

Usabilidad y Accesibilidad

Práctica optativa: Uso de un sistema de reconocimiento de voz

1. Objetivos

- Aprender a utilizar un sistema de reconocimiento de voz.
- Aprender a evaluar la accesibilidad de un sitio web con un sistema de reconocimiento de voz.

2. Introducción

Según la Wikipedia¹, “El reconocimiento automático del habla (RAH) o reconocimiento automático de voz es una disciplina de la inteligencia artificial que tiene como objetivo permitir la comunicación hablada entre seres humanos y computadoras”.

Seguramente todos hemos “sufrido” este tipo de sistema cuando hemos llamado al número de teléfono de contacto de una empresa y una grabación nos ha pedido que digamos la razón de nuestra llamada, para después pasar a una sucesión de grabaciones, preguntas y respuestas hasta finalizar la llamada sin obtener la respuesta que buscábamos. Este tipo de sistema se denomina respuesta de voz interactiva o IVR (del inglés *Interactive Voice Response*)².

Estos sistemas se han popularizado mucho en los últimos años porque los podemos encontrar en dispositivos muy diferentes, como el teléfono móvil y el automóvil. Para el ordenador, los primeros sistemas de reconocimiento de voz aparecieron a finales de los años ochenta, aunque los sistemas comerciales más famosos aparecieron a finales de los años noventa. En concreto, tanto IBM ViaVoice³ como Dragon NaturallySpeaking⁴ fueron lanzados en el año 1997. En la actualidad, Dragon NaturallySpeaking sigue siendo uno de los sistemas de reconocimiento de voz más populares.

Un sistema de reconocimiento de voz también es un producto de apoyo que permite controlar un dispositivo, como un ordenador o un teléfono móvil, por medio de la voz, sin necesidad de un teclado o un ratón. Este tipo de producto de apoyo lo utilizan principalmente las personas con discapacidad motora (por ejemplo, cuadriplejía), las personas a las que les falta algún miembro y las personas con alguna otra condición discapacitante como la artritis severa. A veces, también es empleado por algunas personas con discapacidad cognitiva.

3. ¿Qué tengo que hacer?

Básicamente, existen tres tipos de análisis de la accesibilidad de un sitio web:

- Análisis automático.
- Análisis manual.
- Análisis con usuarios.

En las prácticas anteriores ya has realizado el análisis automático y el análisis manual de tu sitio web. También has realizado el análisis con usuarios, usuarios ciegos, en la práctica en la que has aprendido a usar un lector de pantalla.

En esta práctica tienes que analizar la accesibilidad web de tu sitio web otra vez con usuarios, en concreto, con usuarios que utilicen un sistema de reconocimiento de voz. Pero, en vez de hacerlo con usuarios reales, tú te vas a comportar como si fueras un usuario con discapacidad motora que utiliza un sistema de reconocimiento de voz. Para ello, debes aprender a manejar las principales características de

¹https://es.wikipedia.org/wiki/Reconocimiento_del_habla

²https://es.wikipedia.org/wiki/Respuesta_de_voz_interactiva

³https://en.wikipedia.org/wiki/IBM_ViaVoice

⁴https://en.wikipedia.org/wiki/Dragon_NaturallySpeaking

estos sistemas. Afortunadamente, un sistema de reconocimiento de voz es más fácil de utilizar que un lector de pantalla, así que esta práctica es mucho más sencilla de llevar a cabo.

En esta práctica, debes navegar por tu sitio web con el sistema de reconocimiento de voz y debes prestar atención a los problemas que encuentres. Realiza un informe que detalle los problemas o situaciones curiosas que hayas encontrado. En el informe incluye también la respuesta a las siguientes preguntas:

- ¿Te ha costado mucho aprender a usarlo?
- ¿Qué comandos has aprendido a utilizar?
- ¿Te ha parecido fácil de usar?
- ¿Has logrado rellenar y enviar un formulario?
- ¿Has logrado reproducir un vídeo?

4. ¿Cómo lo hago?

Existen muchos programas de reconocimiento de voz. En “List of speech recognition software”⁵ hay una lista de programas clasificados por sistema operativo y dispositivo.

En el caso de Apple, la aplicación “Voice Control” está disponible de forma nativa para Mac⁶ y para sus dispositivos iPhone, iPad y iPod⁷.

