

Campo / Field	Contenido / Content
Tipo de contribución / Article type	Artículo original de investigación
Título en español	Barreras de accesibilidad detectables automáticamente en comercio electrónico: heterogeneidad y perfiles multivariados en 50 sitios
Title in English	Automatically Detectable Accessibility Barriers in E-Commerce: Heterogeneity and Multivariate Profiles Across 50 Websites
Autorías / Authors	MARLON ELINOVICH TENECELA-CALDERON — UNIVERSIDAD DE GUAYAQUIL, Ecuador — ORCID: https://orcid.org/0009-0005-0559-7365
Correspondencia / Correspondence	MARLON ELINOVICH TENECELA-CALDERON — marlon.tenecela@asesoreseducativos-ec.com

Resumen

La accesibilidad de los comercios electrónicos condiciona el acceso autónomo a bienes y servicios, pero los recuentos producidos por herramientas automáticas suelen interpretarse sin considerar su distribución, coocurrencia y límites. El objetivo fue caracterizar la heterogeneidad y los perfiles multivariados de barreras detectables con WAVE en 50 sitios de comercio electrónico, y derivar prioridades prudentes de diseño inclusivo. Se realizó un estudio observacional transversal retrospectivo mediante reanálisis de un conjunto público auditado el 25 de marzo de 2021. Se reconciliaron 11 criterios WCAG 2.1 y seis indicadores estructurales. Se emplearon estadísticos robustos, correlaciones de Spearman con 5.000 permutaciones y remuestras, ajuste de Benjamini-Hochberg, componentes principales robustamente escalados y agrupamiento jerárquico con análisis de estabilidad. La carga agregada presentó una mediana de 12,5 detecciones por sitio (RIC: 7,0–62,0) y alcanzó 3.051 errores y errores de contraste. Contenido no textual afectó 86 % de los sitios; contraste mínimo, 80 %; y propósito del enlace, 78 %. Por volumen, contraste representó 54,36 % de las 3.019 detecciones asignadas a criterios, seguido de contenido no textual (25,64 %) y propósito del enlace (11,59 %). Los tres primeros componentes explicaron 82,55 % de la variación, pero ninguna solución de dos a cinco grupos produjo una tipología defendible: el patrón de dos grupos separó 48 sitios de dos observaciones atípicas. Ninguna de las 15 asociaciones estructurales sobrevivió al ajuste por multiplicidad; la mayor fue entre carga por criterios y elementos estructurales ($\rho = 0,402$; IC95 %: 0,161–0,611; p ajustada = 0,051). Las prioridades convergieron en contraste, alternativas textuales y propósito de enlaces. Los resultados describen señales automáticas de una captura histórica; no demuestran conformidad WCAG, accesibilidad integral ni experiencia real de usuarios.

Palabras clave: accesibilidad web; comercio electrónico; WCAG; WAVE; diseño inclusivo; análisis multivariado.

Abstract

E-commerce accessibility affects autonomous access to goods and services, yet automated-tool counts are often interpreted without considering their distribution, co-occurrence, and limitations. This study characterized the heterogeneity and multivariate profiles of WAVE-detectable barriers across 50 e-commerce websites and derived cautious inclusive-design priorities. We conducted a retrospective cross-sectional observational reanalysis of a public dataset audited on March 25, 2021. Eleven WCAG 2.1 criteria and six structural indicators were reconciled. Analyses included robust descriptive statistics, Spearman correlations with 5,000 permutations and bootstrap samples, Benjamini-Hochberg adjustment, robustly scaled principal components, and hierarchical clustering with stability assessment. Aggregate burden had a median of 12.5 detections per site (IQR: 7.0–62.0), totaling 3,051 errors and contrast errors. Non-text content affected 86% of websites, minimum contrast 80%, and link purpose 78%. By volume, contrast accounted for 54.36% of the 3,019 criterion-assigned detections, followed by non-text

content (25.64%) and link purpose (11.59%). The first three components explained 82.55% of variance, but no two-to five-cluster solution yielded a defensible typology: the two-group pattern separated 48 websites from two outlying observations. None of 15 structural associations survived multiplicity adjustment; the largest linked criterion burden and structural elements ($\rho = 0.402$; 95% CI: 0.161–0.611; adjusted $p = 0.051$). Priorities converged on contrast, text alternatives, and link purpose. These findings describe automated signals from a historical snapshot; they do not establish WCAG conformance, comprehensive accessibility, or users' lived experience.

Keywords: web accessibility; e-commerce; WCAG; WAVE; inclusive design; multivariate analysis.

1. Introducción

El comercio electrónico transforma actividades cotidianas —buscar productos, comparar información, crear una cuenta, pagar y solicitar asistencia— en secuencias de interacción digital. Cuando esas secuencias dependen exclusivamente de señales visuales, controles sin nombre accesible, enlaces ambiguos o contrastes insuficientes, las barreras no son ornamentales: afectan la posibilidad de completar una compra con autonomía. Esta dimensión resulta especialmente relevante para personas que utilizan lectores de pantalla, ampliación, navegación por teclado, reconocimiento de voz u otras tecnologías de asistencia.

