetas

lingüísticas A, AA or AAA es directamente un término lingüístico perteneciente a S^3 y su 2-Tupla asociada. Específicamente, para cada experto U_k jugando un rol R_l se completa la cuarta columna ($j = 4$) de su matriz asociada $ID^{k,l}$ como se presenta a continuación:

$$ID_{i4}^{k,l} = (Acc_score_i^{k,l}, 0) \quad \forall i = 1, \dots, n \quad (5.13)$$

- $C_5 \cong ELUT$. Un usuario U_k jugando el rol R_l y completando el test ELUT, se recolecta la valoración de la emoción que emerge $ELUT_score_i^{k,l} \in S^5$. El resultado emocional se considera como una variable lingüística para el test C_5 . Específicamente, para un usuario U_k jugando un rol R_l se completa la tercer columna ($j = 5$) en la matriz asociada $ID^{k,l}$ como se presenta a continuación:

$$ID_{i5}^{k,l} = (ELUT_score_i^{k,l}) \quad \forall i = 1, \dots, n \quad (5.14)$$

En resumen, se resaltan las escalas de pertenencia lingüística de los elementos de las matrices $ID^{k,l}$, representadas como 2-Tuplas:

- $ID_{i1}^{k,l} \rightsquigarrow S_{SUS}$
- $ID_{i2}^{k,l} \rightsquigarrow S_{SUS}$
- $ID_{i3}^{k,l} \rightsquigarrow S^5$
- $ID_{i4}^{k,l} \rightsquigarrow S^3$
- $ID_{i4}^{k,l} \rightsquigarrow S^5$
- $ID_{i5}^{k,l} \rightsquigarrow S^5$

5.5. Fase 4. Agregación colectiva

En esta fase, las valoraciones individuales de los usuarios que juegan distintos roles que han sido recolectadas en las matrices $ID^{k,l}$ se agregan en matrices colectivas $r + 1$: una matriz para cada rol R_l , ($l = 1, \dots, r$) y una matriz global que agrega las matrices de cada rol.

Paso 9. Unificación de la información hacia S^9 .

Para agregar las matrices individuales es necesario unificar sus valores. Para ello, se transforma cada matriz $ID^{k,l}\{\emptyset\}$ que contiene las valoraciones originales de los usuarios U_k que juegan el rol R_l en una matriz con las valoraciones unificadas en S^9 , el nivel más profundo de la jerarquía LH . Se denominan a estas matrices *Decisiones Individuales Unificadas (UID)*. Dependiendo de la escala lingüística utilizada en la evaluación de cada criterio, habrá que aplicar una transformación u otra. A continuación, se analiza el método a seguir para cada una de estas situaciones.

- C_1 y C_2 . Las evaluaciones están en la escala S_{SUS} . Se utiliza la función de transformación \mathcal{LH} (véase la sección V.A de [43]) que permite transformar 2-Tuplas de una escala lingüística no balanceada, como S_{SUS} , en una jerarquía lingüística, como $LH = S^3 \cup S^5 \cup S^9$. En específico, se establece $\mathcal{LH}:(S_{SUS} \times [-0, 5, 0, 5]) \rightarrow (LH \times [-0, 5, 0, 5])$.

Después de esta conversión, las evaluaciones lingüísticas se expresan en diferentes dominios lingüísticos, lo que significa que no se pueden procesar directamente. Necesitamos la función de transformación \mathcal{TF} (véase la sección V.B de [43]) para convertir las 2-Tuplas de diferentes dominios en un conjunto de etiquetas de granularidad particular de LH . Se establece $\mathcal{TF}_3^t:(LH \times [-0, 5, 0, 5]) \rightarrow (S^9 \times [-0, 5, 0, 5])$ como una función especial que integra un conjunto de funciones de transformación TF [27] entre niveles de LH hasta el nivel más alto de LH . En este caso, por la definición de SUS, las 2-Tuplas transformadas obtenidas pueden pertenecer tanto a S^5 como a S^9 obteniendo:

$$UID_{ij}^{k,l} = \mathcal{TF}_3^t(\mathcal{LH}(ID_{ij}^{k,l})) \quad \forall j = 1, 2; i = 1, \dots, n \quad (5.15)$$

con t siendo el nivel 2 o 3 de la jerarquía. Si $ID_{ij}^{k,l} = \{\emptyset\}$, entonces $UID_{ij}^{k,l} = \{\emptyset\}$.

- C_3 , C_4 y C_5 . Las valoraciones se encuentran en la escala S^5 , S^3 y S^5 respectivamente. Se utiliza la función de transformación TF [27], que permite transformar una 2-Tupla entre cualquier nivel de la jerarquía lingüística (ver sección 2.2.3) en LH . Específicamente, la tercer y cuarta columna de las matrices $UID^{k,l}$ son obte-

nidas transformando la 2-Tuplas de las matrices $ID^{k,l}$ hacia S^9 como se detalla a continuación:

$$UID_{ij}^{k,l} = TF_3^t(ID_{ij}^{k,l}) = \Delta \left(\frac{\Delta^{-1}(ID_{it}^{k,l}) \times 8}{n(t) - 1} \right) \quad \forall j = 3, 4; i = 1, \dots, n \quad (5.16)$$

con t siendo el nivel 1 o 2 de la jerarquía. Si $ID_{ij}^{k,l} = \{\emptyset\}$, entonces $UID_{ij}^{k,l} = \{\emptyset\}$.

Paso 10. Agregación para cada rol.

Se define la matriz de decisión colectiva unificada para el rol R_l (UCD^l) que contiene las decisiones colectivas unificadas en S^9 incluyendo las decisiones individuales unificadas de todos los usuarios con rol R_l . Para ello, se tiene en cuenta el operador *2-Tuple weighted average* (2TWA, ver sección 2.2.2). Se obtiene (UCD^l) como la agregación de matrices no vacías $UID^{k,l}$, ($k = 1, \dots, u$) mediante el operador 2TWA. Cada elemento UCD_{ij}^l , ($i = 1, \dots, n; j = 1, \dots, m$) se define como:

$$\begin{aligned} UCD_{ij}^l &= 2TWA_{W^l}(UID_{ij}^{1,l}, \dots, UID_{ij}^{u,l}) \\ &= \Delta \left(\frac{\sum_{k=1}^u \Delta^{-1}(UID_{ij}^{k,l}) \times W_k^l}{\sum_{k=1}^u W_k^l} \right) = (s_{ij}^l, \alpha_{ij}^l) \end{aligned} \quad (5.17)$$

donde los elementos del vector de pesos $W^l = (W_1^l, \dots, W_u^l)$ para el rol R_l son definidos por $W_k^l = \frac{W_k^{l'}}{\sum_{k=1}^u W_k^{l'}}$, $k = 1, \dots, u$, tal que:

$$W_k^l = \begin{cases} WU_k & Si \quad UID^{k,l} \neq \{\emptyset\} \\ 0 & Si \quad UID^{k,l} = \{\emptyset\} \end{cases} \quad (5.18)$$

A continuación se agrega con base en el peso de los criterios. Este proceso genera un vector unificado de decisión colectiva (\mathbf{ucd}^l) para cada rol R_l , que es utilizado en el informe de usabilidad. La siguiente expresión lo calcula:

$$\mathbf{ucd}_i^l = \Delta \left(\Delta^{-1}(s_{ij}^l, \alpha_{ij}^l) \times WC_j \right). \quad (5.19)$$

Paso 11. Agregación Global

La matriz de Decisión Colectiva Unificada $UCD^{global} n \times m$ que contiene las decisiones colectivas en S^9 es definida por la agregación de las decisiones colectivas de cada rol R_l . Para este propósito, el operador 2TWA es aplicado a las agregaciones no vacías en las matrices UCD_{ij}^l , ($l = 1, \dots, r$) a través del operador 2TWA. Cada elemento de UCD_{ij}^{global} , ($i = 1, \dots, n$; $j = 1, \dots, m$) esta definido por la ecuación 5.20:

$$\begin{aligned} UCD_{ij}^{global} &= 2TWA_{WR}(UCD_{ij}^1, \dots, UCD_{ij}^r) \\ &= \Delta \left(\sum_{l=1}^r \Delta^{-1}(UCD_{ij}^l) \times WR_l \right) = (s_{ij}^{global}, \alpha_{ij}^{global}) \end{aligned} \quad (5.20)$$

donde WR es el vector de pesos normalizado de roles, donde $\sum_{l=1}^r WR_l = 1$. Enseguida, la agregación se realiza con base en la importancia de los criterios. Este proceso da lugar a la generación de un vector de decisión colectiva global unificada (\mathbf{ucd}^{global}), el cual se utiliza en el informe de usabilidad.

$$\mathbf{ucd}_i^{global} = \Delta \left(\Delta^{-1}(s_{ij}^{global}, \alpha_{ij}^{global}) \times WC_j \right). \quad (5.21)$$

5.6. Fase 5. Explotación de datos

Se aplica el método TOPSIS (explicado en la sección 2.3.2) sobre las matrices de Decisiones Colectivas Unificadas para generar varios rankings de las alternativas, permitiendo derivar un ranking para roles específicos y un ranking general.

Así, el modelo construye $r + 1$ rankings: uno para cada rol R_l basado en las matrices UCD^l , ($l = 1, \dots, r$) y un ranking global basado en la matriz UCD^{global} . La clasificación de las alternativas se establece en función del coeficiente de proximidad relativa a la alternativa ideal. A medida que el coeficiente aumenta, la alternativa A_i mejora en su evaluación. Por lo tanto, todas las alternativas A_i (donde $i = 1, 2, \dots, m$) pueden ser ordenadas en un ranking descendente basado en los valores de cercanía relativa.

Paso 12. Generación de rankings para cada rol

La técnica TOPSIS se aplica sobre las matrices UCD^l , ($l = 1, \dots, r$) cuyos valores son 2-Tuplas $(s_{ij}^l, \alpha_{ij}^l)$. Por tanto, se establece $(s_{ij}, \alpha_{ij}) = (s_{ij}^l, \alpha_{ij}^l)$ para obtener r rankings, uno por cada rol, denotado por $Ranking^l$, ($l = 1, \dots, r$).

Paso 13. Generación del ranking global

La técnica TOPSIS es aplicada sobre la matriz UCD^{global} cuyos valores son 2-Tuplas $(s_{ij}^{global}, \alpha_{ij}^{global})$. Además, se establece $(s_{ij}, \alpha_{ij}) = (s_{ij}^{global}, \alpha_{ij}^{global})$ para obtener el ranking global que se denota como $Ranking^{global}$.

Paso 14. Retranslación.

Los elementos de los vectores \mathbf{ucd}^l , ($l = 1, \dots, r$), y \mathbf{ucd}^{global} son 2-Tuplas en S^9 . Este paso lleva la retranslación de esos valores hacia los valores en S_{SUS} , con el objetivo de ofrecer términos lingüísticos más comprensibles para los usuarios. Por lo tanto, no es necesario llevar a cabo este paso para implementar la metodología LDM4WUE, pero es conveniente para incluir en el informe de usabilidad.

Se construye un vector *adjective usability report* (\mathbf{aur}^l) para cada rol R_l y un vector *adjective usability report global* (\mathbf{aur}^{global}) que contiene la información de los vectores \mathbf{ucd}^l y \mathbf{ucd}^{global} , respectivamente, como términos lingüísticos de S_{SUS} . En primer lugar, se define una función de identidad id para transformar términos lingüísticos 2-Tupla de S^9 a la jerarquía lingüística $LH = S^3 \cup S^5 \cup S^9$ por $id(s, \alpha) = (s, \alpha)$. Particularmente, se establece $id:(S^9 \times [-0.5, 0.5]) \rightarrow (LH \times [-0.5, 0.5])$. Enseguida, se utiliza la función de transformación inversa $\mathcal{L}\mathcal{H}^{-1}$ (véase la sección V.A de [43]) que permite transformar la 2-Tupla lingüística expresada en una jerarquía lingüística, como LH , en un conjunto de términos lingüísticos no balanceados, como S_{SUS} . En particular, se establece $\mathcal{L}\mathcal{H}^{-1}:(LH \times [-0.5, 0.5]) \rightarrow (S_{SUS} \times [-0.5, 0.5])$. Se calculan los vectores \mathbf{aur}^l , ($l = 1, \dots, r$) y \mathbf{aur}^{global} ($i = 1, \dots, n; j = 1, \dots, m$) mediante las ecuaciones

$$\mathbf{aur}_i^l = \mathcal{L}\mathcal{H}^{-1}(id(\mathbf{ucd}_i^l)) \quad \forall l = 1, \dots, r \quad (5.22)$$

$$\mathbf{aur}_i^{global} = \mathcal{L} \mathcal{H}^{-1}(id(\mathbf{ucd}_i^{global})) \quad (5.23)$$

5.7. Resumen de capítulo

- El modelo descrito de Toma de Decisiones Lingüístico para la evaluación de la usabilidad, está centrada en el usuario, permitiendo obtener un mejor acercamiento a la evaluación de la usabilidad.
- La utilización de la escala SUS, permite obtener un esquema resolutivo basado en: la estructuración del problema, elicitación de información de los usuarios, unificación, agregación colectiva, explotación y retransalación.
- En la aplicación de la metodología LDM4WUE se establece al figura de moderador, quien coordina la definición de los elementos como alternativas, criterios, evaluadores (usuarios finales y expertos), roles y ponderaciones, dentro del proceso de evaluación.
- Los criterios y la forma de recolección de la información, de los usuarios finales y expertos, para la evaluación de las alternativas, se realizó por medio de la prueba SUS, NPS, ACC, PU y ELUT.
- La prueba ACC fue realizada solo por expertos, ya que se encarga de evaluar la accesibilidad de las alternativas evaluadas.
- La utilidad de la aplicación de LDM4WUE para la evaluación de la usabilidad se amplifica si se implementa y se pone a disposición de diseñadores y expertos en usabilidad.

Capítulo 6

Herramienta web *USE-AB-Tool* *Decision Support System*

La creación de un software destinado a la evaluación de la usabilidad a través de la recopilación de las opiniones de los usuarios se justifica como una respuesta directa a la creciente importancia de la experiencia del usuario en el entorno digital. En un mundo donde la competencia por la atención y la satisfacción del cliente es cada vez más intensa, entender y mejorar la usabilidad de un software se vuelve fundamental para su éxito. Al permitir que los usuarios expresen sus opiniones, este tipo de software no solo brinda información lingüística valiosa para identificar áreas de mejora, sino que también fomenta la participación del usuario en el proceso de desarrollo, fortaleciendo así la relación entre el equipo de desarrollo y el usuario final. La retroalimentación directa y caracterización de las opiniones de los usuarios facilita la creación de productos más intuitivos, eficientes y adaptados a las necesidades reales de quienes los utilizan.

Este capítulo se conforma con las siguientes secciones. La sección 6.1, describe las características esenciales para desarrollar un software de *A/B Testing*. La sección 6.2, detalla las herramientas utilizadas para la implementación de la herramienta web. La sección 6.3, enumera los pasos necesarios para evaluar la usabilidad mediante *USE-AB-Tool*. Finalmente, la sección 6.4 presenta los puntos mas importantes de capítulo.

6.1. Diseño de una herramienta para A/B testing

Para el desarrollo de un software exitoso, es crucial considerar una variedad de aspectos que abarcan desde la planificación inicial hasta el lanzamiento y mantenimiento continuo. Uno de los pasos para el desarrollo de un software es la recolección de los requisitos de usuario, donde se comprenden las necesidades y expectativas de las personas que harán uso de este, diseñado con base a una metodología centrada en el usuario, la cual implica utilizar la UX como una disciplina fácil e intuitiva para mejorar la experiencia y aumentar el uso del sistema. Una metodología centrada en el usuario como *Design Thinking*, conlleva fases donde se involucra al usuario final (ver sección 3.1).

Por otra parte, es importante diseñar de una arquitectura de software sólida [157], pensando en un proyecto integral que garantice aspectos como: escalabilidad, seguridad, eficiencia y eficacia del software. Por lo tanto, elegir patrones de diseño correcto, así como tecnologías de diseño actuales y estandarizadas, impacta positivamente al momento de implementar el sistema.

De igual manera, al identificar los tipos de información que el sistema maneja, da una pauta para la elección correcta del nivel de seguridad [158] que debe tener, desde la identificación de usuarios, así como el cifrado de datos y, la prevención de ciber-ataques, son atributos que se deben atender.

Además, se debe pensar en la escalabilidad del software [157], que es la capacidad del sistema para crecer y adaptarse eficientemente a medida que aumentan las demandas del usuario y los volúmenes de datos. Se trata de diseñar el software de forma que pueda manejar un aumento en la carga de trabajo o en los recursos sin comprometer su rendimiento o funcionalidad, permitiendo así una expansión suave y rentable a lo largo del tiempo.

Asimismo, un software con calidad [112] debe evitar errores, por lo tanto, se implementan pruebas en cada etapa del desarrollo para detectar y corregir errores. De tal manera, se deben utilizar pruebas unitarias, de integración, de sistema y de aceptación de usuarios.

Por último, es necesario anticipar el mantenimiento constante del software con el fin de rectificar errores, implementar nuevas funcionalidades y modernizar tecnologías, además de desarrollar procedimientos y utilidades para administrar de manera eficiente

las modificaciones y los avances.

Entonces, la implementación de *USE-AB-Tool* esta fundamentado en las mejores prácticas de la ingeniería de software para garantizar la eficacia y la fiabilidad de los resultados obtenidos. Al adherirse a un enfoque metodológico sólido, que incluya la planificación detallada, el diseño robusto, la implementación cuidadosa y el mantenimiento continuo, se establece una base sólida para el aseguramiento de la usabilidad y, por ende, la satisfacción del usuario final.