En el caso de Microsoft, varias versiones de Windows integran de forma nativa la aplicación “Reconocimiento de voz de Windows”. En la actualidad, esta aplicación solo está disponible para los siguientes idiomas: inglés (Estados Unidos, Reino Unido, Canadá, India y Australia), francés, alemán, japonés, mandarín (chino simplificado y chino tradicional) y español.

La página web “Comandos de reconocimiento de voz de Windows”⁸ enumera los comandos que se pueden emplear con la aplicación de Microsoft. Los comandos están clasificados en:

- Comandos comunes del reconocimiento de voz.
- Comandos de dictado.
- Comandos para el teclado.
- Comandos para signos de puntuación y caracteres especiales.
- Comandos para Windows y aplicaciones.
- Comandos para usar el ratón.

En el Panel de control de Windows, dentro de Accesibilidad, Reconocimiento de voz, están disponibles las opciones para configurar “Reconocimiento de voz de Windows”, tal como se puede ver en la Figura 1.

Una vez activado “Reconocimiento de voz de Windows”, en la parte superior de la pantalla se sitúa la interfaz del mismo, tal como se puede ver en la Figura 2. El botón con forma de micrófono permite activar o desactivar el funcionamiento del sistema.

Los sistemas de reconocimiento de voz tienen varios modos de funcionamiento:

1. Activación de un control diciendo su “nombre accesible” (aceptar, cancelar, cerrar, archivo).
2. Comandos para navegación por teclado (inicio, fin, tabulador, avance página).
3. Comandos para controlar el puntero del ratón.

El modo de funcionamiento normal, cuando una aplicación o página web son accesibles, es mediante el primer modo, mediante la activación de un control diciendo su “nombre accesible”. Los otros dos modos se suelen emplear cuando una aplicación o página web presenta problemas de accesibilidad.

Para entender el funcionamiento de estos sistemas, vamos a ver un ejemplo con la página web⁹ que se muestra en la Figura 3.

⁵https://en.wikipedia.org/wiki/List_of_speech_recognition_software

⁶<https://support.apple.com/en-us/HT210539>

⁷<https://support.apple.com/en-us/HT210417>

⁸<https://support.microsoft.com/es-es/windows/comandos-de-reconocimiento-de-voz-de-windows-9d25ef36-994d-f367-a81a-a326160128c7>