Las Pautas de Accesibilidad para el Contenido Web (WCAG) organizan requisitos verificables en cuatro principios: perceptible, operable, comprensible y robusto. WCAG 2.2 amplió la versión 2.1, aunque mantuvo la mayoría de sus criterios y estableció que la conformidad se evalúa sobre páginas completas y mediante una combinación de pruebas automáticas y juicio humano (World Wide Web Consortium [W3C], 2023). Por tanto, una herramienta puede identificar señales de alta probabilidad, pero no convertir un subconjunto de comprobaciones en una declaración integral de conformidad.

WAVE es una herramienta ampliamente utilizada para visualizar errores, errores de contraste, alertas, características de accesibilidad, elementos estructurales y uso de Accessible Rich Internet Applications (ARIA). Su utilidad reside en localizar problemas repetibles con rapidez. Sin embargo, W3C advierte que ninguna herramienta determina por sí sola si un sitio es accesible y que los resultados pueden contener omisiones o señales engañosas (W3C Web Accessibility Initiative [WAI], 2024). Estudios comparativos han documentado diferencias de cobertura, completitud y corrección entre herramientas, así como el riesgo de depender exclusivamente de pruebas automáticas (Vigo et al., 2013; Abduganiev, 2017; Alsaedi, 2020).

La brecha entre criterios, automatización y experiencia vivida también está respaldada por estudios con usuarios. Power et al. (2012) registraron 1.383 problemas durante tareas realizadas por 32 personas ciegas y observaron que aproximadamente la mitad estaba cubierta por criterios WCAG 2.0. Schmutz et al. (2017), en cambio, mostraron que una mayor aplicación de recomendaciones de accesibilidad podía mejorar desempeño y valoraciones tanto de personas con discapacidad visual como de participantes sin discapacidad. La lectura conjunta es importante: las pautas son necesarias y pueden producir beneficios, pero la evaluación automatizada de una parte de ellas no sustituye las pruebas con usuarios.

En comercio electrónico, Santoki y Patvardhan (2021) describieron dificultades de clientes con discapacidad visual en sitios de India. Acosta-Vargas et al. (2022) analizaron con WAVE 50 comercios de alta presencia internacional y documentaron 3.019 barreras asociadas con 11 criterios WCAG 2.1. Contraste mínimo concentró 54,4 % de las detecciones; contenido no textual, 25,6 %; y propósito del enlace, 11,6 %. El conjunto subyacente quedó disponible en Mendeley Data con licencia CC BY 4.0 (Acosta-Vargas et al., 2021).

El panorama más amplio confirma la persistencia del problema. En febrero de 2024, WebAIM evaluó un millón de páginas de inicio y detectó 56,8 errores en promedio; 95,9 % presentó fallos WCAG detectables. Bajo contraste apareció en 81,0 % y texto alternativo ausente en 54,5 % de las páginas (WebAIM, 2024). Estas cifras no son directamente combinables con el conjunto de comercio electrónico debido a diferencias de muestra, fecha, versión y procedimiento, pero permiten contextualizar la recurrencia de categorías similares.

La investigación sobre evaluación web ha avanzado desde simples recuentos hacia análisis de procesos, transparencia y heterogeneidad. Una revisión de 92 estudios destacó la necesidad de integrar accesibilidad durante diseño y evaluación, además de documentar herramientas, métodos y participación de usuarios (Campoverde-Molina et al., 2023). Manca et al. (2023) señalaron que la transparencia de las herramientas es esencial para comprender qué se comprueba, cómo se prioriza y qué permanece fuera de alcance. Laamanen et al. (2022) utilizaron múltiples herramientas y análisis por principios para estudiar páginas universitarias, encontrando diferencias marcadas y agrupamientos de mejor y peor desempeño. No obstante, una tipología solo es útil cuando es estable, tiene tamaños suficientes y no es el artefacto de unos pocos valores extremos.

El vacío abordado en este trabajo no consiste en repetir los totales del estudio fuente. Se propone una relectura multivariada que distingue tres propiedades: extensión entre sitios, volumen de detecciones y coocurrencia entre criterios. También se examina si la complejidad estructural y el uso de ARIA se relacionan con la carga de problemas, y si existen perfiles suficientemente estables para sostener una tipología. Esta aproximación responde a una cuestión de diseño inclusivo: una prioridad no debería depender solo del número agregado, pues un criterio muy concentrado en pocos sitios plantea una intervención distinta de otro extendido en la mayoría.

El objetivo general fue caracterizar los perfiles multivariados de barreras de accesibilidad detectables automáticamente en 50 sitios de comercio electrónico y derivar prioridades prudentes de diseño inclusivo a partir de su prevalencia, volumen y coocurrencia. Los objetivos específicos fueron: (a) auditar y reconciliar la calidad interna del conjunto; (b) describir la distribución robusta de barreras; (c) examinar asociaciones entre carga, indicadores estructurales y ARIA; (d) evaluar la existencia y estabilidad de perfiles; y (e) ordenar prioridades sin equiparar detección automática con conformidad WCAG.

2. Metodología

2.1. Diseño, fuente y unidad de análisis

Se realizó un estudio observacional transversal retrospectivo de datos secundarios públicos. La fuente fue *Dataset Web accessibility e-commerce website*, versión 1, depositada en Mendeley Data bajo licencia CC BY 4.0 (Acosta-Vargas et al., 2021). El archivo original 05042021data_set_ecommerce.xlsx se preservó con SHA-256 36E502D9D6DD827CC1D43CF08CA39ECFD65CF54E35E655649A7A49BC53668EB.