6.2. Aspectos técnicos de *USE-AB-Tool*

La arquitectura de software es el proceso de diseñar y estructurar los componentes de un sistema para satisfacer los requisitos funcionales y no funcionales. Se trata de una disciplina importante en el desarrollo de software, que se enfoca en la organización de los diferentes elementos de un sistema y en cómo estos interactúan entre sí para producir un resultado final.

Como se demuestra anteriormente, la arquitectura de software abarca un conjunto de principios y patrones que guían el diseño y la estructura de los sistemas informáticos. Dentro de este amplio espectro, el modelo cliente-servidor emerge como una de las estructuras más fundamentales y extendidas. Incluye componentes como servidores, bases de datos, aplicaciones cliente y protocolos de comunicación. Esta arquitectura proporciona una estructura organizada para el desarrollo y la implementación de aplicaciones web, facilitando la escalabilidad, la mantenibilidad y la eficiencia del sistema.

Algunas ventajas de utilizar la Arquitectura Web son:

- **Escalabilidad:** Permite escalar vertical y horizontalmente según las necesidades del sistema, lo que facilita la gestión del crecimiento de la demanda de usuarios y el volumen de datos.
- **Mantenibilidad:** Separa claramente las responsabilidades entre los componentes de la arquitectura web facilita la identificación y resolución de problemas, así como la implementación de actualizaciones y mejoras sin afectar a otros componentes del sistema.

- **Servidores web:** Funciona para alojar y servir los archivos de la aplicación web, así como para procesar solicitudes de los clientes.
- **Bases de datos:** Sirve para almacenar y gestionar la información necesaria para la aplicación web.
- **Frameworks de desarrollo web:** Herramientas como Django¹, Ruby on Rails², Laravel³, Vue⁴, entre otros, que proporcionan estructuras y funcionalidades predefinidas para facilitar el desarrollo de aplicaciones web.

Por lo tanto, la arquitectura web es fundamental para el desarrollo exitoso de aplicaciones web modernas. En consecuencia, basado en los principios de la arquitectura web, se eligieron las siguientes tecnologías para el desarrollo de *USE-AB-Tool*:

- **Java Script**⁵: por su ubicuidad, versatilidad, capacidad para crear interfaces de usuario interactivas, amplia gama de bibliotecas y marcos de trabajo, flexibilidad, comunidad activa y evolución continua.
- **JS Vue**⁶: por su facilidad de aprendizaje, flexibilidad, reactividad, sistema de componentes, buena documentación, comunidad activa, integración con otras tecnologías y rendimiento sólido.
- **MySQL**⁷: por su fiabilidad, rendimiento, escalabilidad, facilidad de uso, compatibilidad, comunidad activa y soporte.
- **PHP**⁸: por su facilidad de aprendizaje, amplia disponibilidad de recursos y bibliotecas, integración con bases de datos, flexibilidad, gran comunidad y soporte, rendimiento adecuado y compatibilidad/portabilidad.

¹<https://www.djangoproject.com/>

²<https://rubyonrails.org/>

³<https://laravel.com/>

⁴<https://vuejs.org/>

⁵<https://www.javascript.com/>

⁶<https://vuejs.org/>

⁷<https://www.mysql.com/>

⁸<https://www.php.net/>

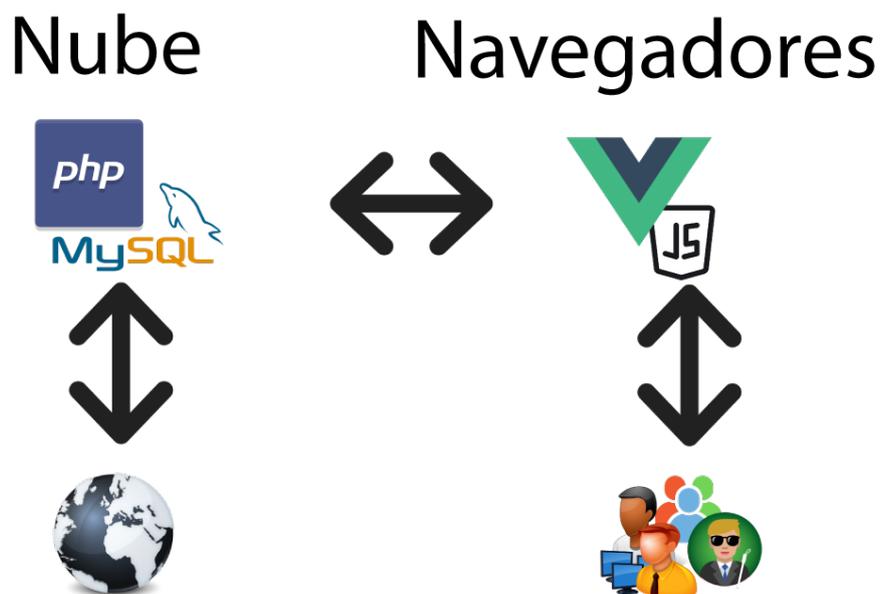


Figura 6.1: Modelo Aplicación-Servidor utilizado para el desarrollo de *USE-AB-Tool*. Fuente: Elaboración propia.

6.3. Manual de uso para LDM4USE

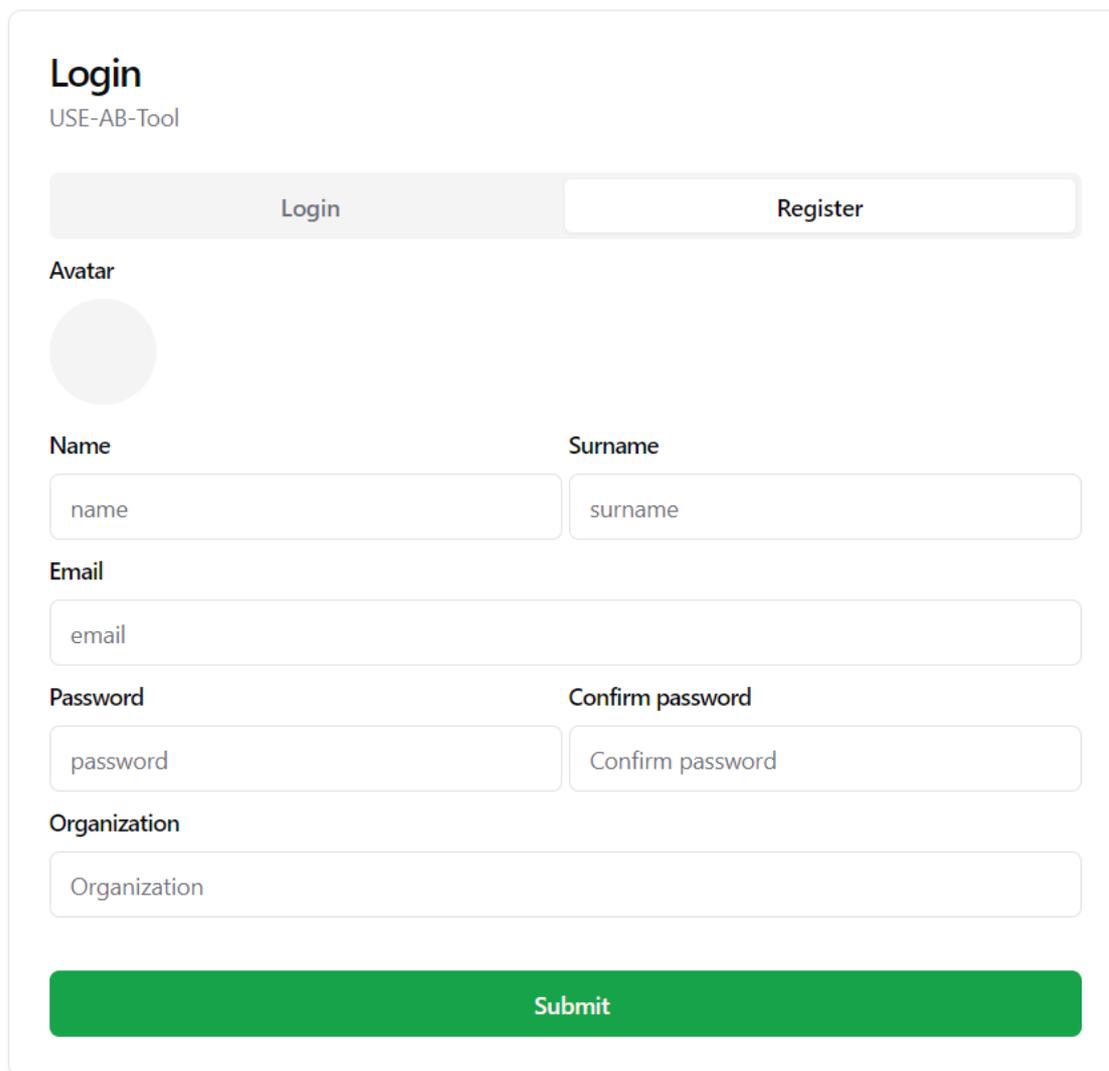
En las siguientes secciones, se explican las funcionalidades de la herramienta *USE-AB-Tool* en tres categorías: 1) añadirse como usuario de sistema, 2) crear un proyecto de evaluación de la usabilidad y, 3) contestar los test en el proyecto de evaluación de la usabilidad.

6.3.1. Ingreso al sistema

A continuación, se describen los pasos para dar de alta un usuario nuevo en *USE-AB-Tool*.

- **Registro de usuario:** Para poder hacer uso del sistema *USE-AB-Tool* ya sea como moderador, experto o usuario final, se debe estar dado de alta en sistema (ver figura 6.2). La página principal, se encuentra con los campos para acceso al sistema. En la sección de *Log in*, un usuario nuevo debe dar clic en *Sing up*, ubicado en la parte inferior del cuadro de *Log in*, enseguida, en la página de registro, se debe subir una fotografía en formato de imagen (.jpg, .jpeg, .png) así como llenar el formulario

que se presenta. Es de importante escribir un correo electrónico válido, ya que este servirá como *username* para ingresar al sistema.



The image shows a user registration form titled "Login" for "USE-AB-Tool". At the top, there are two buttons: "Login" (disabled) and "Register" (active). Below this is an "Avatar" section with a circular placeholder. The form consists of several input fields: "Name" and "Surname" (two columns), "Email" (one column), "Password" and "Confirm password" (two columns), and "Organization" (one column). A large green "Submit" button is at the bottom.

Login
USE-AB-Tool

Login Register

Avatar

Name Surname

name surname

Email

email

Password Confirm password

password Confirm password

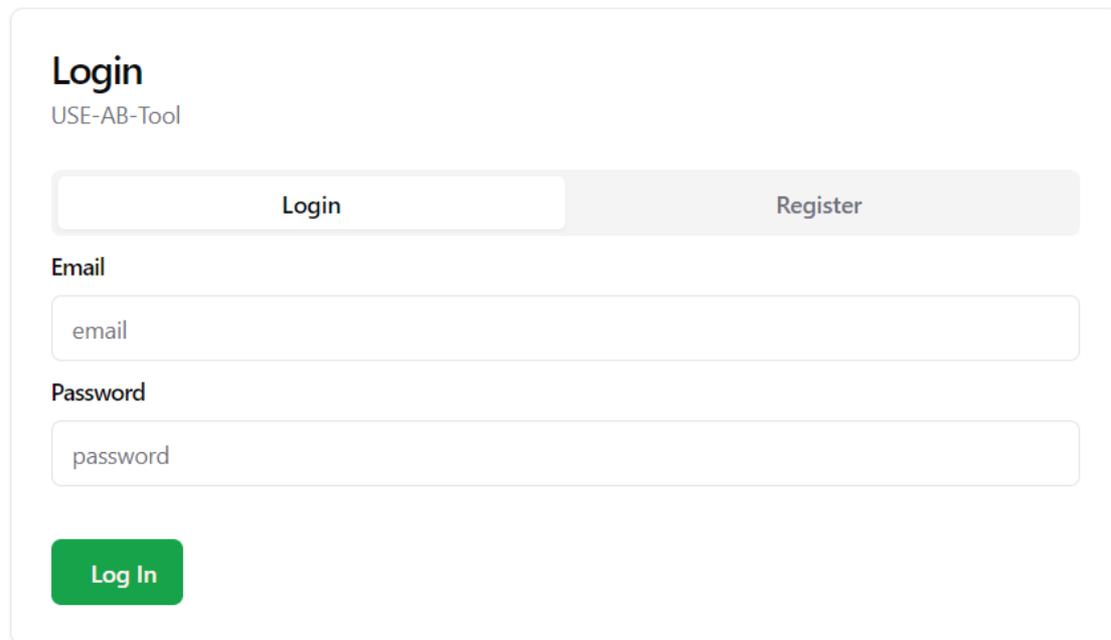
Organization

Organization

Submit

Figura 6.2: Formulario de registro de usuario. Fuente: Elaboración propia.

- **Ingreso al sistema:** El ingreso de sistema se realiza con el *username* y el *Password*, correspondiente al *email* y la contraseña que se ingreso en el paso de **Registro de usuario** (ver figura 6.3).



The screenshot shows a login form titled "Login" for the "USE-AB-Tool". At the top, there are two buttons: "Login" (active) and "Register". Below these are two input fields: "Email" with the placeholder text "email" and "Password" with the placeholder text "password". At the bottom of the form is a green "Log In" button.

Figura 6.3: Vista de *Log in*. Fuente: Elaboración propia.

- **Página de inicio:** Es la página principal que se muestra a los usuarios cuando acceden *USE-AB-Tool*. Esta página contiene información general sobre la aplicación web. El menú principal contiene las siguientes opciones: *Profile, Projects, Test, Users, Roles, Reports* (ver figura 6.4).



Figura 6.4: Vista de inicio. Fuente: Elaboración propia.

- ***Projects*:** En esta sección se encuentran los proyectos de usabilidad que en los que el usuario *logueado* es participe. En la esquina superior derecha, se encuentra el botón *+ Add project* (ver figura 6.5), el cual dirige a un asistente paso a paso para crear un proyecto nuevo acorde a las fases y pasos en del modelo LDM4USE en la sección 5.1. En esta misma sección se puede editar atributos de los proyectos previamente establecidos así como la eliminación de cada uno de ellos.



Figura 6.5: Vista de la sección *Projects*. Fuente: Elaboración propia.

- **Test:** Esta sección de *USE-AB-Tool* incluye información relevante y detallada sobre cómo se realizan los cuestionarios estandarizados para la evaluación de la satisfacción del usuario y la prueba de usabilidad. Entre los elementos que forman parte de esa sección: Nombre del test o cuestionario estandarizado, preguntas o actividades que lo conforman, tipo de evaluación que integra (ver figura 6.6).

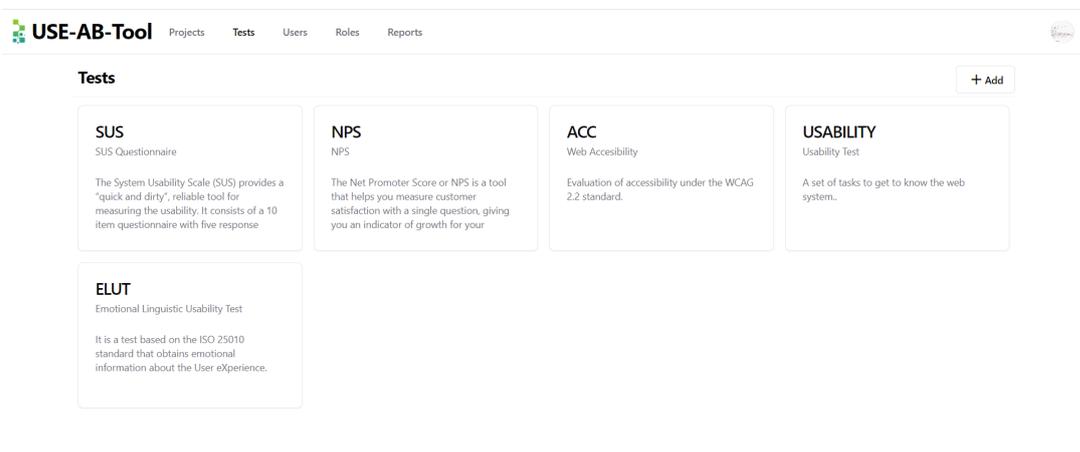


Figura 6.6: Página de la sección *Test*. Fuente: Elaboración propia.

- **Reports:** En esta sección, se encuentra un resumen de los proyectos de usabilidad, que principalmente especifica la cantidad de expertos, usuarios y criterios por los cuales fue evaluado ese proyecto. El detalle del informe se presenta en el apartado 6.3.4 de este capítulo.

6.3.2. Creación de un proyecto de evaluación de usabilidad

En esta sección de *SE-AB-Tool*, se configura un proyecto de evaluación de la usabilidad. Al instante que un usuario da clic en el botón de *+Add project*, se comienza a configurar un nuevo proyecto de evaluación de la usabilidad a través de un asistente paso a paso como se detalla a continuación:

- **Paso preliminar. Configuración básica:** En esta sección se establecen las configuraciones iniciales del proyecto, las cuales se describen a continuación (ver figura 6.7).
 - **Logo:** Es representación visual distintiva de la identidad del proyecto de usabilidad, se sube un archivo de tipo imagen.
 - **Nombre:** Es un texto corto que refleja la naturaleza y el enfoque del proyecto de manera clara y concisa.
 - **Fecha:** Determina el periodo que el proyecto de evaluación se mantiene activo para que los expertos y usuarios finales emitan sus respuestas para cada alternativa sobre el conjunto de criterios.