⁹<http://accesibilidadweb.dlsi.ua.es/>

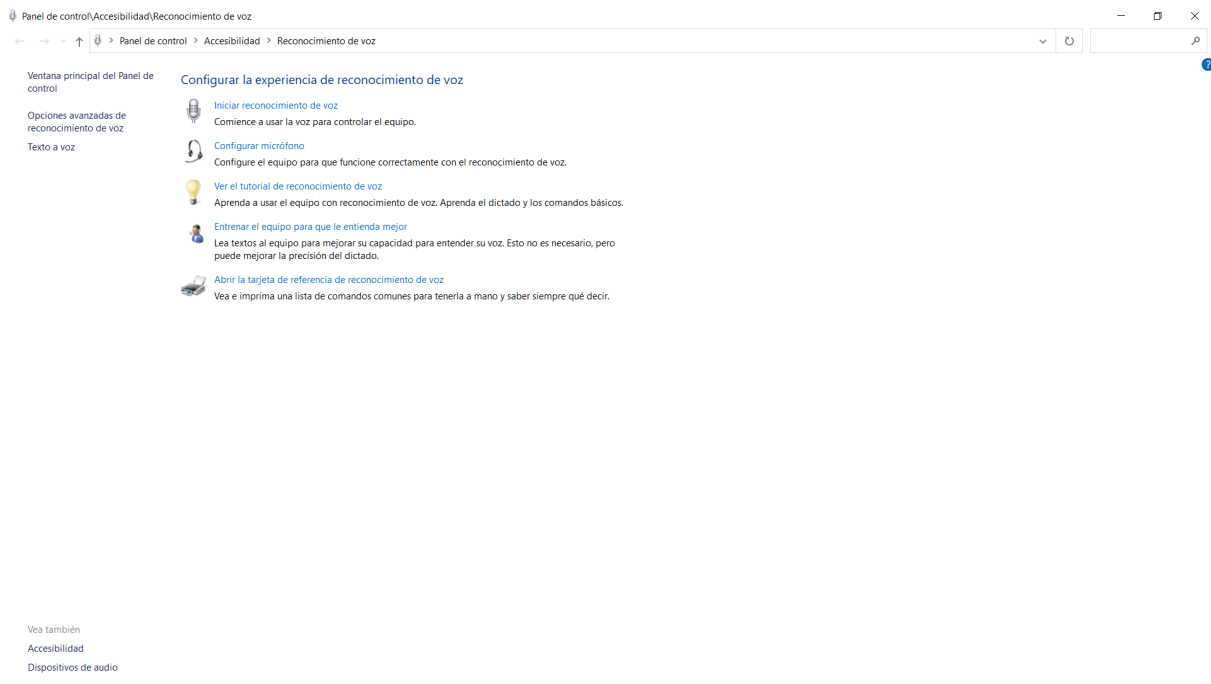


Figura 1: Panel de control de “Reconocimiento de voz de Windows”

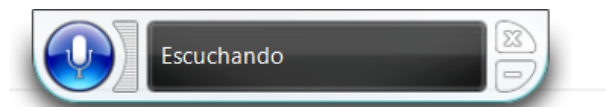


Figura 2: “Reconocimiento de voz de Windows” a la espera de un comando

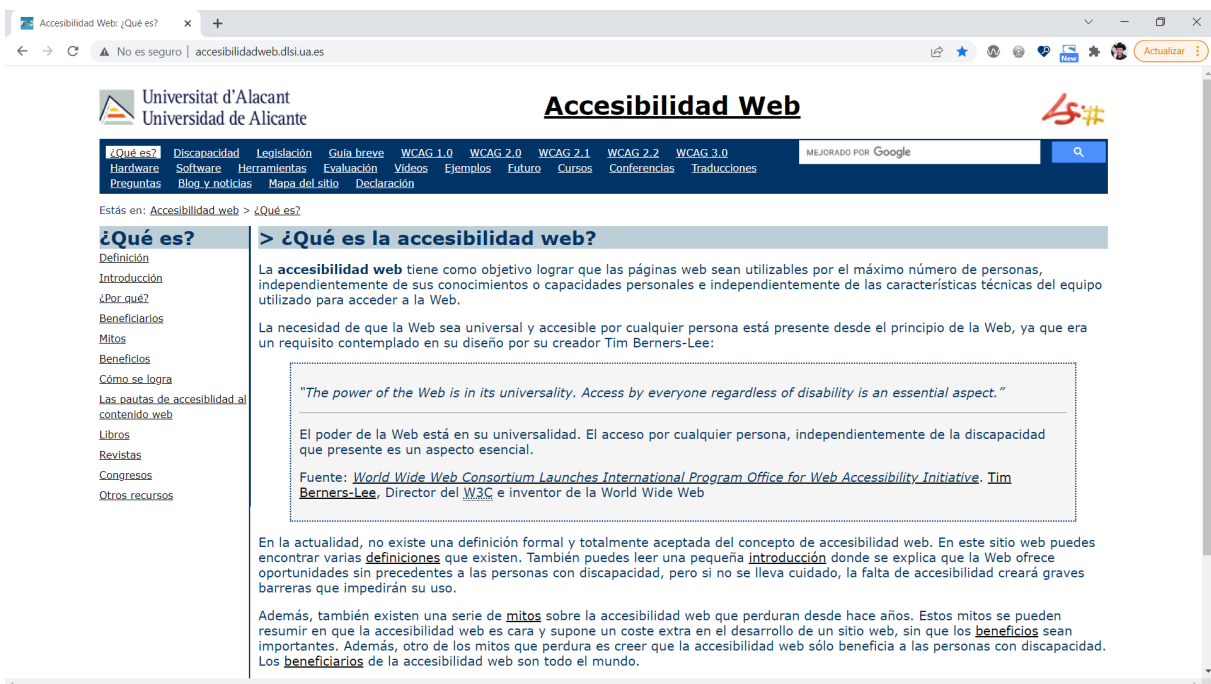


Figura 3: Captura de pantalla de la página “Accesibilidad web”