La unidad fue el sitio o página de comercio electrónico auditado. La matriz incluyó 50 observaciones seleccionadas por la fuente a partir de una clasificación internacional de comercios. La fecha registrada en el libro fue 25 de marzo de 2021. La muestra incluyó sitios de Estados Unidos, Reino Unido, China ampliada, Alemania, Francia, Rusia y otros países, con tamaños nacionales desiguales. El país se trató únicamente como descriptor; no se efectuaron comparaciones inferenciales nacionales.

2.2. Variables y reconciliación

Se recuperaron recuentos para 11 criterios WCAG 2.1: contenido no textual (1.1.1), información y relaciones (1.3.1), contraste mínimo (1.4.3), teclado (2.1.1), evitar bloques (2.4.1), propósito del enlace en contexto (2.4.4), encabezados y etiquetas (2.4.6), idioma de la página (3.1.1), identificación de errores (3.3.1), etiquetas o instrucciones (3.3.2) y nombre, función y valor (4.1.2). Se conservaron además los campos de WAVE: errores, errores de contraste, alertas, características, elementos estructurales y ARIA.

El libro contenía 21 hojas, varias de las cuales repetían, ordenaban o resumían las mismas 50 filas. La matriz se reconstruyó desde la tabla por sitio de la hoja 5, no desde tablas dinámicas ni totales secundarios. Los vacíos en recuentos se interpretaron como cero conforme al patrón del libro. Una celda de Home Depot en el criterio 1.4.3 contenía 41,025641..., único valor no entero entre los criterios. Se redondeó a 41 porque la variable es un recuento y el total publicado del criterio es 1.641; el valor original se conservó en el registro de decisiones.

La suma de criterios fue 3.019; la suma de los campos agregados errores más errores de contraste fue 3.051. No se forzó igualdad entre ambos totales porque las taxonomías del libro no son completamente equivalentes. Se

analizaron por separado: *carga por criterios* y *carga agregada*. La primera fue la suma de los 11 criterios; la segunda, errores más errores de contraste. También se calculó el número de criterios presentes por sitio.

2.3. Descripción y prioridad de diseño

Debido a la asimetría y a los valores extremos, se informaron mediana, rango intercuartílico (RIC), mínimo, máximo, sitios afectados y participación en el total. La prioridad de cada criterio se resumió mediante un índice transparente de tres componentes: 40 % prevalencia entre sitios, 40 % participación en el volumen total y 20 % centralidad media de sus correlaciones absolutas con los demás criterios. El índice ordena problemas para discusión de diseño; no es una escala validada de severidad, impacto humano o riesgo jurídico.

2.4. Asociaciones y multiplicidad

Se calcularon correlaciones de Spearman entre tres resultados —carga agregada, carga por criterios y número de criterios presentes— y cinco predictores: posición en la lista, alertas, características, elementos estructurales y ARIA. Para cada correlación se estimó un intervalo percentil del 95 % mediante 5.000 remuestras y un valor p bilateral mediante 5.000 permutaciones. Las 15 comparaciones se ajustaron con Benjamini-Hochberg. La semilla reproducible fue 20250315.

2.5. Componentes y perfiles

Los 11 recuentos de criterios se transformaron mediante $\log(1+x)$, se centraron por mediana y se escalaron por RIC. Se aplicó descomposición en componentes principales mediante valores singulares. Para explorar perfiles se utilizó agrupamiento jerárquico por enlace promedio y distancia euclidiana sobre las variables transformadas.

Se evaluaron soluciones de dos a cinco grupos mediante silueta, tamaño mínimo y concordancia con la solución obtenida al excluir sucesivamente una observación. Antes del análisis se fijó una regla: una tipología solo se consideraría defendible si todos sus grupos contenían al menos cinco sitios. Un grupo menor se interpretaría como señal atípica, no como segmento generalizable.

2.6. Ética, transparencia y reproducibilidad

No se reclutaron personas ni se trataron datos de compradores. Los datos proceden de páginas públicas y se presentan con fines de reanálisis científico. Los nombres comerciales se mantienen solo para trazabilidad y no sustentan afirmaciones sobre el estado actual, la conformidad legal o la calidad global de las empresas. No fue aplicable la revisión ética de investigación con seres humanos.

El expediente conserva el archivo original, protocolo, código, matriz reconciliada, estadísticos, correlaciones, cargas, figuras y semillas. Se siguió STROBE de manera adaptada a un análisis transversal de datos secundarios. Las decisiones de depuración fueron registradas y no se modificó el archivo fuente.

3. Resultados

3.1. Calidad, cobertura y distribución

Se analizaron 50 sitios sin excluir observaciones. Los 11 criterios sumaron 3.019 detecciones, mientras errores y errores de contraste totalizaron 3.051. La mediana de la carga agregada fue 12,5 por sitio (RIC: 7,0–62,0; rango: 1–406). La carga por criterios también presentó mediana de 12,5 (RIC: 7,0–75,25; rango: 2–419). Cada sitio tuvo al menos un criterio detectado; la mediana fue cinco criterios presentes (RIC: 3–6; rango: 1–11).