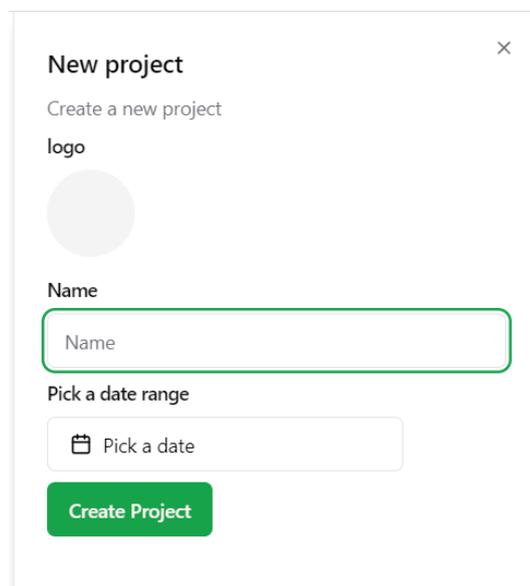


Figura 6.7: Vista para crear un proyecto nuevo. Fuente: Elaboración propia.

- **Paso 1: Definición del conjunto de alternativas.** En este paso, el moderador ingresa cada una de las alternativas que se evalúan con el modelo LDM4USE. Para

cada alternativa los campos necesarios son: logo, alias y url. Una vez agregadas las alternativas a la lista, el *check* en el encabezado **Visible** sirve para habilitar o deshabilitar esa alternativa en los cálculos del informe final de usabilidad (ver figura 6.8). Al menos dos alternativas deben ser seleccionadas.

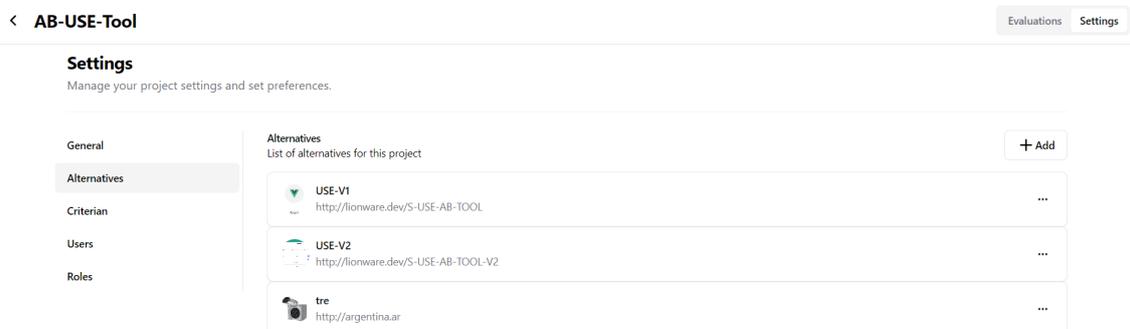


Figura 6.8: Configuración de las alternativas. Fuente: Elaboración propia.

- **Paso 2: Definición del conjunto de criterios y su importancia.** En este paso, el moderador selecciona los test con los que desea evaluar este proyecto del conjunto de test que están dados de alta previamente en el sistema (ver figura 6.9). Al menos dos test deben ser seleccionados.

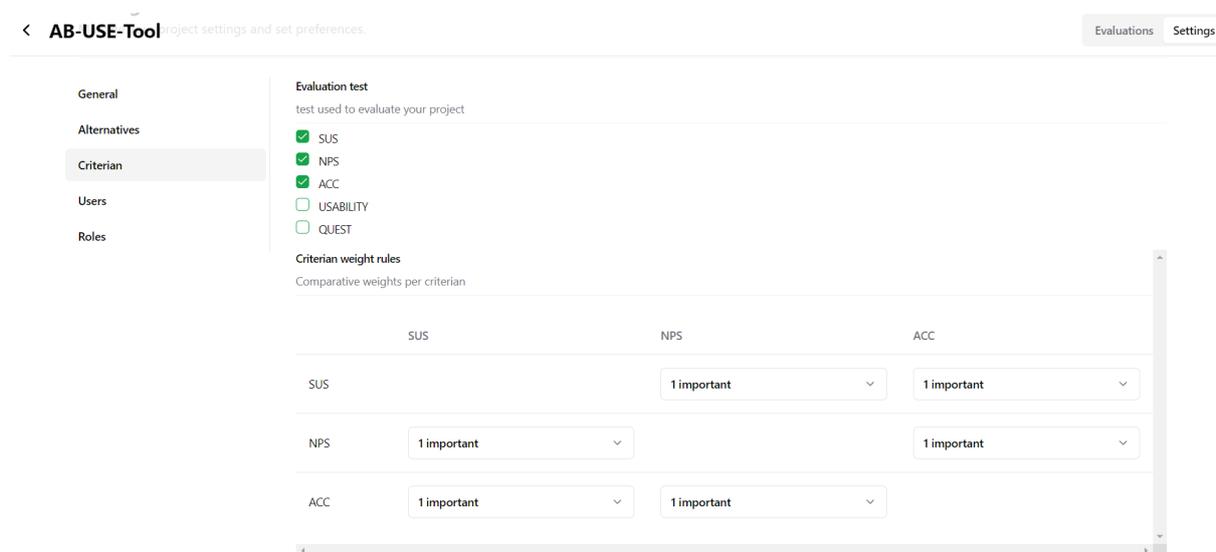


Figura 6.9: Configuración de los criterios. Fuente: Elaboración propia.

Enseguida, se debe determinar la importancia de cada criterio, el moderador elige la importancia de un criterio sobre otro al elegir en la matriz el nivel de importancia

que tiene el criterio en la fila sobre el criterio en la columna. Automáticamente, *USE-AB-Tool* determina el valor de campo inverso (ver figura 6.10). El cálculo del peso de cada criterio debe tener un valor de consistencia $< .10$, en caso de ser verdadero, el texto mostrado como $CI =$ se resalta en color verde, y el botón de *Next*, en caso contrario el campo de texto $CI =$, se resalta en color rojo y el botón *Next* se encuentra inhabilitado.

Por defecto, todos los criterios tienen igual importancia a menos que el moderador configure lo contrario, en caso de querer regresar al valor inicial, se debe dar clic en el botón *reset*.

< Usability Evaluations Settings

Settings
Manage your project settings and set preferences.

General
Alternatives
Criteria
Users
Roles

Evaluation test
test used to evaluate your project

- SUS
- NPS
- ACC
- USABILITY
- ELUT

Criterion weight rules
Comparative weights per criterion

	SUS	NPS	PU	ACC	ELUT
SUS		Moderate...	Equally important	Moderate...	Equally important
NPS	Moderate...		Equally important	Equally important	Moderate...
PU	Equally important	Equally important		Moderate...	Equally important
ACC	Moderate...	Equally important	Moderate...		Moderate...
ELUT	Equally important	Moderate...	Equally important	Moderate...	

Figura 6.10: Configuración de la importancia de los criterios. Fuente: Elaboración propia.

- **Paso 3: Definición del conjunto de usuarios y su importancia.** En este paso se determina si, para el proyecto que se está configurando, existen expertos y usuarios finales (ver figura 6.11). Se debe agregar los usuarios que se desea invitar como evaluadores al dar clic en el botón *Add users*, esta función busca tanto por nombre como por *email* registrado, al seleccionar el usuario se debe elegir el tipo de usuario con el cual realiza la evaluación (*User/Expert*). La evaluación de la importancia se encuentra en un valor eterno positivo entre 0 y 100. Por defecto, se presume que

se considera con el apoyo de los dos tipos de usuario, por lo tanto, se les asigna la importancia de 100 a expertos y 90 a usuarios finales. Sin embargo, la importancia de las opiniones de cada tipo de usuario se puede configurar por el moderador según las características y necesidades del proyecto.

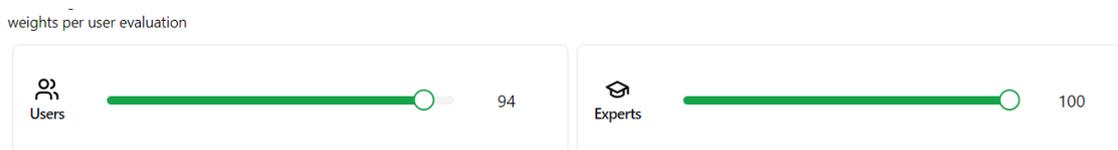


Figura 6.11: Configuración del peso de la opinión de expertos y usuarios finales. Fuente: Elaboración propia.

- Paso 4: Definición del conjunto de roles y su importancia.** El motor de *USE-AB-Tool* es centrado en el usuario, por lo tanto se determinó un conjunto de roles, de los cuales el moderador puede elegir para obtener una evaluación caracterizada. Por defecto, *USE-AB-Tool* tiene en su sistema cuatro roles: visual, motriz, auditivo y oral, cada uno de ellos sub-categorizado en discapacidades: permanentes, temporales y situacionales (ver figura 6.12).

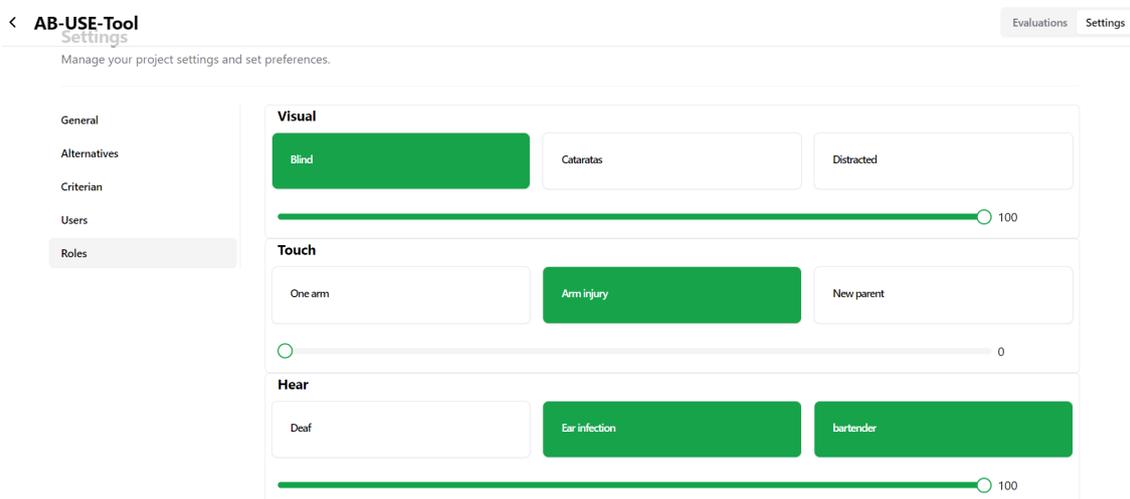


Figura 6.12: Configuración de los roles para la evaluación de la usabilidad. Fuente: Elaboración propia.

Una vez que el moderador eligió los roles que serán elegidos para la evaluación del proyecto, el siguiente paso es asignarles la importancia a cada uno. La importancia de cada rol está definida por un valor numérico entero entre 0 y 100. Por defecto,

todos los roles están configurados con importancia de 100, lo que significa que cada uno de ellos es igual de importante que el resto. El moderador puede configurar ese parámetro dependiendo de las características y necesidades del proyecto (ver figura 6.13).



Figura 6.13: Configuración de la importancia de los roles. Fuente: Elaboración propia.

6.3.3. Realización de los test por los evaluadores

En la siguiente sección se detallan los pasos que, un usuario final o experto, deben seguir para responder los cuestionarios de manera efectiva dentro de las posibles elecciones en cada una de las pruebas.

- **Paso 1 Ingreso al sistema:** Al ingresar al sistema en la *Página de inicio*, se encuentra el menú proyectos, dentro del cual se encuentran los proyectos de los cuales se es participe para evaluación.
- **Paso 2 Selección de proyecto:** Para iniciar o continuar con la evaluación de un proyecto, se da clic en la tarjeta del proyecto que se desea evaluar. Al ingresar, aparecen las pruebas con los que se pretende evaluar ese proyecto que previamente fueron seleccionados por el moderador.
- **Paso 3 Selección de prueba:** Para iniciar o continuar con la evaluación de un proyecto, se da clic en la tarjeta del proyecto que se desea evaluar. Al ingresar, aparecen las pruebas con los que se pretende evaluar ese proyecto que previamente fueron seleccionados por el moderador. Para iniciar la evaluación de una prueba, se debe dar clic en la tarjeta relativa a dicha prueba, enseguida se abre una ventana modal con las preguntas que se deben responder, a continuación, se explica brevemente como contestar cada una de las pruebas:

1. **Prueba de usabilidad:** Este tipo de prueba esta diseñada para el reconocimiento de la web que se evalúa. Esta compuesta por un conjunto de tareas en

las que al finalizar cada una se debe recolectar tres respuestas: 1) *Time*: es el tiempo en segundos que el usuario necesita para realizar la tarea; 2) *Success*: el usuario debe elegir entre las opciones *Done* en caso de finalizar la tarea con éxito, *Not done*, en caso contrario; 4) *Satisfaction*: el usuario debe elegir entre el conjunto de términos lingüísticos { *unsatisfied*, *dissatisfied*, *indifferent*, *satisfied*, and *very satisfied* }.

2. **Cuestionario SUS**: Este cuestionario estandarizado se contesta posterior al reconocimiento total del sistema. Cuando el usuario da clic sobre esta tarjeta, se abre una ventana modal con diez preguntas, cada una se responde eligiendo los números entre 1 y 5.
3. **Cuestionario NPS**: Esta pregunta estandarizada se contesta posterior al reconocimiento total del sistema. Cuando el usuario da clic sobre esta tarjeta, se abre una ventana modal con una sola pregunta, la cual se contesta seleccionando un número entre 0 y 10.
4. **Accesibilidad**: Esta pregunta está restringida solo a usuarios de tipo **Experto**. Se contesta posterior al reconocimiento total del sistema. Cuando el usuario da clic sobre esta tarjeta, se abre una ventana modal con una sola pregunta sobre el nivel de accesibilidad que se considera que tiene la web, posterior a una inspección de accesibilidad con alguna herramienta autorizada por la W3C. Se contesta seleccionando una etiqueta entre el conjunto { *A*, *AA*, *AAA* }.
5. **ELUT**: Este test de propia autoría, está alineado a la Norma ISO 25010. Cuando el usuario da clic en esta tarjeta, se abre una ventana modal con 12 preguntas y el usuario debe elegir una sola respuesta entre el siguiente conjunto de opciones { *Enojado*, *Frustrado*, *Indiferente*, *Alegre*, *Feliz* }.

6.3.4. Obtención del informe de usabilidad

Entre las propuestas de *AB-Testing-DSS* es un informe de usabilidad altamente interpretable. De tal manera que los resultados obtenidos en la evaluación de los sitios web son mediante etiquetas lingüísticas en escala SUS (ver figura 5.3). El informe de usabilidad se compone de dos partes: informe general de alternativas y criterios e informe de la Prueba de Usabilidad que contiene la eficiencia, eficacia y satisfacción por cada tarea que conforma

la PU.

La figura 6.14 contiene el informe general de la evaluación de tres alternativas a través de un conjunto de test (SUS, NPS, PU, ACC, ELUT). Cada alternativa esta representada con una tabla en donde se identifica en la parte superior el nombre del sitio web, enseguida la posición que la web obtuvo entre el total de sitios evaluados. Las siguientes filas están compuestas por los test y separados por rol, de tal manera que se puede identificar la calificación obtenida por test-rol para cada alternativa.

Las calificaciones están expresadas como una 2-Tupla, donde el primer argumento es la etiqueta lingüística dentro del conjunto de adjetivos de SUS, y el segundo argumento es un valor llamado α que representa la distancia del resultado hacia la etiqueta lingüística resultante. A continuación, se obtiene un resultado por rol que determina un puntaje para el colectivo de personas que jugaron ese rol, de esta manera, se identifica las áreas de carencia que tienen los sitios web y en que sentido se debe mejorar el software. Finalmente, se encuentra la etiqueta que presenta el resultado final que engloba todos los roles. Este valor de usabilidad se considera para realizar el ranking entre el conjunto de sitios evaluados.

La web mejor evaluada se identifica fácilmente por una estrella al lado del nombre, por el contrario, el sitio peor evaluado se caracteriza por un emoticón de *cara triste*. De la misma forma, internamente cada sitio web realiza un ranking de los roles con los cuales el conjunto de usuarios y expertos evaluó la alternativa, señalando la mejor y peor evaluación para cada rol.