Figura 4: Desambiguación del enlace “mitos” en una página web

En esta página se puede ver que en el menú del panel izquierdo hay un enlace que pone “Mitos”. Para activarlo, simplemente se puede decir su nombre. Sin embargo, en la página hay dos enlaces con el texto “Mitos”, en el último párrafo se puede ver la oración “Además, también existen una serie de mitos sobre la accesibilidad web. . .” y la palabra “mitos” aparece subrayada porque es un enlace. Para desambiguar el enlace que se desea activar, el sistema de reconocimiento de voz numera los posibles enlaces que son coincidentes, tal como se puede ver en la Figura 4. Para activar uno u otro, simplemente se tiene que decir el número correspondiente, tal como se puede ver en la interfaz del sistema que pone “Diga un número y, después, diga Aceptar”.

Otra forma de activar un elemento de la interfaz es mediante el comando “Mostrar números”. Cuando se ejecuta, el sistema asigna un número a todos los elementos que se puedan activar, tal como se puede ver en la Figura 5.

Otra vez, el sistema queda a la espera y muestra el mensaje “Diga un número y, después, diga Aceptar”. Cuando se dice un número, se ilumina para confirmar que se desea activar ese número, tal como se puede ver en la Figura 6. En este ejemplo, se ha dicho el número 23 para activar el enlace que pone “Accesibilidad Web”.

Otra forma de activar un elemento de la interfaz es mediante una simulación de activación mediante el puntero de ratón. Esta forma se inicia diciendo “Cuadrícula de mouse” o “Cuadrícula de ratón”. Una vez ejecutado, sobre la pantalla se muestra una cuadrícula numerada, tal como se puede ver en la Figura 7.

Al decir un número, el sistema dibuja una nueva cuadrícula en el sector correspondiente al número que se ha dicho. Por ejemplo, en la Figura 8 se puede ver la cuadrícula que se dibuja al decir el número uno (1) en la Figura 7.

Si en la cuadrícula anterior se dice nueve (9), el sistema vuelve a dibujar una nueva cuadrícula, tal como se puede ver en la Figura 9. El proceso finaliza cuando se llegue a una configuración de la cuadrícula en la que no haya ambigüedad, es decir, que solo haya un posible elemento de interacción en el número que se diga. El proceso también se puede finalizar en cualquier momento diciendo “Cancelar”.

Muy importante: no debes emplear ni el ratón ni el teclado cuando utilices el sistema de reconocimiento de voz, emplea únicamente los comandos por voz. Al emplear el ratón o el teclado, no estás simulando el verdadero comportamiento de una persona con discapacidad motora y los resultados pueden ser incorrectos.

En la página “Testing with speech recognition”¹⁰ se explica cómo realizar una evaluación de la accesibilidad con un sistema de reconocimiento de voz. Al final de la página se indican los criterios de conformidad