La distribución fue altamente desigual. Los cinco sitios por encima del percentil 90 de carga por criterios fueron Wildberries (419), Kaola (296), Auchan (294), Magazine Luiza (242) y Otto (238). Esta concentración justificó estadísticos robustos y análisis de sensibilidad en lugar de medias como única medida.

Tabla 1. Distribución de las barreras detectadas por criterio

Criterio	Barrera	Sitios afectados, n (%)	Mediana (RIC)	Máximo	Total	% del volumen
1.4.3	Contraste mínimo	40 (80)	5,0 (1,0–20,0)	264	1.641	54,36
1.1.1	Contenido no textual	43 (86)	2,0 (1,0–19,75)	115	774	25,64
2.4.4	Propósito del enlace	39 (78)	2,5 (1,0–9,75)	45	350	11,59
1.3.1	Información y relaciones	33 (66)	1,0 (0,0–2,0)	25	94	3,11
4.1.2	Nombre, función y valor	9 (18)	0,0 (0,0–0,0)	20	51	1,69
2.4.6	Encabezados y etiquetas	29 (58)	1,0 (0,0–1,0)	6	44	1,46
3.3.2	Etiquetas o instrucciones	25 (50)	0,5 (0,0–1,0)	6	35	1,16
2.4.1	Evitar bloques	11 (22)	0,0 (0,0–0,0)	2	12	0,40
2.1.1	Teclado	8 (16)	0,0 (0,0–0,0)	1	8	0,26
3.1.1	Idioma de la página	8 (16)	0,0 (0,0–0,0)	1	8	0,26
3.3.1	Identificación de errores	2 (4)	0,0 (0,0–0,0)	1	2	0,07

Contenido no textual alcanzó la mayor extensión entre sitios (86 %), aunque contraste mínimo concentró más del doble de detecciones. Propósito del enlace combinó alta extensión (78 %) con el tercer mayor volumen. Los tres criterios reunieron 91,59 % de todas las detecciones asignadas.

3.2. Prioridades inclusivas

El índice compuesto ordenó primero contraste mínimo (0,589), contenido no textual (0,510) y propósito del enlace (0,406). Información y relaciones ocupó el cuarto lugar (0,345) por su extensión y centralidad, pese a representar solo 3,11 % del volumen. Encabezados y etiquetas (0,304) y etiquetas o instrucciones (0,272) completaron el grupo medio.

Este orden distingue volumen de propagación. Por ejemplo, nombre, función y valor acumuló 51 detecciones, más que encabezados y etiquetas, pero apareció en solo nueve sitios; su prioridad compuesta fue menor. La lectura es operativa: contraste, alternativas textuales y significado de enlaces son candidatos para controles sistemáticos tempranos, mientras los problemas menos extendidos requieren revisión focalizada y humana.

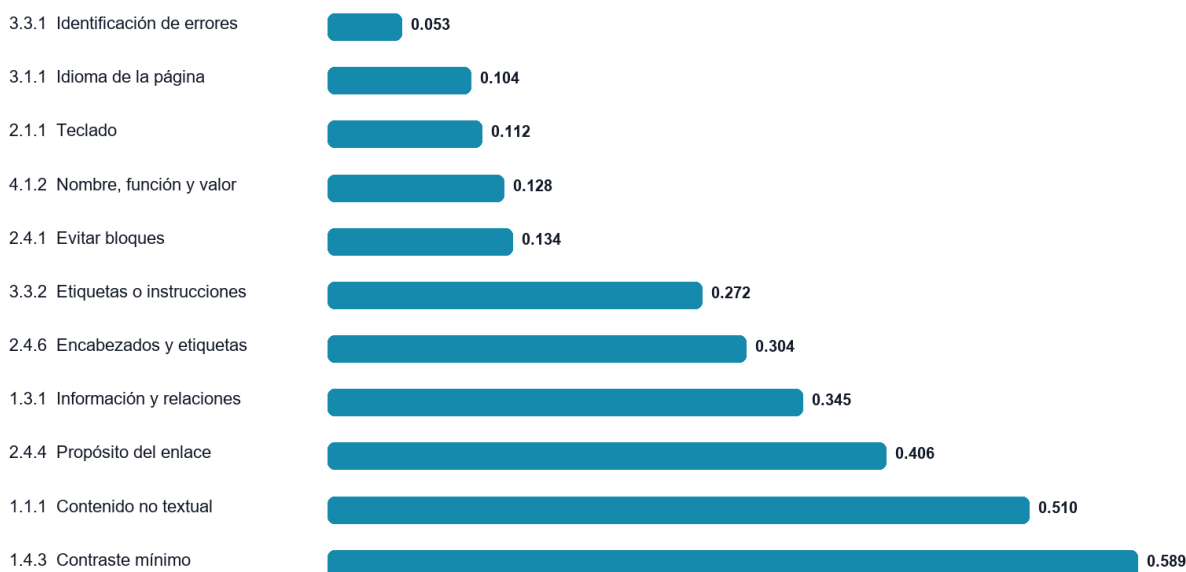
Tabla 2. Primeras seis prioridades del índice compuesto