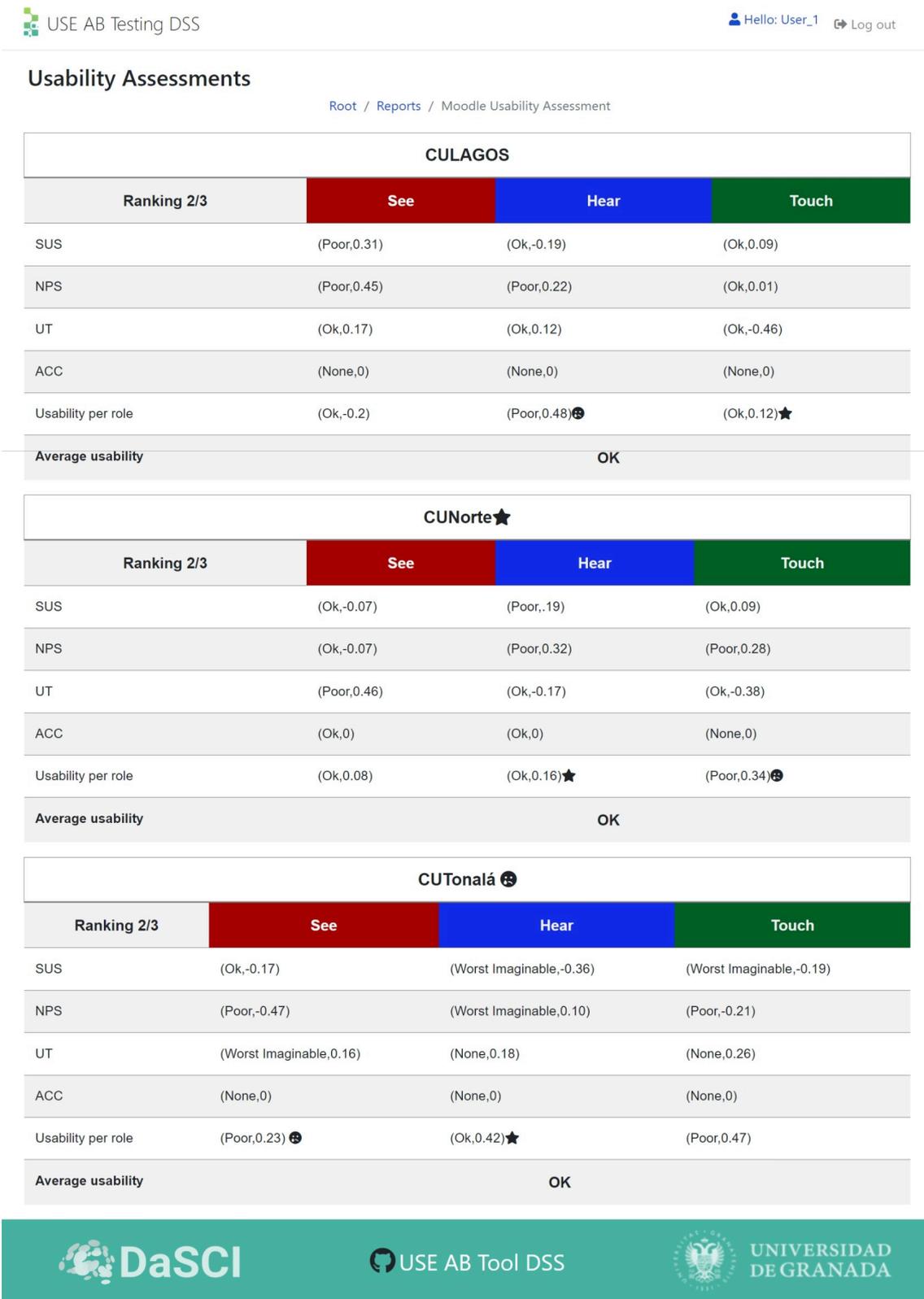


Figura 6.14: Vista del informe de usabilidad en *USE-AB-Tool*. Fuente: Elaboración propia.

En el informe de la prueba de usabilidad podemos obtener datos estadísticos sobre las tareas realizó el conjunto de expertos para conocer el entorno de los sitios web. Este

informe determina por cada tarea la tasa de éxito, tasa de eficiencia y nivel de satisfacción por cada uno de los roles con los cuales se evaluaron los sitios así como de manera general.

6.4. Resumen de capítulo

- Es necesario contemplar elementos de usabilidad en los entornos digitales para mejorar la experiencia del usuario, por ello, se utilizan tecnologías, como software para identificar las áreas de mejora y así crear productos adaptados al usuario final.
- Existen características esenciales para desarrollar un software de *A/B Testing*, basado en la arquitectura de software, identificando patrones de diseño, tipos de información para el sistema, identificación de usuarios, seguridad, escalabilidad, entre otros.
- La adecuada implementación de *USE-AB-Tool* ayuda a mejorar las prácticas de ingeniería de software.
- Las herramientas utilizadas para la implementación del *USE-AB-Tool* fueron Java Script, JS Vue, MySQL y PHP. Se realizó la evaluación de la usabilidad mediante *USE-AB-Tool*, desde los tres enfoques de usuario del sistema (moderador, experto o usuario final).
- El informe de usabilidad con base a las evaluaciones de los usuarios o expertos es importante para determinar las áreas de oportunidad que tiene un software desde el enfoque centrado en el usuario.

Capítulo 7

Caso de estudio: Aplicación web de *USE-AB-Tool*

Este capítulo se conforma con las siguientes secciones. La sección 7.1, describe de manera global la necesidad de una herramienta web para la evaluación de usabilidad. La sección 7.2, detalla los requerimientos utilizadas para la evaluación de *USE-AB-Tool*. La sección 7.4, presenta los diferentes test con los que se solicita información a los usuarios. La sección 7.5, describe la manera en que se opera con la información recolectada. La sección 7.6, describe los pasos con los que obtuvo el resultado final y ranking de usabilidad. Finalmente, la sección 7.7 presenta los puntos importantes de *USE-AB-Tool*.

7.1. Caso de estudio: Descripción de *USE-AB-Tool*

Para facilitar la evaluación de la usabilidad y accesibilidad a través de un modelo lingüístico de Toma de Decisiones (LDM), este trabajo propone una solución software que facilite la aplicación de distintas herramientas de evaluación de manera asíncrona y su integración en un modelo de toma de decisiones. Para que este sistema sea en si mismo usable y accesible, se evalúa con el modelo propuesto en esta investigación LDM4WUE.

Debido a que la herramienta es en línea, gratuita y libre para cualquier persona que necesite una evaluación de la usabilidad, es importante que se considere un buen nivel de usabilidad y accesibilidad, mismo que tendrá un impacto directo en la intuición en el uso de l sistema, por lo tanto, la mínima necesidad de solicitudes de ayuda al equipo de

soporte técnico y desmotivación por parte de los usuarios.

La evaluación de la usabilidad a dos versiones del *USE-AB-Tool* pretende adecuar tres aspectos importantes:

1. **La usabilidad:** Identificar la versión más usable de las propuestas para el *USE-AB-Tool* desde una perspectiva centrada en el usuario.
2. **La accesibilidad:** El acceso a cada sección del software considerando tanto el uso de tecnologías de asistencia, como la posibilidad de acceder a través solo del uso del teclado.
3. **La interpretabilidad:** El informe de usabilidad debe ser altamente interpretable, con ello, identificar las áreas de mejora del sistema que se está evaluando.

7.2. Fase 1. Estructuración del problema

Se presenta el caso de uso para la evaluación de *USE-AB-Tool* desde la metodología LDM4WUE propuesta en el capítulo 5. El objetivo de esta evaluación es elegir la versión más usable y accesible.

Paso 1. Definición del conjunto de alternativas.

Se establecen dos versiones *USE-AB-Tool* como conjunto de alternativas: Para identificar las versiones propuestas para la evaluación se les asigna un nombre que será utilizado como identificador, la versión *Green*¹, y la versión *Light*². La figura 7.1, muestra el acceso principal al sitio web de cada alternativa.

¹<https://lionware.dev/S-USE-AB-TOOL-DSS/>

²<https://lionware.dev/use/>

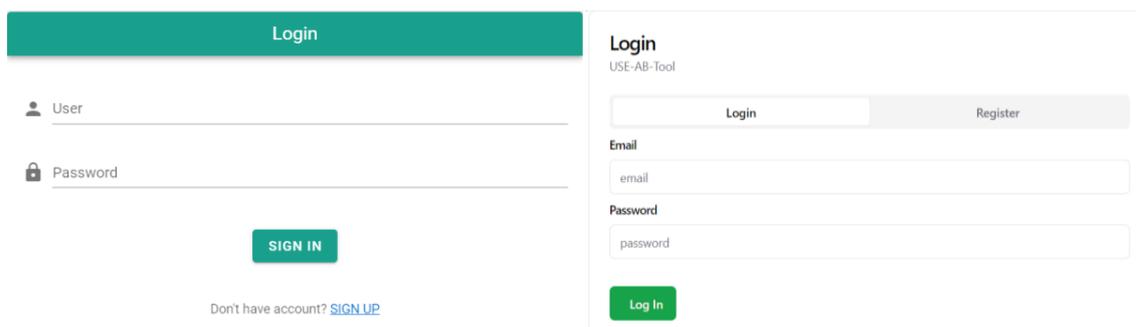


Figura 7.1: Vista de la página de inicio de las dos versiones del sistema *USE-AB-Tool*.

Fuente: Elaboración propia.

Paso 2. Definición del conjunto de criterios y derivación al vector de importancia.

Se establecen los criterios para evaluar la usabilidad del conjunto de alternativas A , sea $C = \{SUS, NPS, UT, ACC, ELUT\}$. En el caso del C_3 , como moderador del proyecto de evaluación se define el conjunto de tareas que conforman la prueba de usabilidad UT y el tiempo estimado máximo para finalizar cada una de ellas. Esta prueba de usabilidad tiene como objetivo que un grupo de usuarios utilicen las diversas secciones del *USE-AB-Tool* a través del recorrido y navegación que realizan en cada una de las tareas. La prueba de usabilidad UT se define por $UT = \{q_1, \dots, q_{10}\}$. A continuación se define el peso de los criterios.

Paso 2.1 Obtener la evaluación de la importancia de los criterios y completar la matriz de comparación de criterios CP .

El moderador, determina el vector de importancia de cada criterio C_j . En este paso se establece la importancia de un criterio sobre otro, utilizando una evaluación lingüística $s \in s_{CP}^5$ (ver sección 2.3.1) que expresa la importancia relativa de cada criterio del test. La tabla 7.1, detalla la evaluación del moderador en cada par de criterios. La tabla 7.2, muestra la transformación que se realizó de etiquetas lingüísticas a su equivalente en números triangulares difusos (TFN). Además, se completa la información en la matriz de comparación CP por medio de la ecuación (2.6).

Paso 2.2 Calcular la extensión sintética difusa.

La extensión sintética difusa es calculada por medio de las ecuaciones 2.7, 2.8 y 2.9. La tabla 7.3 muestra los valores que se derivan en este paso.

Paso 2.3 Cálculo del grado de posibilidad. El grado de posibilidad es calculado a

	C_1	C_2	C_3	C_4	C_5
C_1	Equally important	Moderate important	Equally important	Moderate important ⁻¹	Equally important
C_2		Equally important	Equally important	Equally important	Moderate important ⁻¹
C_3			Equally important	Moderate important ⁻¹	Equally important
C_4				Equally important	Moderate important ⁻¹
C_5					Equally important

Tabla 7.1: Valores lingüísticos asignados por el moderador para obtener la importancia de cada criterio.

	C_1			C_2			C_3			C_4			C_5		
	l	m	u	l	m	u	l	m	u	l	m	u	l	m	u
C_1	1	1	1	1	3	5	1	1	1	$\frac{1}{5}$	$\frac{1}{3}$	1	1	1	1
C_2	$\frac{1}{5}$	$\frac{1}{3}$	1	1	1	1	1	1	1	1	1	1	$\frac{1}{5}$	$\frac{1}{3}$	1
C_3	1	1	1	1	1	1	1	1	1	$\frac{1}{5}$	$\frac{1}{3}$	1	1	1	3
C_4	1	3	5	1	1	1	1	3	5	1	1	1	$\frac{1}{5}$	$\frac{1}{3}$	1
C_5	1	1	1	1	3	5	$\frac{1}{3}$	1	1	1	3	5	1	1	1

Tabla 7.2: La matriz CP se configura con las evaluaciones del moderador para cada criterio.

través de la ecuación 2.10, resultando en los valores dados en la tabla 7.4.

Paso 2.4 Obtención del vector de pesos para el conjunto de criterios C .

El vector de pesos $WC' = \{0.806, 0.477, 0.631, 0.963, 1.000\}$, es obtenido a través de la ecuación 2.11. Finalmente, con la ecuación 2.12, se deriva la normalización del vector de importancia de criterios

$$WC = \{0.208, 0.123, 0.163, 0.248, 0.258\}.$$

Una vez obtenido el vector, es importante determinar la consistencia de las evaluaciones realizadas sobre los criterios valorando el índice de consistencia (CI). En este proyecto de evaluación, $CI = -0.087 < .10$, lo cual demuestra una consistencia correcta según []. En caso contrario, la evaluación debe ser realizada de nueva cuenta hasta que se encuentre un índice de consistencia correcto.

Paso 3. Definición del conjunto de usuarios.

Sea $E = \{E_1, \dots, E_4\}$ el conjunto de expertos y $D = \{D_1, \dots, D_8\}$ el conjunto de usuarios-finales, entonces sea $U = \{U_1, \dots, U_8\}$ la unión de expertos y usuarios-finales, tal que $U = E \cup D$. Cada usuario U_k evaluó la usabilidad de cada alternativa A_i sobre cada criterio C_j .

Eq. 2.9			
Criteria	l	m	u
C_1	4.20	6.33	9.00
C_2	3.40	3.67	5.00
C_3	4.20	4.33	7.00
C_4	4.20	8.33	13.00
C_5	4.33	9.00	13.00
Eq. 2.8			
	l	m	u
	0.021	0.032	0.049
Eq. 2.7			
	l	m	u
C_1	0.09	0.20	0.44
C_2	0.07	0.12	0.25
C_3	0.09	0.14	0.34
C_4	0.09	0.26	0.64
C_5	0.09	0.28	0.64

Tabla 7.3: Cálculo de *Fuzzy sintetic extension*.

$C_1 < C_j$	PD	$C_2 < C_j$	PD	$C_3 < C_j$	PD	$C_4 < C_j$	PD	$C_5 < C_j$	PD
SC1>SC2	1.00	SC2>SC1	0.65	SC3>SC1	0.80	SC4>SC1	1.00	SC5>SC1	1.00
SC1>SC3	1.00	SC2>SC3	0.88	SC3>SC2	1.00	SC4>SC2	1.00	SC5>SC2	1.00
SC1>SC4	0.85	SC2>SC4	0.52	SC3>SC4	0.67	SC4>SC3	1.00	SC4>SC3	1.00
SC1>SC5	0.81	SC2>SC5	0.48	SC3>SC5	0.63	SC4>SC5	0.96	SC5>SC4	1.00

Tabla 7.4: Cálculo del *possibility degree*.

Además, el conjunto de expertos E se compone de personas con conocimiento en alguna disciplina de *User eXperience*, de tal manera que su opinión tiene mas importancia para la evaluación general que el conjunto de usuarios-finales D . Para la evaluación de la herramienta *Use-AB-Tool*, se establece que $WE = 100$ y $WD = 90$, y se considera que cada usuario pertenes a uno de los dos grupos, el siguiente vector se almacena.

$$WU = \{0.263, 0.237, 0.237, 0.263, 0.237, 0.263, 0.263, 0.237\}$$

Enseguida, se determinan los roles y su importancia que son seleccionados por los expertos y usuarios para evaluar el proyecto de usabililidad, se calcula el vector normalizado de pesos de los usuarios W^l a través de la ecuación 5.18.

Paso 4. Definición del conjunto de roles.

El moderador selecciona los siguientes roles $R = \{\text{Discapacidad motora, Discapacidad visual}\}$ como el conjunto de posibles opciones y el máximo número de veces que un experto o usuario puede contestar el *AB-Testing* jugando roles sin repetir. Entonces, el moderador asigna la importancia de los roles estableciendo un vector de importancia de roles $WR' = \{70, 100\}$, donde la importancia normalizada es $WR = \{0.41, 0.59\}$.

Es importante señalar, que los usuarios deben iniciar respondiendo la prueba de usabilidad ($C_3 \cong UT$), y posteriormente, los usuarios son libre de elegir la opción que deseen de los test restantes hasta que sean respondidos la totalidad de test definidos por el moderador. Lo anterior, se debe al tiempo que se debe dedicar en realizar las tareas de la prueba de usabilidad *UT* en cada alternativa, de la manera que este criterio es más de reconocimiento que el resto de los test.

Al conocer el rol que juega cada usuario, es posible calcular el vector de pesos normalizado para cada rol. En la evaluación del *USE-AB-Tool*, se agruparon los usuarios conforme a las discapacidades que ellos tienen, de tal manera que en sistema los usuarios U_1, U_2, U_3, U_4 seleccionaron R_1 debido a alguna discapacidad motora existente, mientras los usuarios U_5, U_6, U_7, U_8 seleccionaron R_2 . Una vez que un usuario U_k elige un rol, este se mantiene durante la evaluación de todos los test C_j para todas las alternativas A_i .

7.3. Fase 2. Empatía y role-playing

Pasos 5 y 6. Sesión informativa y juego de roles.

Las personas que participan en una PU necesitan una breve introducción para saber de previamente las tareas que van a realizar. La metodología propuesta establece tres pruebas estandarizadas, la lista de tareas de la Tabla 7.10 además de un cuestionario alineado a las dimensiones de Norma ISO 25010. El moderador de las pruebas también determina que el grupo de usuarios expertos son los profesores del curso de Diseño de Gráficos e interfaces. Los alumnos que pertenecen al colectivo de alumnos discapacitados inscrito en la Coordinación de Control Escolar de la Universidad de Gaudalajara actúan como usuarios finales. Los usuarios tienen las siguientes discapacidades: discapacidades visuales como daltonismo de 2do grado y personas con problemas motores en alguna extremidad. Las pruebas se proponen a los alumnos como ejercicios prácticos para realizar

pruebas de accesibilidad y usabilidad. Las pruebas A/B completas (evaluaciones para cada alternativa y cada prueba) duran menos de dos horas, incluyendo el tiempo dedicado a dar instrucciones, elegir roles y dos breves descansos. De esta manera, se determina a los expertos y usuarios finales que participaron en la evaluación del LDM4WUE.