Posición	Criterio	Barrera	Prevalencia, %	Volumen, %	Centralidad	Índice
1	1.4.3	Contraste mínimo	80,0	54,36	0,257	0,589
2	1.1.1	Contenido no textual	86,0	25,64	0,317	0,510
3	2.4.4	Propósito del enlace	78,0	11,59	0,238	0,406
4	1.3.1	Información y relaciones	66,0	3,11	0,343	0,345

Posición	Criterio	Barrera	Prevalencia, %	Volumen, %	Centralidad	Índice
5	2.4.6	Encabezados y etiquetas	58,0	1,46	0,330	0,304
6	3.3.2	Etiquetas o instrucciones	50,0	1,16	0,336	0,272

Prioridad de barreras detectadas con WAVE

Índice compuesto: prevalencia, volumen y coocurrencia (n = 50 sitios)



Fuente: elaboración propia a partir de Acosta-Vargas et al. (2021), DOI 10.17632/s33r57h5zm.1.

Figura 1. Prioridad de las barreras detectadas con WAVE según prevalencia, volumen y coocurrencia.

3.3. Asociación con estructura, alertas y ARIA

La mayor asociación se observó entre carga por criterios y elementos estructurales ($\rho = 0,402$; IC95 %: 0,161–0,611; p por permutación = 0,003; p ajustada = 0,051). La carga agregada mostró una asociación menor con elementos estructurales ($\rho = 0,361$; IC95 %: 0,079–0,597; p ajustada = 0,071). Alertas se relacionó con carga por criterios ($\rho = 0,343$; IC95 %: 0,077–0,562; p ajustada = 0,071).

ARIA presentó correlaciones negativas de magnitud pequeña a moderada con carga agregada ($\rho = -0,308$; IC95 %: -0,577–0,003) y por criterios ($\rho = -0,289$; IC95 %: -0,548–0,004), pero ambos intervalos incluyeron cero y las p ajustadas fueron 0,095. La posición en la lista y las características no mostraron asociaciones relevantes.

Tabla 3. Asociaciones principales con carga de barreras

Resultado	Predictor	ρ	IC95 %	p permutación	p ajustada
Carga por criterios	Elementos estructurales	0,402	0,161–0,611	0,003	0,051
Carga agregada	Elementos estructurales	0,361	0,079–0,597	0,012	0,071
Carga por criterios	Alertas	0,343	0,077–0,562	0,014	0,071
Carga agregada	ARIA	-0,308	-0,577–0,003	0,030	0,095
Carga por criterios	ARIA	-0,289	-0,548–0,004	0,038	0,095

Resultado	Predictor	rho	IC95 %	p permutación	p ajustada
Carga agregada	Alertas	0,294	−0,009–0,551	0,037	0,095

Ninguna de las 15 asociaciones sobrevivió al ajuste con $p < 0,05$. La evidencia sugiere, sin confirmar, que páginas con más elementos ofrecen más oportunidades para errores detectables. El número de elementos ARIA no puede interpretarse como calidad: puede representar una implementación útil, redundante o incorrecta.

3.4. Componentes y ausencia de tipología estable

El primer componente explicó 39,12 % de la variación y presentó cargas del mismo signo para los principales criterios de volumen; se interpretó como gradiente general de carga. El segundo explicó 23,97 % y estuvo influido positivamente por nombre, función y valor, y negativamente por encabezados/etiquetas y etiquetas/instrucciones. El tercer componente explicó 19,46 %. En conjunto, los tres primeros representaron 82,55 % de la variación robustamente escalada.

La mejor silueta se obtuvo con dos grupos (0,487), pero sus tamaños fueron 48 y 2. Los dos sitios separados —Kohl's y Nordstrom— se distinguieron por una combinación de criterios menos común, no por la mayor carga total. Las soluciones de tres a cinco grupos generaron al menos un grupo unitario. Como ninguna cumplió el tamaño mínimo preespecificado, no se aceptó una tipología de sitios.

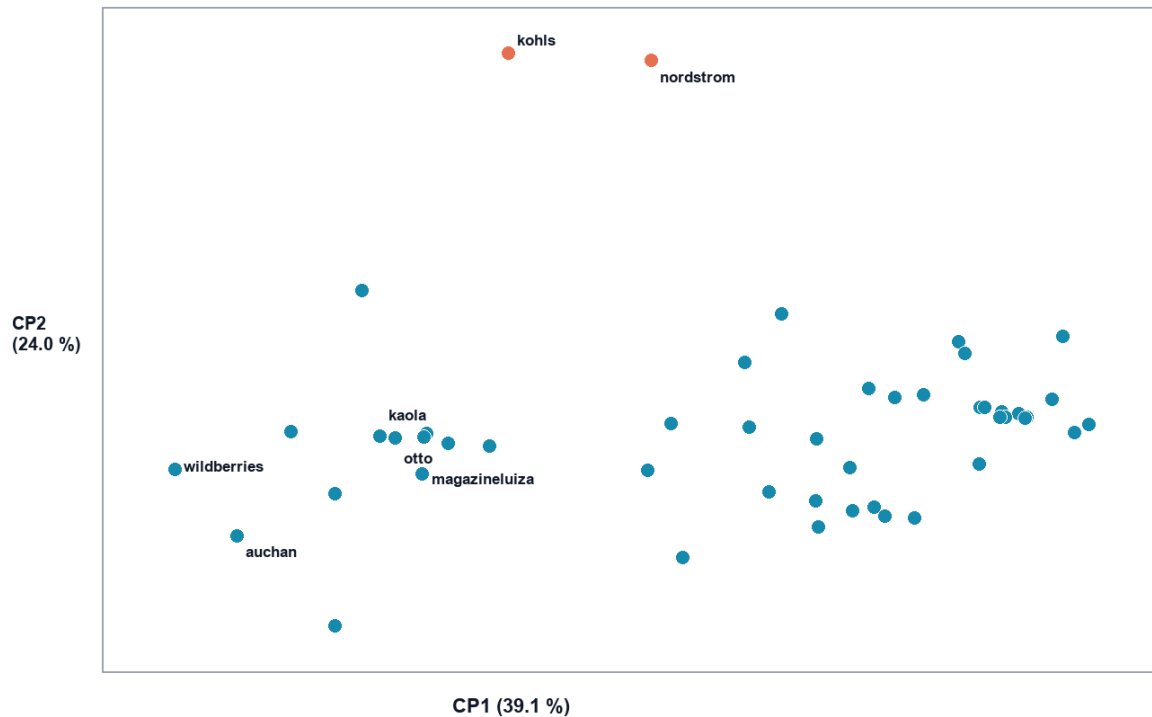
Tabla 4. Evaluación de soluciones de agrupamiento

Grupos	Silueta	Tamaño mínimo	Mediana ARI por exclusión	Dictamen
2	0,487	2	1,000	No defendible; dos observaciones atípicas
3	0,386	1	1,000	No defendible; grupo unitario
4	0,412	1	1,000	No defendible; grupo unitario
5	0,393	1	1,000	No defendible; grupo unitario

La concordancia alta por exclusión no compensó la falta de tamaño: una partición puede repetirse y seguir siendo sustantivamente frágil. El resultado favorece describir gradientes continuos y observaciones atípicas, en lugar de asignar etiquetas de “baja”, “media” o “alta” accesibilidad.

Perfiles multivariados de barreras

Componentes robustamente escalados; colores según agrupamiento exploratorio



Azul: perfil dominante (n = 48). Naranja: dos observaciones multivariadas atípicas; no constituyen una tipología estable.

La estabilidad se evaluó mediante exclusión de una observación y tamaño mínimo de grupo.

Figura 2. Distribución de los sitios en los dos primeros componentes y observaciones multivariadas atípicas.

4. Discusión

El reanálisis confirmó una concentración intensa en tres problemas: contraste mínimo, contenido no textual y propósito del enlace. La contribución adicional fue mostrar que el orden se sostiene cuando se consideran simultáneamente volumen, extensión y coocurrencia. Contraste dominó el volumen; contenido no textual afectó más sitios; y propósito del enlace combinó ambas dimensiones. Para el diseño inclusivo, las tres categorías deben entrar en los componentes reutilizables, reglas de contenido y pruebas continuas, no tratarse como correcciones al final del desarrollo.

La coincidencia con WebAIM (2024) es contextual, no una replicación directa. Contraste y alternativas textuales también estuvieron entre los fallos más frecuentes en un millón de páginas de inicio. La persistencia entre muestras y años sugiere que son problemas sistémicos vinculados con decisiones recurrentes de color, imágenes, iconografía y contenido. Sin embargo, la comparación no permite estimar tendencias en comercio electrónico porque cambian el universo, la herramienta, las reglas y la versión de WCAG.

El contraste constituye una prioridad de alto volumen, pero corregirlo exige algo más que cambiar un color aislado. Los sistemas de diseño deben incorporar pares de color verificados, estados interactivos, texto sobre imágenes, deshabilitados y variaciones responsivas. El contenido no textual requiere decidir qué imágenes son informativas, funcionales o decorativas y redactar alternativas acordes con el propósito. En comercio electrónico, una fotografía de producto, una muestra de color y un icono de acción no cumplen la misma función. El propósito del enlace exige que el nombre accesible sea comprensible en contexto, especialmente en listas repetitivas de productos.

La ausencia de una tipología estable es un resultado relevante. Los recuentos varían ampliamente, pero esa heterogeneidad no se organiza en segmentos robustos con 50 observaciones. Forzar dos grupos habría convertido dos combinaciones atípicas en una supuesta “clase”. Laamanen et al. (2022) hallaron agrupamientos en páginas universitarias mediante otro conjunto y múltiples herramientas; la diferencia recuerda que los perfiles dependen del dominio, variables, muestreo y procedimiento. La segmentación debe probarse, no presuponerse.

Los elementos estructurales presentaron la relación más consistente con la carga, aunque quedó justo por encima del umbral después del ajuste. La dirección es plausible: una página con más controles, regiones, enlaces, imágenes y contenido dinámico ofrece más oportunidades de errores. WebAIM (2024) también informó mayor complejidad promedio y advirtió que la densidad de errores puede ser engañosa. No obstante, este estudio no dispone del denominador exacto para cada criterio ni de medidas de tamaño de página homogéneas. La asociación debe considerarse hipótesis para nuevas capturas, no evidencia causal.