7.4. Fase 3. Solicitud de los datos en el caso de uso

Como propósito de esta investigación, se llevó a cabo un A/B testing con personas reales. La prueba de usabilidad $C_3 \cong UT$, se determina específicamente para que el conjunto de usuarios U navegue y conozca las áreas y funciones del *USE-AB-Tool*. El tiempo de realización del *A/B-Testing* completo es de menos de 2 horas, incluyendo el reconocimiento y la realización de los test.

Para aquellos estudiantes que desearon hacerlo de manera presencial, se habilitó un laboratorio de cómputo en un horario de 10:00 am a 12:00 pm donde tuvieron el equipamiento necesario para realizar la evaluación de la usabilidad. Por el contrario, para aquellos que tenían alguna actividad en ese horario establecido, se realizó la actividad de manera remota cumpliendo la condición que el cierre de la evaluación sería a las 20:00 horas.

El conjunto de respuestas puede ser descargado desde el repositorio del proyecto en GitHub³. Específicamente para la prueba de usabilidad UT , se obtiene información relevante acerca de las tareas q_v que no se encuentra en la matriz de evaluación lingüística tal como la eficiencia y eficacia de cada una de ellas, tal que al analizar estos datos se determine el desarrollo de una mejor prueba de usabilidad o la posibilidad de identificar el flujo de un proceso en el sistema que podría ser mejorado.

Paso 7. Recolección la evaluación de los usuarios.

La solución software *USE-AB-Tool* recolecta la información de cada usuario U_k que evalúa las versiones *Green* y *Light* del software. Las respuestas recolectadas para el cuestionario SUS (C_1) se presentan en la tabla 7.5. La ponderación de NPS (C_2) se detalla en la tabla 7.6. La totalidad de las respuestas derivadas de la prueba de usabilidad (C_3) se presentan en el anexo, sin embargo, la tabla 7.7 muestra las respuestas para C_3 obtenías

³<https://github.com/ari-dasci/S-USE-AB-Tool>

del usuario U_k jugando el rol R_l . La tabla 7.9 muestra los datos obtenidos del cuestionario propio validado $ELUT$. Finalmente, y solo para el grupo de expertos D , el puntaje para el nivel de accesibilidad interpretado lingüísticamente en términos de números de errores y advertencia con apoyo de la herramienta WAVE y una inspección manual se presenta en la tabla 7.8.

Alternatives A_i	U_1^1	U_2^1	U_3^1	U_4^1	U_5^2	U_5^6	U_{72}	U_8^2
A_1	92.5	67.5	62.5	55	52.5	40	75	82.5
A_2	95	67.5	70	90	67.5	70	97.5	85

Tabla 7.5: Vista de las respuestas de $C_1 \cong SUS$ para cada A_i .

	U_1^1	U_2^1	U_3^1	U_4^1	U_5^2	U_6^2	U_7^2	U_8^2
A_1	9	7	6	6	8	6	8	7
A_2	9	6	9	9	9	7	9	8

Tabla 7.6: Vista de las respuestas de $C_2 \cong NPS$ con NPS_LTR_{score} para cada A_i .

UT	A_1					A_2			
	$MaxTime$	$Time_1^{4,1}$	$Effic_1^{4,1}$	$Success_1^{4,1}$	$Satis_1^{4,1}$	$Time_2^{4,1}$	$Effic_2^{4,1}$	$Success_2^{4,1}$	$Satis_2^{4,1}$
q_1	60	103	0	1	$(s_3^5, 0)$	16	1	1	$(s_3^5, 0)$
q_2	20	22	0	1	$(s_3^5, 0)$	14	1	1	$(s_3^5, 0)$
q_3	35	27	0	1	$(s_3^5, 0)$	2	1	1	$(s_3^5, 0)$
q_4	15	9	1	1	$(s_2^5, 0)$	5	1	1	$(s_3^5, 0)$
q_5	240	290	0	1	$(s_2^5, 0)$	146	1	1	$(s_2^5, 0)$
q_6	250	90	1	1	$(s_1^5, 0)$	162	1	1	$(s_2^5, 0)$
q_7	100	43	1	1	$(s_2^5, 0)$	38	1	1	$(s_3^5, 0)$
q_8	90	83	0	1	$(s_0^5, 0)$	80	1	1	$(s_3^5, 0)$
q_9	60	72	0	1	$(s_1^5, 0)$	30	1	1	$(s_3^5, 0)$
q_{10}	120	132	0	1	$(s_2^5, 0)$	31	1	1	$(s_3^5, 0)$
q_{11}	30	23	0	1	$(s_2^5, 0)$	21	1	1	$(s_4^5, 0)$
q_{12}	120	35	1	1	$(s_3^5, 0)$	30	1	1	$(s_3^5, 0)$
q_{13}	35	24	1	1	$(s_2^5, 0)$	18	1	1	$(s_4^5, 0)$
Promedio			38.46 %	100.00 %	$(s_3^5, 0)$		100.00 %	100.00 %	$(s_3^5, 0)$

Tabla 7.7: Un usuario U_4 del conjunto E evalúa C_3 jugando el rol R_1 .

Paso 8. Construcción de las matrices de decisión individual

En este paso se construyen las matrices de decisión individual con las respuestas del conjunto de usuarios U . Con la finalidad de realizar una agregación lingüística colectiva, los datos deben pasar por un proceso de homogeneización. Para que el proceso sea claro y entendible, a continuación se explica la transformación de los datos del usuario U_1 con el rol R^l evaluando la alternativa A_1 sobre el conjunto de criterios C .

	Expertos					Usuarios finales				
U_k^l	U_1^1	U_2^1	U_3^1	U_4^2	U_5^2	U_6^2	U_7^2	U_7^2	U_8^2	
A_1	AA	-	-	A	-	A	AA	-		
A_2	AA	-	-	AA	-	AA	AA	-		

Tabla 7.8: Vista de las respuestas de $C_4 \cong ACC$ con etiquetas lingüísticas correspondientes a Accesibilidad.

U_k^l	A_1	A_2
U_1^1	$(s_2^5, 0)$	$(s_3^5, -0.42)$
U_2^1	$(s_2^5, 0.33)$	$(s_2^5, 0.25)$
U_3^1	$(s_2^5, 0.17)$	$(s_3^5, -0.42)$
U_4^1	$(s_2^5, 0.33)$	$(s_2^5, 0)$
U_5^2	$(s_2^5, -0.17)$	$(s_2^5, 0.42)$
U_6^2	$(s_2^5, 0.17)$	$(s_3^5, 0.08)$
U_7^2	$(s_2^5, 0.17)$	$(s_3^5, -0.25)$
U_8^2	$(s_2^5, 0.08)$	$(s_2^5, 0.42)$

Tabla 7.9: Vista de las respuestas de $C_5 \cong ELUT$ con *ELUT_score* para cada A_i .

- **Respuestas de $C_1 \cong SUS$.** Las respuestas para las 10 preguntas del cuestionario SUS son $\{5, 1, 5, 1, 4, 2, 4, 1, 4, 1\}$. Además, usando la ecuación 5.1 se obtiene un $SUS_score_2^{4,1} = 90$. Este valor numérico es transformado a través de la ecuación 5.9, obteniendo $ID_{2,1}^{4,1} = TF_{SUS}(90) = (s_5^{sus}, 0.2) \in S^{SUS}$.
- **Respuestas de $C_2 \cong NPS$.** La respuesta directa a la pregunta LTS (*¿Qué tanto recomendarías esta web a un amigo?*) es 9. Este valor numérico es transformado en una 2-Tupla lingüística a través de la ecuación 5.10, entonces $NPS_SUS_score = 95.875$, Finalmente al aplicar la ecuación 5.11 se obtiene $ID_{2,2}^{4,1} = TF_{SUS}(95.875) = (s_6^{sus}, -0.33) \in S^{SUS}$.
- **Respuestas de $C_3 \cong PU$.** Se aplica una *PU* diseñada para utilizar todas las secciones de la solución software. Esta prueba consiste en 13 tareas (enlistadas en la tabla 7.10). El usuario U_4^1 da los siguientes resultados:

 1. *Efficiency rate*: Se finalizan 13 actividades de 13 propuestas en la *PU* por debajo del *MaxTime* definido por el moderador, del tal manera que $Efficiency_score_2^{4,1} = 100.00$ (by eq. 5.2).
 2. *Success rate*: De las 13 tareas se logra realizar las 13, obteniendo $Success_score_2^{4,1} = 100$ (por medio de la eq. 5.3).

q_v	Tareas de la Prueba de usabilidad PU
q_1	Registrar usuario.
q_2	Iniciar sesión.
q_3	Abrir el proyecto de evaluación A/B testing-DSS.
q_4	Seleccionar un rol de los que se encuentran habilitados por el moderador.
q_5	Contestar la prueba de usabilidad para cada alternativa insertando el tiempo en segundos que se tardó en realizar cada actividad en cada una de las alternativas, además, si se logró o no la culminación de la actividad y el nivel de satisfacción.
q_6	Contestar las demás pruebas. La evaluación de la accesibilidad es solo para expertos.
q_7	Crear proyecto de evaluación con tu nombre.
q_8	Agregar dos páginas diferentes como alternativas con un mismo objetivo.
q_9	Seleccionar tres tipos de test para evaluar los sitios web.
q_{10}	Agregar tres usuarios finales y tres usuarios expertos.
q_{11}	Asignar importancia de 100 a expertos y de 90 a usuarios finales.
q_{12}	Agregar dos tipos de roles (discapacidades) y asignarle el peso a cada una de ellas.
q_{13}	Ver informe de usabilidad.

Tabla 7.10: Prueba de usabilidad C_3 para evaluar $USE-AB-Tool$.

3. *Satisfaction level*: el nivel de satisfacción depende de cada tarea de la PU , sin embargo, el promedio obtenido es (aplicando las eq. 5.5 y 5.12) para el usuario U_4 es $Satisfaction_score_1^{4,1} = (s_2^5, -0.036)$. Además, $ID_{2,3}^{4,1} = (s_3^5, 0) \in S^5$.

- **Respuestas de $C_4 \cong ACC$** . El usuario U_4 que participó en la evaluación como experto, utilizó la herramienta WAVE en el buscador sobre la alternativa A_2 . La interpretación del resumen provisto por la herramienta resulta en una etiqueta de AA , Entonces, se obtiene $ID_{2,4}^{4,1} = (AA, 0) = (s_1, 0) \in S^3$.
- **Respuestas de $C_5 \cong ACC$** . El usuario U_4 , contestó el cuestionario ELUT. resolviendo mediante el operador de agregación de media aritmética, se obtiene $ID_{2,5}^{4,1} = (s_2, 0.5) \in S^5$.

7.5. Fase 4. Agregación colectiva del caso de uso

Como se puede observar, la información proviene de distintos test, por lo que debido a la particularidad de cada uno de ellos la información resultante es heterogénea. Para agregar la información obtenida previamente del conjunto de usuarios U , es necesario realizar un proceso de unificación previo a la agregación de la información. A continuación se detallan los pasos necesarios para calcular la UDC^{global} a partir de las matrices $ID^{k,l}$.

Paso 9. Unificación de la información hacia S^9 .

Los valores de la tabla 7.11 son unificados hacia S^9 aplicando las ecuaciones 5.15 y 5.16. Los resultados obtenidos de los procedimientos para cada alternativa se presentan en la tabla 7.12.

	$A_1 = \{USE - AB - Tool - v1\}$					$A_2 = \{USE - AB - Tool - v2\}$				
U_r^l	$C_1 \cong SUS$	$C_2 \cong NPS$	$C_3 \cong UT$	$C_4 \cong ACC$	$C_5 \cong ELUT$	$C_1 \cong SUS$	$C_2 \cong NPS$	$C_3 \cong UT$	$C_4 \cong ACC$	$C_5 \cong ELUT$
U_1^1	$(s_5^{SUS}, 0.4)$	$(s_5^{SUS}, -0.33)$	$(s_3^5, 0)$	$(s_1^3, 0)$	$(s_2^5, 0)$	$(s_6^{SUS}, -0.4)$	$(s_6^{SUS}, -0.33)$	$(s_3^5, 0.38)$	$(s_1^3, 0)$	$(s_3^5, -0.42)$
U_2^1	$(s_4^{SUS}, -0.3)$	$(s_4^{SUS}, -0.16)$	$(s_2^5, 0.17)$		$(s_2^5, 0.33)$	$(s_4^{SUS}, -0.3)$	$(s_3^{SUS}, 0.33)$	$(s_3^5, 0.33)$		$(s_2^5, 0.25)$
U_3^1	$(s_4^{SUS}, -0.5)$	$(s_3^5, 0.33)$	$(s_2^5, -0.17)$		$(s_2^5, 0.17)$	$(s_4^{SUS}, -0.2)$	$(s_6^{SUS}, -0.33)$	$(s_3^5, 0.42)$		$(s_3^5, -0.42)$
U_4^1	$(s_2^5, 0.2)$	$(s_2^5, 0.33)$	$(s_2^5, 0)$	$(s_0^3, 0)$	$(s_2^5, 0.33)$	$(s_5^{SUS}, 0.2)$	$(s_6^{SUS}, -0.33)$	$(s_3^5, 0)$	$(s_3^3, 0)$	$(s_2^5, 0)$
U_5^1	$(s_2^5, 0.1)$	$(s_5^{SUS}, -0.33)$	$(s_2^5, 0.33)$		$(s_2^5, -0.17)$	$(s_2^5, -0.3)$	$(s_6^{SUS}, -0.33)$	$(s_2^5, 0.33)$		$(s_2^5, 0.42)$
U_6^1	$(s_2^5, 0.2)$	$(s_2^5, 0.33)$	$(s_3^5, -0.15)$	$(s_0^3, 0)$	$(s_2^5, 0.17)$	$(s_4^{SUS}, -0.2)$	$(s_5^{SUS}, -0.33)$	$(s_3^5, 0.08)$	$(s_3^3, 0)$	$(s_2^5, 0.08)$
U_7^1	$(s_4^{SUS}, 0)$	$(s_5^{SUS}, -0.33)$	$(s_2^5, 0.15)$	$(s_1^3, 0)$	$(s_2^5, 0.17)$	$(s_6^{SUS}, -0.2)$	$(s_6^{SUS}, -0.33)$	$(s_3^5, 0.38)$	$(s_1^3, 0)$	$(s_3^5, -0.25)$
U_8^1	$(s_5^{SUS}, -0.4)$	$(s_4^{SUS}, -0.16)$	$(s_2^5, -0.42)$		$(s_2^5, 0.08)$	$(s_5^{SUS}, -0.2)$	$(s_4^{SUS}, -0.16)$	$(s_3^5, 0.17)$		$(s_2^5, 0.42)$

Tabla 7.11: Elementos de las matrices $ID^{k,l}$, representadas como 2-Tuplas.

	$A_1 = \{USE - AB - Tool - v1\}$					$A_2 = \{USE - AB - Tool - v2\}$				
U_r^l	$C_1 \cong SUS$	$C_2 \cong NPS$	$C_3 \cong UT$	$C_4 \cong ACC$	$C_5 \cong ELUT$	$C_1 \cong SUS$	$C_2 \cong NPS$	$C_3 \cong UT$	$C_4 \cong ACC$	$C_5 \cong ELUT$
U_1^1	$(s_7^9, 0.40)$	$(s_8^9, -0.33)$	$(s_6^9, 0)$	$(s_4^9, 0)$	$(s_4^9, 0)$	$(s_8^9, -0.40)$	$(s_8^9, -0.33)$	$(s_7^9, -0.23)$	$(s_4^9, 0)$	$(s_6^9, 0.26)$
U_2^1	$(s_5^9, 0.4)$	$(s_6^9, -0.32)$	$(s_4^9, 0.33)$		$(s_5^9, -0.34)$	$(s_5^9, 0.40)$	$(s_5^9, -0.34)$	$(s_7^9, -0.33)$		$(s_6^9, 0.24)$
U_3^1	$(s_5^9, 0)$	$(s_5^9, -0.34)$	$(s_4^9, -0.33)$		$(s_4^9, 0.34)$	$(s_6^9, -0.4)$	$(s_8^9, -0.33)$	$(s_7^9, -0.17)$		$(s_6^9, 0.24)$
U_4^1	$(s_4^9, 0.40)$	$(s_5^9, -0.34)$	$(s_4^9, 0)$	$(s_0^9, 0)$	$(s_5^9, -0.34)$	$(s_4^9, 0.20)$	$(s_8^9, -0.33)$	$(s_6^9, 0)$	$(s_4^9, 0)$	$(s_6^9, 0.26)$
U_5^1	$(s_4^9, 0.20)$	$(s_7^9, -0.33)$	$(s_5^9, -0.33)$		$(s_4^9, -0.33)$	$(s_5^9, 0.40)$	$(s_8^9, -0.33)$	$(s_5^9, -0.33)$		$(s_6^9, 0.24)$
U_6^1	$(s_3^9, 0.20)$	$(s_5^9, -0.34)$	$(s_6^9, -0.31)$	$(s_0^9, 0)$	$(s_4^9, 0.33)$	$(s_6^9, -0.4)$	$(s_7^9, -0.33)$	$(s_6^9, 0.15)$	$(s_4^9, 0)$	$(s_6^9, 0.26)$
U_7^1	$(s_6^9, 0)$	$(s_7^9, -0.33)$	$(s_0^9, 0)$	$(s_4^9, 0)$	$(s_4^9, 0.33)$	$(s_8^9, -0.20)$	$(s_8^9, -0.33)$	$(s_7^9, -0.23)$	$(s_4^9, 0)$	$(s_6^9, 0.26)$
U_8^1	$(s_7^9, -0.4)$	$(s_6^9, -0.32)$	$(s_3^9, 0.17)$		$(s_4^9, 0.17)$	$(s_7^9, -0.20)$	$(s_6^9, -0.32)$	$(s_6^9, 0.33)$		$(s_6^9, 0.24)$

Tabla 7.12: Matrices de Decisión Unificadas (UID) expresadas en S^9 .