Los resultados de ARIA requieren especial prudencia. Un recuento alto puede indicar aplicaciones ricas y esfuerzos de accesibilidad, pero ARIA mal aplicada puede introducir nombres, estados o relaciones incorrectas. La correlación negativa observada no sobrevivió multiplicidad y su intervalo incluyó cero. W3C establece que la semántica nativa debe preferirse cuando existe y que la conformidad depende del comportamiento completo, no del número de atributos. Por ello, usar “más ARIA” como indicador de accesibilidad sería un error de medición.

La dependencia de una herramienta automática limita el alcance. Vigo et al. (2013) mostraron que las herramientas cubren solo una parte de los criterios y pueden intercambiar completitud por falsos positivos. Alsaeedi (2020) llegó a una conclusión convergente al comparar marcos y herramientas. Manca et al. (2023) añadió que la falta de transparencia dificulta interpretar resultados. En este manuscrito, los recuentos se nombran como *detecciones o señales*, no como experiencia de barrera confirmada en cada caso.

La prioridad automática tampoco sustituye la severidad vivida. Un solo control de compra sin nombre accesible puede bloquear una tarea completa, mientras decenas de contrastes marginales pueden tener efectos diferentes según tamaño y contexto. Power et al. (2012) mostró que numerosos problemas experimentados por personas ciegas quedan fuera de los criterios o no se resuelven con una técnica aplicada superficialmente. En consecuencia, el siguiente nivel de evaluación debe combinar auditoría experta y tareas críticas con personas que utilicen tecnologías de asistencia.

Para educación tecnológica, el conjunto puede convertirse en un laboratorio de lectura crítica. Los equipos pueden comparar volumen, prevalencia y severidad; reconstruir flujos de compra; revisar ejemplos; y observar por qué un total no equivale a accesibilidad. El protocolo de diseño debería integrar accesibilidad desde ideación, arquitectura de información y sistema de componentes, pero medir su efecto requiere un estudio prospectivo real con prototipos y usuarios. El presente reanálisis no afirma haber ejecutado esa intervención.

La matriz de prioridad propuesta es deliberadamente transparente. Sus ponderaciones no son universales y podrían modificarse con evidencia de impacto humano, criticidad de tarea o costo de reparación. Su ventaja es hacer visible que una prioridad depende de varios criterios. Futuras investigaciones pueden validar ponderaciones mediante Delphi, análisis de tareas y participación de personas con discapacidad, además de comparar varias herramientas sobre páginas preservadas.

5. Conclusiones

Las barreras detectables en los 50 comercios electrónicos mostraron una distribución muy asimétrica. Contraste mínimo concentró 54,36 % del volumen; contenido no textual afectó 86 % de los sitios; y propósito del enlace apareció en 78 %. Las tres categorías reunieron 91,59 % de las detecciones asignadas a criterios y ocuparon las primeras posiciones de prioridad.

La complejidad estructural mostró la mayor asociación con la carga, pero ninguna correlación superó el ajuste por multiplicidad. Los componentes describieron un gradiente continuo y combinaciones secundarias, sin una tipología estable: las particiones generaron grupos de uno o dos sitios. La interpretación responsable es heterogeneidad con observaciones atípicas, no clases generalizables.

Los resultados respaldan la incorporación temprana de contraste, alternativas textuales y nombres de enlace en sistemas de diseño, plantillas de contenido y pruebas automatizadas continuas. También muestran que la automatización debe complementarse con revisión experta y pruebas de tareas con usuarios diversos. Una captura WAVE identifica prioridades potenciales; no demuestra conformidad WCAG, accesibilidad integral ni desempeño actual.

6. Limitaciones

La muestra fue intencional y se basó en una clasificación de comercios; no representa probabilísticamente todos los sitios ni permite estimaciones por país. Los datos corresponden a una única captura del 25 de marzo de 2021. Las páginas cambian, por lo que ningún resultado debe atribuirse al estado actual de una empresa.

El libro contiene tablas derivadas y una celda no entera que requirió reconciliación documentada. Aunque el total por criterios coincide con el artículo fuente, los campos agregados suman 32 detecciones adicionales; las taxonomías no se asumieron equivalentes. No se dispuso de HTML preservado, versiones de navegador, configuración exacta de WAVE ni repetición independiente de la captura.

WAVE no cubre todos los criterios ni evalúa por sí sola significado, contexto, orden lógico, facilidad de tarea o compatibilidad completa con tecnologías de asistencia. Los recuentos no se normalizaron por número de elementos susceptibles de cada criterio. El índice de prioridad usa ponderaciones analíticas no validadas y no representa severidad experimentada.

El tamaño de 50 sitios limita la estabilidad de análisis multivariados y la potencia de asociaciones después de multiplicidad. Los componentes y grupos son exploratorios. La regla de cinco sitios evitó sobreinterpretar grupos diminutos, pero no sustituye una muestra mayor. Futuros trabajos deberían preservar páginas, usar varias herramientas, realizar auditoría humana y evaluar tareas críticas con personas con discapacidad.

Agradecimientos

No se declaran agradecimientos.