Paso 10. Agregación por cada rol.

El vector unificado de decisiones colectivas se calcula para cada función. En primer lugar, las valoraciones de S^9 se agrupan en una matriz UCD^l para cada función R_l . Sea el caso de $l = 1$ y las valoraciones de los usuarios $U_k, k = \{1, 2, 3, 4\}$ representadas en la matriz UCD^1 . A continuación, se aplicamos la ecuación 5.17 utilizando el vector de pesos de usuario W^1 para obtener los elementos UCD_{ij}^l . Posteriormente, mediante la ecuación 5.19 y el vector de ponderaciones de criterios (véase el Paso 2.4. en la Sección 7.2) se calcula el vector de decisión colectiva unificada $u\mathbf{cd}^l$ por rol. Los resultados de ambos procedimientos se muestran en la tabla 7.13. Esto debe repetirse posteriormente para cubrir $UCD^l (l = 2, \dots, r)$.

Paso 11. Agregación global.

Se debe examinar la matriz UCD_{ij}^l con el fin de integrar toda la información. Posteriormente, se agrega la información teniendo en cuenta las ponderaciones asignadas a los roles, tal como se indica en la Sección 7.2, denotada como WR . Se aplica la ecuación 5.20 para calcular cada elemento de decisión colectiva global unificada UCD_{ij}^{global} , que se utiliza

A_i	$C_1 \cong SUS$ $WC^1 = 0.208$	$C_2 \cong NPS$ $WC^2 = 0.123$	$C_3 \cong UT$ $WC^3 = 0.163$	$C_4 \cong ACC$ $WC^4 = 0.248$	$C_5 \cong ELUT$ $WC^5 = 0.258$	ucd_i^l
A_1	$(s_6^9, -0.43)$	$(s_6^9, -0.31)$	$(s_5^9, -0.47)$	$(s_2^9, 0)$	$(s_4^9, 0.41)$	$(s_4^9, 0.44)$
A_2	$(s_7^9, -0.5)$	$(s_7^9, -0.04)$	$(s_7^9, -0.44)$	$(s_5^9, -0.42)$	$(s_5^9, -0.3)$	$(s_6^9, -0.14)$
A_1	$(s_5^9, -0.02)$	$(s_6^9, -0.09)$	$(s_3^9, 0.35)$	$(s_2^9, 0)$	$(s_4^9, 0.14)$	$(s_4^9, 0.07)$
A_2	$(s_6^9, 0.42)$	$(s_7^9, -0.06)$	$(s_6^9, 0.01)$	$(s_4^9, 0)$	$(s_5^9, 0.36)$	$(s_6^9, -0.26)$

Tabla 7.13: *Role playing* derivado en ucd_i^1 , el vector 2-Tuplas con la evaluación de la usabilidad para cada alternativa.

para informar de una puntuación lingüística para cada prueba de usabilidad y para cada alternativa.

Además, se puede asignar una puntuación lingüística a cada alternativa mediante la ecuación 5.21. Se anota como decisión colectiva unificada el vector ucd_i^{global} y se representa la *usabilidad* de los sitios web. La tabla 7.14 muestra 2-Tuplas en S^9 que son la representación colectiva de las evaluaciones de usabilidad dadas por todos los usuarios que juegan con los sitios web alternativos a través de un conjunto de pruebas.

A_i	$ucd_{i,1}^{global}$	$ucd_{i,2}^{global}$	$ucd_{i,3}^{global}$	$ucd_{i,4}^{global}$	$ucd_{i,5}^{global}$	ucd_i^{global}
A_1	$(s_5^9, 0.21)$	$(s_6^9, -0.18)$	$(s_4^9, -0.18)$	$(s_4^9, -0.45)$	$(s_5^9, -0.41)$	$(s_4^9, 0.49)$
A_2	$(s_6^9, 0)$	$(s_7^9, 0)$	$(s_6^9, 0)$	$(s_4^9, 0)$	$(s_5^9, 0)$	$(s_6^9, -0.43)$

Tabla 7.14: Para cada alternativa, el modelo LDM4WUE puede calcular e informar puntajes combinados de usabilidad basados en la importancia de los roles en cada test.

7.6. Fase 5. Explotación de la información del caso de uso

Para clasificar las alternativas evaluadas se utiliza la técnica TOPSIS (ver sección 2.3.2). A continuación se detalla el procedimiento para los dos tipos de clasificación obtenidos en el modelo propuesto:

Paso 12: Generación de rankings por rol.

Para los pasos siguientes, los cálculos se basan en la matriz UCD^l para cada función R_l , que aparece en la tabla 7.15. Obsérvese que UCD^1 se repite en tabla 7.13. A continuación se describe el procedimiento paso a paso, la solución ideal positiva A^{+l} y la solución ideal

negativa A^{-l} para R_1 R_2 y se determinan mediante las ecuaciones 2.13 - 2.16 y se muestran a continuación.

$$A^{+l} = \{1.352, 0.856, 1.124, 1.138, 1.213\}, A^{-l} = \{1.158, 0.701, 1.067, 0.497, 1.137\}, R_1$$

$$A^{+l} = \{1.334, 0.853, 0.977, 0.993, 1.382\}, A^{-l} = \{1.035, 0.727, 0.833, 0.497, 1.067\}, R_2$$

A continuación, las medidas de separación, D_i^{+l} y D_i^{-l} , de cada alternativa A_i respecto a las soluciones ideales positivas A^{+l} y las soluciones ideales negativas A^{-l} se calculan mediante las ecuaciones 2.17-2.18, como se expresa en la tabla 7.16. Asimismo, los coeficientes de proximidad relativa $RC_i^l (i = 1, 2, 3)$ se calculan mediante la ecuación 2.19. Estos se utilizan para clasificar las alternativas como $Ranking^l$ y los resultados se muestran en la tabla 7.15.

R^l	A_i	$C_1 \cong SUS$ $WC^1 = 0.208$	$C_2 \cong NPS$ $WC^2 = 0.123$	$C_3 \cong UT$ $WC^3 = 0.163$	$C_4 \cong ACC$ $WC^4 = 0.248$	$C_5 \cong ELUT$ $WC^5 = 0.258$	ucd_i^l	$Ranking^l$
$R^1 \cong Touch$	A_1	$(s_6^9, -0.43)$	$(s_6^9, -0.31)$	$(s_5^9, -0.47)$	$(s_2^9, 0)$	$(s_4^9, 0.41)$	$(s_4^9, 0.44)$	2
	A_2	$(s_7^9, -0.5)$	$(s_7^9, -0.04)$	$(s_7^9, -0.44)$	$(s_5^9, -0.42)$	$(s_5^9, -0.3)$	$(s_6^9, -0.14)$	1
$R^2 \cong See$	A_1	$(s_5^9, -0.02)$	$(s_6^9, -0.09)$	$(s_3^9, 0.35)$	$(s_2^9, 0)$	$(s_4^9, 0.14)$	$(s_4^9, 0.07)$	2
	A_2	$(s_6^9, 0.42)$	$(s_7^9, -0.06)$	$(s_6^9, 0.01)$	$(s_4^9, 0)$	$(s_5^9, 0.36)$	$(s_6^9, -0.26)$	1

Tabla 7.15: Ranking de alternativas por roles.

Según $Ranking^1 = A_2 \succ A_1$ que se centra en el rol $R_1 = \{Touch\}$, la mejor usabilidad se muestra en el sitio web A_2 y la alternativa A_1 tiene el menor grado de usabilidad para los usuarios con este rol. Con respecto al rol $R_2 = \{See\}$, se obtiene un resultado similar, $Ranking^2 = A_2 \succ A_1$, confirmando que A_2 tiene mejor nivel de usabilidad sobre la otra alternativas.

Paso 13: Generación del ranking global .

Para establecer una clasificación global, se consideran como base las matrices UCD^l (ver tabla 7.15). A continuación, se calcula la solución ideal positiva A^+ y la solución ideal negativa A^- mediante las ecuaciones 2.13 - 2.16 y se presentan a continuación.

$$A^{+l} = \{6.449, 6.944, 6.227, 4.233, 5.096\}, A^{-l} = \{5.215, 5.821, 3.822, 3.550, 4.585\}.$$

Posteriormente, las medidas de separación, D_i^+ y D_i^- , de cada alternativa A_i respecto a la solución ideal positiva A^+ y la solución ideal negativa A^- se calculan mediante las ecuaciones 2.17-2.18. Del mismo modo, los coeficientes relativos de proximidad $RC_i (i = 1, 2)$ se calculan mediante la ecuación 2.19. Todos estos resultados se presentan en la tabla 7.16.

A_i	D_i^+	D_i^-	RC_i	$Ranking^{global}$
A_1	3.049	0.000	0.000	2
A_2	0.000	3.049	1.000	1

Tabla 7.16: Valores utilizados para ordenar las alternativas en $Ranking^{global}$

De acuerdo al $Ranking^{global} = A_2 \succ A_1$, de las dos versiones de *USE-AB-Tool*, A_2 muestra que es más adecuado para usuarios reales teniendo en cuenta discapacidades reales.

Paso 14. Retranslación

Se calcula el vector *informe adjetivo de usabilidad* (\mathbf{aur}^l) para cada rol R_l y el vector *informe adjetivo de usabilidad* (\mathbf{aur}^{global}) global mediante las ecuaciones 5.22 y 5.23, respectivamente. Los resultados se muestran en la tabla 7.17. Según las clasificaciones y los informes de usabilidad del adjetivo, es evidente que A_2 es la mejor opción teniendo en cuenta los aspectos de usabilidad de la interfaz. Lingüísticamente la alternativa mejor evaluada tiene una etiqueta adjetiva *Good* mientras que la otra alternativa tiene una etiqueta adjetiva *OK*, gracias al modelo propuesto podemos conseguir una mejor comprensión de las experiencias de los usuarios en estos tres sitios.

A_i	R_l	ucd_i^l	aur_i^l	ucd_i	aur_i	<i>Usability</i>
A_1	$R^1 \cong Touch$	$(S_4^9, 0.44)$	$(s_3^{sus}, 0.44)$	$(s_4^9, 0.49)$	$(s_3^{sus}, 0.49)$	<i>OK</i>
	$R^1 \cong See$	$(S_4^9, 0.07)$	$(s_3^{sus}, 0.07)$			
A_2	$R^1 \cong Touch$	$(S_6^9, -0.14)$	$(s_4^{sus}, -0.14)$	$(s_6^9, -0.43)$	$(s_4^{sus}, -0.43)$	<i>Good</i>
	$R^1 \cong See$	$(S_6^9, -0.26)$	$(s_4^{sus}, -0.26)$			

Tabla 7.17: Puntuaciones finales del informe dado bajo escala adjetiva S^{SUS} .

7.7. Resumen de capítulo

- Se eligieron dos grupos de usuarios, el primer grupo fue aquel cuyos integrantes tienen algún tipo de discapacidad motriz y visual, y el segundo grupo fue aquel que el conjunto de expertos y cuyos integrantes tienen conocimiento en algún área del HCI, de tal manera que la evaluación se realizó a través de usuarios con necesidades reales y expertos.
- Se evaluó la usabilidad de dos versiones del *USE-AB-Tool* con la finalidad de encontrar cual de esas versiones es la más usable para usuarios con algún tipo de discapacidad.
- La flexibilidad del *USE-AB-Tool* permitió realizar la evaluación remotamente o en sitio. Esta evaluación requería contestar un conjunto de tareas en el sistema de la manera mas natural posible en ellos, utilizando alguna tecnología de asistencia en caso de ser necesario o simplemente con con la computadora.
- Ambas versiones se evaluaron a través de cinco pruebas, tres de ellas estandarizadas, una prueba de usabilidad para el reconocimiento del sistema y un cuestionario nuevo basado en la Norma ISO 25010.
- Una vez aplicado el modelo LDM4WUE en las dos versiones de *USE-AB-Tool*, los resultados obtenidos muestran que la versión 2 tiene mejor grado de usabilidad, basándose en la perspectiva de usuarios, así como mejor nivel de accesibilidad siendo más compatibles con tecnologías de asistencia para personas con discapacidad visual.
- Se destaca la facilidad de la evaluación de los sistemas por la flexibilidad de *USE-AB-Tool* en pruebas remotas y en sitio así como síncronas y asíncronas.

Capítulo 8

Conclusiones y trabajo Futuro

Este proyecto de investigación se basa en utilizar las opiniones de los usuarios para medir la usabilidad y la accesibilidad web, utilizando instrumentos, tanto de elaboración propia, como cuestionarios estandarizados. Es así que este trabajo tiene como enfoque la medición de la usabilidad y la accesibilidad como un problema de Toma de Decisiones, que debe estar centrado en el usuario.

De los puntos principales a resaltar, es con respecto a las métricas para medir la usabilidad y accesibilidad, ya que las formas clásicas de evaluación son costosas y de larga duración. Aunado a ello, las métricas clásicas no reflejan la diversidad de los usuarios y de los espacios o contextos reales. Esto puede ser un obstáculo para las evaluaciones, por lo tanto, en este proyecto de investigación se propuso una opción gratuita y centrada en el usuario para medir la usabilidad.

Por ello, como una de las principales aportaciones de este trabajo, se realizó la propuesta del Test de Usabilidad Lingüístico Emocional (ELUT, por sus siglas en inglés), de elaboración propia, acorde a las seis dimensiones de la Norma ISO 25010 para medir la usabilidad de los sitios web. Para alcanzar la confiabilidad y la validez de dicho instrumento, se realizaron varios pasos. Primeramente, se realizó la evaluación por expertos para alcanzar la validez de contenido del instrumento, se trabajó con un panel de expertos muy diverso, ya que no todos los jueces tenían el mismo nivel de pericia en todas las áreas. Se utilizaron las opiniones de 13 expertos para la validez de contenido, en donde se desarrollaron conjuntos de términos lingüísticos multigranulares diferentes para que cada experto eligiera el que considerara pertinente respecto a su grado de experiencia; de

esa manera aquellos que tenían más experiencia en el área eligieron aquellos con mayor granularidad para ser más precisos. Como aporte principal, se utilizó el método *2-Tuple Fuzzy Linguistic Delphi* (2TFLD) para validar el consenso de los expertos con respecto a los ítem del test. La flexibilidad del método propuesto como DM, basado en el consenso mediante el uso de conjuntos de términos lingüísticos multigranulares, permite dar y recibir información comprensible. Se concluyó que es importante realizar la validación de contenido del mismo, ya que asegura que cada ítem contribuya significativa y coherentemente al constructo medido.

Además, para evaluar la fiabilidad de ELUT, se utilizó el Coeficiente Alfa de Cronbach, como medida crucial de la consistencia interna del instrumento, estableciendo un umbral deseable. Se tomaron los resultados de la prueba piloto con estudiantes de Ingeniería en Ciencias Computacionales, del Centro Universitario de Tonalá, en la interacción con un sitio web, obteniendo datos favorables, dado que se obtuvo un Alfa de Cronbach de .833, logrando el objetivo de elaborar y validar dicho test.

Finalmente, como bondades de ELUT, se resalta su confiabilidad, validez y facilidad en su aplicación, ya que se basa en preguntas sencillas y concretas, enfocadas en las dimensiones de la Norma ISO 25010, siendo de rápida aplicación y basada en términos lingüísticos, lo que facilita que los usuarios lo respondan, además de su posterior interpretación más cercana al razonamiento humano. Por lo tanto, se utilizó el modelo de Toma de Decisiones Lingüístico (LDM, por sus siglas en inglés), ya que utiliza variables lingüísticas aptas para abordar la incertidumbre e imprecisión del lenguaje natural humano.

Como aporte, se construyó una herramienta como apoyo para el proceso de LDM, como soporte en la aplicación del método 2TFLD de forma cómoda. Las funcionalidades de las herramientas del DSS, como las opciones de recorte, son muy útiles para comprender los niveles umbrales de discrepancia, por lo que es una herramienta muy útil para el moderador que interpreta favorablemente los niveles colectivos de confianza y consenso. Así, se presentó como una evaluación lingüística altamente interpretable para cada criterio (claridad, redacción, pertinencia, escala evaluativa), que tiene cada ítem del cuestionario.

Con respecto a la valoración de la usabilidad de la solución software, dicha propuesta incorporó técnicas lingüísticas que son tendencia en la evaluación de la usabilidad de los sistemas informáticos. Se le dio importancia a los roles de usuario, enfocado en la

caracterización de las condiciones o necesidades particulares del usuario final, ya que son aspectos de importancia crítica en el campo de la *User eXperience* (UX) y la evaluación de la usabilidad.

Por otro lado, el presente proyecto de investigación contribuyó a la identificación de áreas de oportunidad para mejorar la satisfacción del usuario final, así como para reducir la frustración de los usuarios en el uso de los sistemas informáticos. Retomando la perspectiva de testear la usabilidad bajo la opinión de satisfacción del usuario como colectivo y el uso de tests estándar como SUS y NPS se realizó una propuesta para realizar *A/B testing* que no estuviera limitada por un número de alternativas, ni limitada al número de test. Siguiendo el paradigma del *Design Thinking*, se realizó una Prueba de Usabilidad habilitando el uso de *personas* y *juegos de rol*, ponderando las opiniones de los usuarios, para equilibrar el punto de vista del desarrollo con respecto al resto de *stakeholders* y usuarios finales.

El modelo propuesto de evaluación de la usabilidad con usuarios finales, es fácil de desarrollar en línea, y puede ayudar a definir nuevas pruebas y funciones. Para ejecutar un *A/B Testing* como la que se describe en la Sección 5.1. El *USE-AB-Tool*¹ se diseñó para poder cubrir las cuatro fases necesarias para resolver el problema LDM de evaluación de la usabilidad: 1) definir el *A/B Testing*, 2) recopilar información del usuario, 3) unificación y agregación colectiva, y 4) explotación a través de rankings y la generación del informe de *feedback* de usabilidad. Ese informe describió las puntuaciones de usabilidad por rol y por prueba, y también globalmente para todo el sitio web alternativo utilizando la escala adjetiva SUS, que es un estándar en el campo de la UX. Se destaca la aplicación del modelo a un caso de uso real relativo a la evaluación de la herramienta.

La utilización de variables lingüísticas simplifica la calificación de la percepción del usuario. Además, el modelo lingüístico computacional de 2-Tuplas, junto con técnicas como Fuzzy AHP y Fuzzy TOPSIS, permite procesar la información lingüística sin pérdidas. Mediante la incorporación de opiniones colectivas de los responsables de la toma de decisiones, incluidos tanto expertos como usuarios finales, y la aplicación de diversos métodos, se identificaron áreas susceptibles de mejora.

Finalmente, se propone como trabajo futuro, diferenciar las dimensiones de usabilidad y utilizar el modelo en escenarios más complejos con usuarios reales que requieran el

¹USE-AB-Tool: <https://github.com/ari-dasci/S-USE-AB-Tool/>

apoyo de tecnologías de asistencia en línea. En tal caso, no solo se aplicaría el modelo propuesto, si no comprobaría si dichas tecnologías son habilitadoras o, por el contrario, son inhabilitadoras por no estar correctamente integradas con el sistema evaluado.

Bibliografía

- [1] Y.-C. Pan, Y.-C. Chiu, and Y.-H. Lin, “Systematic review and meta-analysis of epidemiology of internet addiction,” *Neuroscience & Biobehavioral Reviews*, vol. 118, pp. 612–622, 2020.
- [2] M. Denzer, T. Schank, and R. Upward, “Does the Internet increase the job finding rate? evidence from a period of expansion in internet use,” *Information Economics and Policy*, vol. 55, p. 100900, 2021.
- [3] ITU [Internet], “Discapacidad [internet],” Available: <https://datahub.itu.int/>, [Consultado el 11 de marzo de 2024].
- [4] Statista, “El uso de Internet a nivel mundial— datos estadísticos [internet],” Available: <https://es.statista.com/estadisticas/541434/numero-mundial-de-usuarios-de-internet/>, [Consultado el 28 de febrero 2024].
- [5] S. Glencross, J. Mason, M. Katsikitis, and K. M. Greenwood, “Internet use by people with intellectual disability: Exploring digital inequality—a systematic review,” *Cyberpsychology, Behavior, and Social Networking*, vol. 24, no. 8, pp. 503–520, 2021.
- [6] L. Rainie and B. Wellman, “The Internet in daily life,” *Society and the Internet: how networks of information and communication are changing our lives*, p. 27, 2019.
- [7] C. Faulkner, *The essence of human-computer interaction*. Prentice-Hall, Inc., 1998.
- [8] J. J. Garrett, “Customer loyalty and the elements of user experience,” *Design management review*, vol. 17, no. 1, pp. 35–39, 2006.
- [9] G. B. Ilyas, A. R. Munir, H. Tamsah, H. Mustafa, and Y. Yusriadi, “The influence of digital marketing and customer perceived value through customer satisfaction on

- customer loyalty,” *Journal of Legal, Ethical and Regulatory Issues*, vol. 24, pp. 1–14, 2021.
- [10] G. Zarei, B. Asgarnezhad Nuri, and N. Noroozi, “The effect of Internet service quality on consumers’ purchase behavior: The role of satisfaction, attitude, and purchase intention,” *Journal of Internet Commerce*, vol. 18, no. 2, pp. 197–220, 2019.
- [11] J. Nielsen, R. L. Mack *et al.*, *Usability inspection methods*. Wiley New York, 1994, vol. 1.
- [12] J. A. Martínez Usero, “Directrices para mejorar la accesibilidad a los recursos electrónicos en los servicios de información públicos,” *BiD: textos universitarios de biblioteconomía i documentació*, no. 17, 2006.
- [13] S. Firmenich, A. Garrido, J. Grigera, J. M. Rivero, and G. Rossi, “Usability improvement through A/B testing and refactoring,” *Software Quality Journal*, vol. 27, no. 1, pp. 203–240, 2019.
- [14] B. Martin, B. Hanington, and B. M. Hanington, *Universal methods of design: 100 ways to research complex problems, develop innovative ideas, and design effective solutions*. Rockport Pub, 2012.
- [15] J. Nielsen, “Designing Web Usability,” *New Riders*, no. 667, 2000. [En línea]. Disponible: <https://books.google.com.mx/books?id=aExqAAAAMAAJ>
- [16] M. C. Harrell and M. A. Bradley, “Data collection methods. semi-structured interviews and focus groups,” Rand National Defense Research Inst santa monica ca, Tech. Rep., 2009.
- [17] K. Braunsberger and R. Gates, “Developing inventories for satisfaction and likert scales in a service environment,” *Journal of Services Marketing*, vol. 23, no. 4, pp. 219–225, 2009.
- [18] L. A. Zadeh, “Fuzzy logic= computing with words,” in *Computing with Words in Information/Intelligent Systems 1*. Springer, 1999, pp. 3–23.
- [19] L. Martínez, D. Ruan, and F. Herrera, “Computing with Words in Decision Support Systems: An overview on Models and Applications,” *International Journal of Computational Intelligence Systems*, vol. 3, no. 4, pp. 382–395, 2010.

- [20] E. Triantaphyllou, *Multi-Criteria Decision Making Methods*. Boston, MA: Springer US, 2000, pp. 5–21.
- [21] S. A. Adepoju, I. O. Oyefolahan, M. B. Abdullahi, and A. A. Mohammed, “A survey of research trends on university websites’ usability evaluation,” *i-manager’s Journal on Information Technology*, 2019.
- [22] B. K. Eldrandaly, A. A. Al, R. K. Chakraborty, and M. Abdel-Basset, “An efficient framework for evaluating the usability of academic websites: Calibration, validation, analysis, and methods,” *Neutrosophic Sets and Systems*, vol. 53, no. 1, p. 12, 2023.
- [23] G. Agrawal, A. Dumka, and M. Singh, “Usability and accessibility-based quality evaluation of Indian airline websites: An MCDM approach,” *Universal Access in the Information Society*, 2022.
- [24] H. Li and V. C. Yen, *Fuzzy sets and fuzzy decision-making*. CRC press, 1995.
- [25] R. Montes, C. Zuheros, J. Morales, N. Zermeño, J. Duran, and F. Herrera, “Design and consensus content validity of the questionnaire for b-learning education: A 2-tuple fuzzy linguistic delphi based decision support tool,” *Applied Soft Computing*, vol. 147, p. 110755, 2023.
- [26] M. C. Zuheros Montes, “Procesos de toma de decisiones en grupos de multitud considerando opiniones en redes sociales analizadas mediante deep learning,” Tesis Univ. Granada., 2023. [En línea]. Disponible: <https://hdl.handle.net/10481/82555>
- [27] F. Herrera and L. Martínez, “A model based on linguistic 2-tuples for dealing with multigranular hierarchical linguistic contexts in multi-expert decision-making,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 31, no. 2, pp. 227–234, 2001.
- [28] L. Zadeh, “The Concept of a Linguistic Variable and its Applications to Approximate Reasoning I, II and III. J. of Information Sciences 8 and 9,” 1975.
- [29] L. A. Zadeh, “Outline of a new approach to the analysis of complex systems and decision processes,” *IEEE Transactions on systems, Man, and Cybernetics*, no. 1, pp. 28–44, 1973.

- [30] Z. Zhang, C. Guo, and L. Martínez, “Managing multigranular linguistic distribution assessments in large-scale multiattribute group decision making,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 11, pp. 3063–3076, 2016.
- [31] J. Brooke, “SUS: A Retrospective,” *J. Usability Studies*, vol. 8, no. 2, pp. 29–40, Feb. 2013.
- [32] T. Malhotra and A. Gupta, “A systematic review of developments in the 2-tuple linguistic model and its applications in decision analysis,” *Soft Computing*, pp. 1–35, 2020.
- [33] F. Herrera and L. Martínez, “A 2-tuple fuzzy linguistic representation model for computing with words,” *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 6, pp. 746–752, Dec 2000.
- [34] T. Malhotra and A. Gupta, “A new 2-tuple linguistic approach for unbalanced linguistic term sets,” *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 8, pp. 2158–2168, 2021.
- [35] P. K. Muhuri and P. K. Gupta, “A novel solution approach for multiobjective linguistic optimization problems based on the 2-tuple fuzzy linguistic representation model,” *Applied Soft Computing*, vol. 95, p. 106395, 2020. [En línea]. Disponible: <https://www.sciencedirect.com/science/article/pii/S1568494620303355>
- [36] F. Herrera and L. Martínez, “A 2-tuple fuzzy linguistic representation model for computing with words,” *IEEE Transactions on fuzzy systems*, vol. 8, no. 6, pp. 746–752, 2000.
- [37] L. Martínez and F. Herrera, “An overview on the 2-tuple linguistic model for computing with words in decision making: Extensions, applications and challenges,” *Information Sciences*, vol. 207, pp. 1–18, 2012. [En línea]. Disponible: <http://www.sciencedirect.com/science/article/pii/S0020025512002897>
- [38] S.-P. Wan, “2-tuple linguistic hybrid arithmetic aggregation operators and application to multi-attribute group decision making,” *Knowledge-Based Systems*, vol. 45, pp. 31–40, 2013.

- [39] G. Wei and X. Zhao, “Some dependent aggregation operators with 2-tuple linguistic information and their application to multiple attribute group decision making,” *Expert Systems with Applications*, vol. 39, no. 5, pp. 5881–5886, 2012.
- [40] Y. Xu and H. Wang, “Approaches based on 2-tuple linguistic power aggregation operators for multiple attribute group decision making under linguistic environment,” *Applied Soft Computing*, vol. 11, no. 5, pp. 3988–3997, 2011.
- [41] O. Cordón, F. Herrera, and I. Zwir, “Linguistic modeling by hierarchical systems of linguistic rules,” *IEEE Transactions on fuzzy systems*, vol. 10, no. 1, pp. 2–20, 2002.
- [42] F. Herrera and L. Martínez, “A model based on linguistic 2-tuples for dealing with multigranular hierarchical linguistic contexts in multi-expert decision-making,” *IEEE Transactions on fuzzy systems*, vol. 31, no. 2, p. 227–234, 2001.
- [43] F. Herrera, E. Herrera-Viedma, and L. Martínez, “A fuzzy linguistic methodology to deal with unbalanced linguistic term sets,” *IEEE Transactions on fuzzy Systems*, vol. 16, no. 2, pp. 354–370, 2008.
- [44] M. Espinilla, J. Liu, and L. Martinez, “An extended hierarchical linguistic model for decision-making problems,” *Computational Intelligence*, vol. 27, no. 3, pp. 489–512, 2011.
- [45] A. Bangor, P. Kortum, and J. Miller, “Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale,” *Usability Studies*, vol. 4, no. 3, pp. 114–123, may 2009.
- [46] C.-T. Chen, “Extensions of the topsis for group decision-making under fuzzy environment,” *Fuzzy Sets and Systems*, vol. 114, no. 1, pp. 1 – 9, 2000.
- [47] K. Kumar and H. Garg, “Topsis method based on the connection number of set pair analysis under interval-valued intuitionistic fuzzy set environment,” *Computational and Applied Mathematics*, vol. 37, no. 2, pp. 1319–1329, May 2018.
- [48] D. Siregar, H. Nurdiyanto, S. Sriadhi, D. Suita, U. Khair, R. Rahim, D. Napitupulu, A. Fauzi, A. Hasibuan, M. Mesran, and A. P. U. Siahaan, “Multi-attribute decision making with VIKOR method for any purpose decision,” *Journal of Physics: Conference Series*, vol. 1019, p. 012034, jun 2018.

- [49] S. Tadić, S. Zečević, and M. Krstić, “A novel hybrid mcdm model based on fuzzy dematel, fuzzy anp and fuzzy vikor for city logistics concept selection,” *Expert Systems with Applications*, vol. 41, no. 18, pp. 8112 – 8128, 2014.
- [50] T. L. Saaty, *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*. McGraw-Hill International Book Company, 1980.
- [51] Z. Xu and R. R. Yager, “Dynamic intuitionistic fuzzy multi-attribute decision making,” *International Journal of Approximate Reasoning*, vol. 48, no. 1, pp. 246 – 262, 2008, special Section: Perception Based Data Mining and Decision Support Systems.
- [52] R. Zhang, B. Ashuri, and Y. Deng, “A novel method for forecasting time series based on fuzzy logic and visibility graph,” *Advances in Data Analysis and Classification*, vol. 11, no. 4, pp. 759–783, Dec 2017.
- [53] Y.-C. Chou, H.-Y. Yen, V. T. Dang, and C.-C. Sun, “Assessing the Human Resource in Science and Technology for Asian Countries: Application of Fuzzy AHP and Fuzzy TOPSIS,” *Symmetry*, vol. 11, no. 2, 2019.
- [54] R. de FSM Russo and R. Camanho, “Criteria in AHP: a systematic review of literature,” *Procedia Computer Science*, vol. 55, pp. 1123–1132, 2015.
- [55] D.-Y. Chang, “Applications of the extent analysis method on fuzzy AHP,” *European journal of operational research*, vol. 95, no. 3, pp. 649–655, 1996.
- [56] C. Kahraman, S. C. Onar, and B. Oztaysi, “Fuzzy multicriteria decision-making: a literature review,” *International Journal of Computational Intelligence Systems*, vol. 8, no. 4, pp. 637–666, 2015.
- [57] F. R. L. Junior, L. Osiro, and L. C. R. Carpinetti, “A comparison between Fuzzy AHP and Fuzzy TOPSIS methods to supplier selection,” *Applied Soft Computing*, vol. 21, pp. 194–209, 2014.
- [58] S. H. Zyoud, L. G. Kaufmann, H. Shaheen, S. Samhan, and D. Fuchs-Hanusch, “A framework for water loss management in developing countries under fuzzy environment: Integration of Fuzzy AHP with Fuzzy TOPSIS,” *Expert Systems with Applications*, vol. 61, pp. 86–105, 2016.

- [59] F. Torfi, R. Z. Farahani, and S. Rezapour, “Fuzzy AHP to determine the relative weights of evaluation criteria and Fuzzy TOPSIS to rank the alternatives,” *Applied Soft Computing*, vol. 10, no. 2, pp. 520–528, 2010.
- [60] C.-L. Hwang and K. Yoon, “Methods for multiple attribute decision making,” in *Multiple attribute decision making*. Springer, 1981, pp. 58–191.
- [61] C. Yu, Y. Shao, K. Wang, and L. Zhang, “A group decision making sustainable supplier selection approach using extended topsis under interval-valued pythagorean fuzzy environment,” *Expert Systems with Applications*, vol. 121, pp. 1–17, 2019.
- [62] S. Peng, T. Li, M. Li, Y. Guo, J. Shi, G. Z. Tan, and H. Zhang, “An integrated decision model of restoring technologies selection for engine remanufacturing practice,” *Journal of Cleaner Production*, vol. 206, pp. 598–610, 2019.
- [63] A. Singh, S. Kumari, H. Malekpoor, and N. Mishra, “Big data cloud computing framework for low carbon supplier selection in the beef supply chain,” *Journal of Cleaner Production*, vol. 202, pp. 139–149, 2018.
- [64] R. Kuo, C. Hsu, and Y. Chen, “Integration of fuzzy ANP and fuzzy TOPSIS for evaluating carbon performance of suppliers,” *International journal of environmental science and technology*, vol. 12, no. 12, pp. 3863–3876, 2015.
- [65] H.-C. Liu, M.-L. Ren, J. Wu, and Q.-L. Lin, “An interval 2-tuple linguistic MCDM method for robot evaluation and selection,” *International Journal of Production Research*, vol. 52, no. 10, pp. 2867–2880, 2014.
- [66] J.-Y. Mao, K. Vredenburg, P. W. Smith, and T. Carey, “The state of user-centered design practice,” *Communications of the ACM*, vol. 48, no. 3, pp. 105–109, 2005.
- [67] T. Lowdermilk, *User-centered design: a developer’s guide to building user-friendly applications*. .o’Reilly Media, Inc.”, 2013.
- [68] T. Brown *et al.*, “Design thinking,” *Harvard business review*, vol. 86, no. 6, p. 84, 2008.
- [69] M. Berkowitz, “Product shape as a design innovation strategy,” *Journal of product innovation management*, vol. 4, no. 4, pp. 274–283, 1987.