Declaraciones

Campo / Field	Contenido / Content
Contribuciones CRediT	Conceptualización; metodología; software; validación; análisis formal; investigación; curación de datos; visualización; redacción del borrador original; redacción, revisión y edición; administración del proyecto.
Financiamiento	La investigación no recibió financiación externa específica.
Conflictos de interés	La persona autora declara no tener conflictos de interés.
Aprobación ética	Análisis de páginas web públicas; no involucró participantes ni datos personales de compradores. No fue aplicable la revisión ética de investigación con seres humanos.
Consentimiento	No aplicable al análisis de páginas web públicas.
Disponibilidad de datos	Mendeley Data: https://doi.org/10.17632/s33r57h5zm.1 . Código y derivados en el expediente reproducible.

Campo / Field	Contenido / Content
Uso de inteligencia artificial	Codex de OpenAI apoyó la programación reproducible, la organización del manuscrito, la traducción académica y los controles editoriales. La persona autora revisó el contenido y asume la responsabilidad íntegra por la exactitud, originalidad, interpretación y versión final del artículo.

Referencias

- Abduganiev, S. G. (2017). Towards automated web accessibility evaluation: A comparative study. *International Journal of Information Technology and Computer Science*, 9(9), 18–44. <https://doi.org/10.5815/ijitcs.2017.09.03>
- Abu Doush, I., Sultan, K., Al-Betar, M. A., Almeraj, Z., Alyasseri, Z. A. A., & Awadallah, M. A. (2023). Web accessibility automatic evaluation tools: To what extent can they be automated? *CCF Transactions on Pervasive Computing and Interaction*, 5(3), 288–320. <https://doi.org/10.1007/s42486-023-00127-8>
- Acosta-Vargas, P., Salvador-Acosta, B., Salvador-Ullauri, L., Jadán-Guerrero, J., & Gonzalez, M. (2021). *Dataset Web accessibility e-commerce website* (Version 1) [Data set]. Mendeley Data. <https://doi.org/10.17632/s33r57h5zm.1>
- Acosta-Vargas, P., Salvador-Acosta, B., Salvador-Ullauri, L., Jadán-Guerrero, J., & Gonzalez, M. (2022). Accessibility challenges of e-commerce websites. *PeerJ Computer Science*, 8, e891. <https://doi.org/10.7717/peerj-cs.891>
- Alsaeedi, A. (2020). Comparing web accessibility evaluation tools and evaluating the accessibility of webpages: Proposed frameworks. *Information*, 11(1), 40. <https://doi.org/10.3390/info11010040>
- Campoverde-Molina, M., Luján-Mora, S., & García, L. V. (2023). Accessibility engineering in web evaluation process: A systematic literature review. *Universal Access in the Information Society*. <https://doi.org/10.1007/s10209-023-00967-2>
- Laamanen, M., Ladonlahti, T., Puupponen, H., & Kärkkäinen, T. (2022). Does the law matter? An empirical study on the accessibility of Finnish higher education institutions' web pages. *Universal Access in the Information Society*. <https://doi.org/10.1007/s10209-022-00931-6>
- Manca, M., Palumbo, V., Paternò, F., & Santoro, C. (2023). The transparency of automatic web accessibility evaluation tools: Design criteria, state of the art, and user perception. *ACM Transactions on Accessible Computing*, 16(1), 1–36. <https://doi.org/10.1145/3556979>
- Power, C., Freire, A. P., Petrie, H., & Swallow, D. (2012). Guidelines are only half of the story: Accessibility problems encountered by blind users on the web. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 433–442). ACM. <https://doi.org/10.1145/2207676.2207736>
- Santoki, S., & Patvardhan, N. (2021). An examination on ‘website accessibility’ for active engagement of visually impaired e-commerce customers. *Technology and Disability*, 33(1), 5–16. <https://doi.org/10.3233/TAD-200293>
- Schmutz, S., Sonderegger, A., & Sauer, J. (2017). Implementing recommendations from web accessibility guidelines: A comparative study of nondisabled users and users with visual impairments. *Human Factors*, 59(6), 956–972. <https://doi.org/10.1177/0018720817708397>
- Vigo, M., Brown, J., & Conway, V. (2013). Benchmarking web accessibility evaluation tools: Measuring the harm of sole reliance on automated tests. In *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility*. ACM. <https://doi.org/10.1145/2461121.2461124>

WebAIM. (2024). *The WebAIM Million: The 2024 report on the accessibility of the top 1,000,000 home pages*.
<https://webaim.org/projects/million/2024>

World Wide Web Consortium. (2023). *Web Content Accessibility Guidelines (WCAG) 2.2*.
<https://www.w3.org/TR/2023/REC-WCAG22-20231005/>

World Wide Web Consortium, Web Accessibility Initiative. (2024). *Selecting web accessibility evaluation tools*.
<https://www.w3.org/WAI/test-evaluate/tools/selecting/>

Anexos

No se incluyen anexos.

Información editorial

Campo / Field	Contenido / Content
Recepción / Received	10 de marzo de 2025 / March 10, 2025
Revisión / Revised	5 de abril de 2025 / April 5, 2025
Aceptación / Accepted	28 de abril de 2025 / April 28, 2025
Publicación / Published	29 de junio de 2025 / June 29, 2025
Volumen, número y año / Volume, issue and year	Vol. 1, Núm. 1, enero-junio de 2025

Licencia prevista: Creative Commons Atribución 4.0 Internacional (CC BY 4.0), salvo indicación aplicable al material de terceros.