- [70] R. Kijima, M. Yang-Yoshihara, and M. S. Maekawa, “Using Design Thinking to cultivate the next generation of female STEAM thinkers,” *International Journal of STEM Education*, vol. 8, no. 1, pp. 1–15, 2021.
- [71] N. H. Razali, N. N. N. Ali, S. K. Safiyuddin, and F. Khalid, “Design Thinking approaches in education and their challenges: A systematic literature review,” *Creative Education*, vol. 13, no. 7, pp. 2289–2299, 2022.
- [72] H. Plattner, “Guía del proceso creativo. mini guía: una introducción al design thinking+ bootcamp bootleg,” 2018.
- [73] M. Chesler and R. Fox, “Role-playing methods in the classroom.” *ERIC*, 1966.
- [74] A. Luebbe and M. Weske, *Bringing Design Thinking to Business Process Modeling*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 181–195.
- [75] M. C. Angelides and R. J. Paul, “A methodology for specific, total enterprise, role-playing, intelligent gaming-simulation environment development,” *Decision Support Systems*, vol. 25, no. 2, pp. 89–108, 1999.
- [76] E. K. Armstrong *et al.*, “Applications of role-playing in tourism management teaching: An evaluation of a learning method,” *Journal of Hospitality, Leisure, Sport and Tourism Education*, vol. 2, no. 1, pp. 5–16, 2003.
- [77] T. Berners-Lee and M. Fischetti, *Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor*. Harper San Francisco, 1999.
- [78] T. Cobeña, T. Z. Solorzano, F. Z. Pico, and M. P. Ortega, “Análisis de normas de accesibilidad web en el sitio web de la facultad de informática de la universidad laica “eloy alfaró” de manabí,” *Informática y Sistemas: Revista de Tecnologías de la Informática y las Comunicaciones*, vol. 3, no. 1, pp. 23–31, 2019.
- [79] C. Thorén, “Nordic guidelines for computer accessibility,” in *Inclusive Design Guidelines for HCI*. CRC Press, 2001, pp. 141–148.
- [80] A. E. de Normalización y Certificación, *UNE 139802 Aplicaciones informáticas para personas con discapacidad. Requisitos de accesibilidad al ordenador. Software:*

- Septiembre 2003*. AENOR, 2003, [Consultado: 12 de marzo de 2024]. [En línea]. Disponible: <https://books.google.com.mx/books?id=oQNUQwAACAAJ>
- [81] D. Abril Abadín, S. Gil González, and M. Sebastián Herranz, “Mi interfaz de acceso al ordenador,” 2014.
- [82] C. Rodríguez, “Historia de la accesibilidad en españa,” *Madrid: Ceapat-imserso. Fotografía VII, Ceip de practiques Fotografía VIII, Ceip de pràctiques*, 2014.
- [83] M. L. Guenaga, A. Barbier, and A. Eguiluz, “La accesibilidad y las tecnologías en la información y la comunicación,” *TRANS. Revista de traductología*, no. 11, pp. 155–169, 2007.
- [84] W3C, “World Wide Web Consortium,” [Consultado: 12 de marzo de 2024]. [En línea]. Disponible: <http://w3c.org/>
- [85] Accesible Digital Learning Portal, “Assistive technology guide: Low cost hardware for access.” [Consultado: 13 de marzo de 2024]. [En línea]. Disponible: <https://accessibledigitallearning.org/resource/free-and-low-cost-hardware-for-access/>
- [86] Microsoft, “Accesibilidad (conceptos básicos de diseño).” 2022, [Consultado: 12 de marzo de 2024]. [En línea]. Disponible: <https://learn.microsoft.com/es-es/windows/win32/uxguide/inter-accessibility>
- [87] G. M. Alsalem and I. A. Doush, “Access education: What is needed to have accessible higher education for students with disabilities in jordan?,” *International journal of special education*, vol. 33, no. 3, pp. 541–561, 2018.
- [88] U. N. Publications, *United Nations Treaty*. United Nations Publications, 2006, vol. 2258.
- [89] UNE Normalización Española, “UNE 139803:2012,” 2012, [Consultado: 10 de marzo de 2024]. [En línea]. Disponible: <https://www.une.org/encuentra-tu-norma/busca-tu-norma/norma/?c=N0049614>
- [90] Pascual Almenara, “Accesibilidad en entornos web interactivos: superación de las barreras digitales [Tesis de doctorado, Universitat de Lleida],” 2015, [Consultado: 13

- de marzo de 2024]. [En línea]. Disponible: <https://mpiua.invid.udl.cat/accessibilidad/esfuerzos-normativos-en-el-ambito-de-la-accesibilidad-web/>
- [91] Marco de Desarrollo de la Junta de Andalucía, “Herramientas de edición de contenidos (ATAG),” s.f., [Consultado: 28 de febrero de 2024]. [En línea]. Disponible: <http://www.juntadeandalucia.es/servicios/madeja/contenido/subsistemas/interfaz-usuario/herramientas-edicion-contenidos-atag/>
- [92] Jan Richards, Jeanne Spellman, Jutta Treviranus, “Authoring Tool Accessibility Guidelines (ATAG) 2.0.” 2015, [Consultado: 10 de enero de 2023]. [En línea]. Disponible: <https://www.w3.org/TR/ATAG20/#abstract>
- [93] W3C, “Web Content Accessibility Guidelines (WCAG) 2.0,” Available: <http://www.w3.org/TR/WCAG20/>, 2008.
- [94] E. Fernández-Díaz, M. C. Jambrino Maldonado, P. P. Iglesias Sánchez *et al.*, “Accesibilidad Web. La nueva era de las WCAG 2.1, la transición a las futuras WCAG 3.0,” 2019.
- [95] “Web Content Accessibility Guidelines (WCAG 2.0, ISO/IEC 40500: 2012), author=Initiative, Web Accessibility.”
- [96] C. R. Barrera, S. Núñez Amaro, and D. Motola Pedroso, “Evaluación de sitios web en internet: Propuestas para la evaluación de sitios web de bibliotecas públicas y de salud,” *Acimed*, vol. 14, no. 4, pp. 0–0, 2006.
- [97] J. López-Zambrano, J. Moreira-Pico, and N. Alava-Cagua, “Metodología para valorar y clasificar herramientas de evaluación de accesibilidad web,” *e-Ciencias de la Información*, vol. 8, no. 1, pp. 172–189, 2018.
- [98] I. A. Doush, F. Alkhateeb, E. Al Maghayreh, and M. A. Al-Betar, “The design of RIA accessibility evaluation tool,” *Advances in Engineering Software*, vol. 57, pp. 1–7, 2013.
- [99] F. F Reichheld, “The one number you need to grow,” *Harvard business review*, vol. 81, pp. 46–54, 124, 06 2004.

- [100] T. F. Waddell, B. Zhang, and S. S. Sundar, “Human–computer interaction,” *The International Encyclopedia of Interpersonal Communication*, pp. 1–9, 2015.
- [101] J. Kolko, *Thoughts on interaction design*. Morgan Kaufmann, 2010.
- [102] M. C. Albornoz, “Diseño de interfaz gráfica de usuario,” in *XVI Workshop de Investigadores en Ciencias de la Computación*, 2014.
- [103] A. Marcus, S. Abromowitz, and M. Abulhair, *Design, user experience, and usability*. Springer, 2013.
- [104] C. Abras, D. Maloney-Krichmar, J. Preece *et al.*, “User-centered design,” *Bainbridge, W. Encyclopedia of Human-Computer Interaction*. Thousand Oaks: Sage Publications, vol. 37, no. 4, pp. 445–456, 2004.
- [105] W. Sánchez, “La usabilidad en ingeniería de software: definición y características. ing-novación. reporte de investigación, 2, 7–21,” 2011.
- [106] J. Nielsen, *Usability engineering*. Academic Press, 1993.
- [107] I. ISO, “Iec 9216,” *Information Technology–Software Quality Characteristics and Metrics*, 1998.
- [108] H. A. Delgado Rodríguez *et al.*, “Ux-user experience¿ qué es? y su importancia en diseño web,” 2020. [En línea]. Disponible: <https://disenowebakus.net/ux.php>
- [109] I. Jiménez, “Aplicación de Metodologías Ágiles al Diseño de la UX,” 05 2014.
- [110] M. Jeon, “Emotions and affect in human factors and human–computer interaction: Taxonomy, theories, approaches, and methods,” *Emotions and affect in human factors and human-computer interaction*, pp. 3–26, 2017.
- [111] A. Muhammad, A. Siddique, Q. N. Naveed, U. Khaliq, A. M. Aseere, M. A. Hasan, M. R. N. Qureshi, and B. Shahzad, “Evaluating usability of academic websites through a fuzzy analytical hierarchical process,” *Sustainability*, vol. 13, no. 4, p. 2040, 2021.
- [112] O. I. de Normalización, *ISO/IEC 25000: Software Engineering-Software Product Quality Requirements and Evaluation (SQuaRE)-Guide to SQuaRE*. ISO, 2005.

- [113] M. Azuma, “SQuaRE: the next generation of the ISO/IEC 9126 and 14598 international standards series on software product quality,” in *ESCOM (European software control and metrics conference)*. sn, 2001, pp. 337–346.
- [114] D. C. Mex-Alvarez, L. M. Hernández-Cruz, C. E. Uc-Rios, and J. R. Cab-Chan, “Análisis de usabilidad web a través de métricas estandarizadas y su aplicación práctica en la plataforma SAEFI.” *Revista de Tecnologías Computacionales*, pp. 15–24, 2019.
- [115] J. Nielsen, “Usability testing of international interfaces,” in *Designing user interfaces for international use*, 1990, pp. 39–44.
- [116] C. Wharton, “The cognitive walkthrough method: A practitioner’s guide,” *Usability Inspection Methods, New York*, pp. 105–140, 1994.
- [117] M. H. Blackmon, P. G. Polson, M. Kitajima, and C. Lewis, “Cognitive walkthrough for the web,” in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2002, pp. 463–470.
- [118] D. P. Freedman and G. M. Weinberg, *Handbook of walkthroughs, inspections, and technical reviews*. Dorset House Publishing Company, Incorporated, 1990.
- [119] D. A. Wheeler, B. Brykczynski, and R. N. Meeson Jr, *Software Inspection: An Industry Best Practice for Defect Detection and Removal*. IEEE Computer Society Press, 1996.
- [120] S. Prefasi Gomar, T. Magal Royo, F. Garde Calvo, J. L. Giménez López *et al.*, “Tecnologías de la información y de la comunicación orientadas a la educación de personas con discapacidad cognitiva,” *RELATEC: revista latinoamericana de tecnología educativa*, 2010.
- [121] R. Fidel, “Qualitative methods in information retrieval research,” *Library and information science research*, vol. 15, pp. 219–219, 1993.
- [122] C. Snyder, *Paper prototyping: The fast and easy way to design and refine user interfaces*. Morgan Kaufmann, 2003.
- [123] J. Nielsen and K. Pernice, *Eyetracking web usability*. New Riders, 2010.

- [124] D. D. Salvucci and J. H. Goldberg, “Identifying fixations and saccades in eye-tracking protocols,” in *Proceedings of the 2000 symposium on Eye tracking research & applications*, 2000, pp. 71–78.
- [125] Y. Hassan Montero and F. J. Martín Fernández, “Card sorting: Técnica de categorización de contenidos,” *No Solo Usabilidad*, no. 3, 2004.
- [126] M. Hertzum, *Usability testing: A practitioner’s guide to evaluating the user experience*. Springer Nature, 2022.
- [127] J. Meneses, “El cuestionario,” 2016.
- [128] P. Jordan, B. Thomas, I. McClelland, and B. Weerdmeester, *Usability Evaluation In Industry*. Taylor & Francis, 1996. [En línea]. Disponible: <https://books.google.es/books?id=IfUsRmzAqvEC>
- [129] R. Owen, *Net Promoter Score and Its Successful Application*. Singapore: Springer Singapore, 2019, pp. 17–29.
- [130] P. Stirling, P. J. Jenkins, N. D. Clement, A. D. Duckworth, and J. E. McEachan, “The net promoter scores with friends and family test after four hand surgery procedures,” *Journal of Hand Surgery (European Volume)*, vol. 44, no. 3, pp. 290–295, 2019.
- [131] D. Hamilton, J. V. Lane, P. Gaston, J. Patton, D. Macdonald, A. Simpson, and C. Howie, “Assessing treatment outcomes using a single question: the net promoter score,” *The bone & joint journal*, vol. 96, no. 5, pp. 622–628, 2014.
- [132] K. Chin, V. Diehl, and K. Norman, “Questionnaire for user interface satisfaction,” *Dev. an Instrum. Meas. User Satisf. Human-Computer Interface*, no. 1988, pp. 213–218, 1988.
- [133] D. F. Sittig, G. J. Kuperman, and J. Fiskio, “Evaluating physician satisfaction regarding user interactions with an electronic medical record system.” in *Proceedings of the AMIA Symposium*. American Medical Informatics Association, 1999, p. 400.
- [134] H. X. Lin, Y.-Y. Choong, and G. Salvendy, “A proposed index of usability: a method for comparing the relative usability of different software systems,” *Behaviour & information technology*, vol. 16, no. 4-5, pp. 267–277, 1997.

- [135] J. G. Ponterotto and D. E. Ruckdeschel, “An overview of coefficient alpha and a reliability matrix for estimating adequacy of internal consistency coefficients with psychological research measures,” *Perceptual and Motor Skills*, vol. 105, no. 3, pp. 997–1014, 2007.
- [136] C. Manterola, L. Grande, T. Otzen, N. García, P. Salazar, and G. Quiroz, “Confiabilidad, precisión o reproducibilidad de las mediciones.” *Revista chilena de infectología*, vol. 35, pp. 680 – 688, 00 2018.
- [137] A. Evers, J. Muñiz, C. Hagemeister, A. Høstmælingen, P. Lindley, A. Sjöberg, and D. Bartram, “Assessing the quality of tests: Revision of the efpa review model,” *Psicothema*, vol. 25, no. 3, pp. 283–291, 2013.
- [138] H. F. Kaiser, “An index of factorial simplicity,” *Psychometrika*, vol. 39, no. 1, pp. 31–36, 1974.
- [139] M. S. Bartlett, “Tests of significance in factor analysis,” *British Journal of Mathematical and Statistical Psychology*, vol. 3, no. 2, pp. 77–85, 1950.
- [140] C. S. Ding and S. L. Hershberger, “Assessing content validity and content equivalence using structural equation modeling,” *Structural Equation Modeling*, vol. 9, no. 2, pp. 283–297, 2002.
- [141] K. Hyrkäs, K. Appelqvist-Schmidlechner, and L. Oksa, “Validating an instrument for clinical supervision using an expert panel,” *International Journal of nursing studies*, vol. 40, no. 6, pp. 619–625, 2003.
- [142] C. Werner, T. Bedford, R. M. Cooke, A. M. Hanea, and O. Morales-Nápoles, “Expert judgement for dependence in probabilistic modelling: A systematic literature review and future research directions,” *European Journal of Operational Research*, vol. 258, no. 3, pp. 801–819, 2017.
- [143] R. Hernández Sampieri, C. Fernández Collado, P. Baptista Lucio *et al.*, *Metodología de la investigación*. McGraw-Hill Interamericana México, 2018, vol. 4.
- [144] R. Sampieri, *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta*. McGraw-Hill Education, 2018, Consultado: 2023-09-12. [En línea]. Disponible: <https://books.google.com.mx/books?id=5A2QDwAAQBAJ>

- [145] M. Urrutia Egaña, S. Barrios Araya, M. Gutiérrez Núñez, and M. Mayorga Camus, “Métodos óptimos para determinar validez de contenido,” *Educación médica superior*, vol. 28, no. 3, pp. 547–558, 2014.
- [146] M. Barrios, G. Guilera, L. Nuño, and J. Gómez-Benito, “Consensus in the delphi method: What makes a decision change?” *Technological Forecasting and Social Change*, vol. 163, p. 120484, 2021.
- [147] C. Okoli and S. D. Pawlowski, “The delphi method as a research tool: an example, design considerations and applications,” *Information & management*, vol. 42, no. 1, pp. 15–29, 2004.
- [148] D. L. Streiner, “Starting at the beginning: an introduction to coefficient alpha and internal consistency,” *Journal of personality assessment*, vol. 80, no. 1, pp. 99–103, 2003.
- [149] E. López Gómez *et al.*, “El método delphi en la investigación actual en educación: una revisión teórica y metodológica,” *Educación XXI: revista de la Facultad de Educación*, 2018.
- [150] C.-T. Chen, “Extensions of the TOPSIS for group decision-making under fuzzy environment,” *Fuzzy Sets and Systems*, vol. 114, no. 1, pp. 1–9, 2000.
- [151] Microsoft, “Inclusive 101 design toolkit,” [Online] Available: https://download.microsoft.com/download/b/0/d/b0d4bf87-09ce-4417-8f28-d60703d672ed/inclusive_toolkit_manual_final.pdf, 2016.
- [152] T. Lockwood, *Design thinking: Integrating innovation, customer experience, and brand value*. Simon and Schuster, 2010.
- [153] J. Allen, J. Chudley, A. Maier, and M. Kammerer, *Smashing UX Design: Foundations for Designing Online User Experiences*. John Wiley & Sons, 2012.
- [154] S. Krug, *Don't make me think!: a common sense approach to Web usability*. Pearson Education India, 2000.
- [155] F. Herrera, S. Alonso, F. Chiclana, and E. Herrera-Viedma, “Computing with words in decision making: Foundations, trends and prospects,” *Fuzzy Optimization and Decision Making*, vol. 8, no. 4, pp. 337–364, Dec. 2009.

- [156] J. Sauro and J. R. Lewis, *Quantifying the user experience: Practical statistics for user research*. Morgan Kaufmann, 2016.
- [157] C. B. Reynoso, “Introducción a la arquitectura de software,” *Universidad de Buenos Aires*, vol. 33, 2004.
- [158] G. McGraw, “Software security,” *IEEE Security & Privacy*, vol. 2, no. 2, pp. 80–83, 2004.