¹⁰<https://www.tpgi.com/testing-with-speech-recognition/>

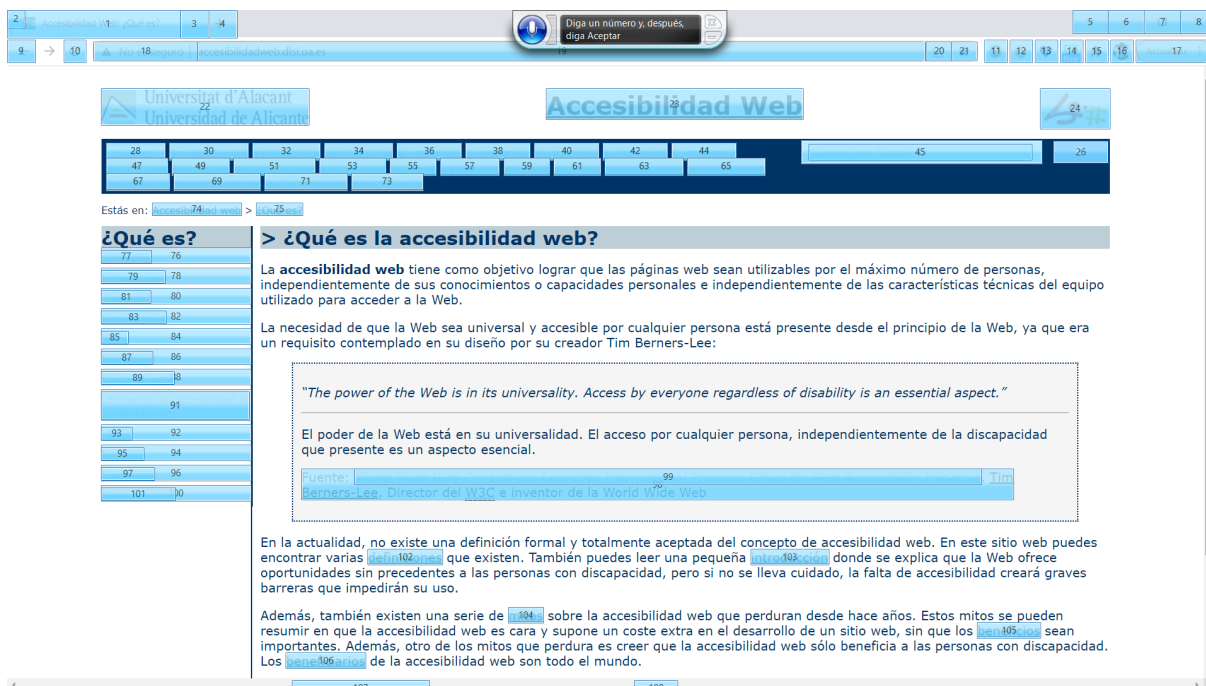


Figura 5: Comando “Mostrar números” en una página web

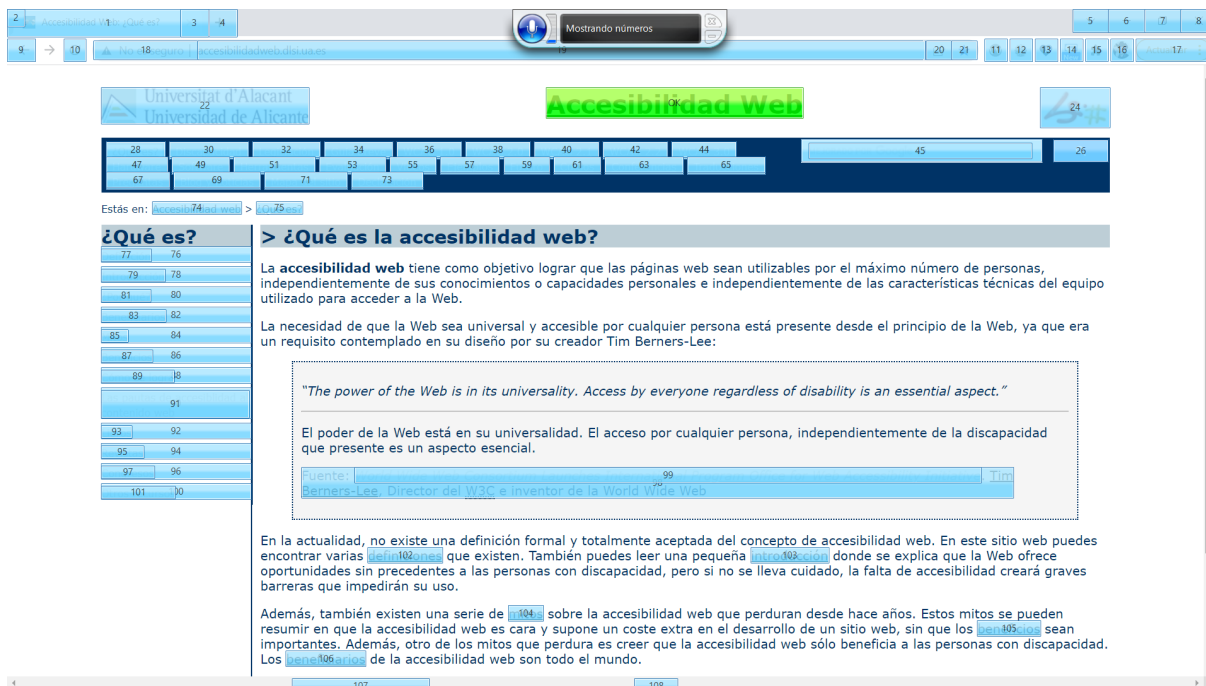


Figura 6: Comando “Mostrar números” en una página web

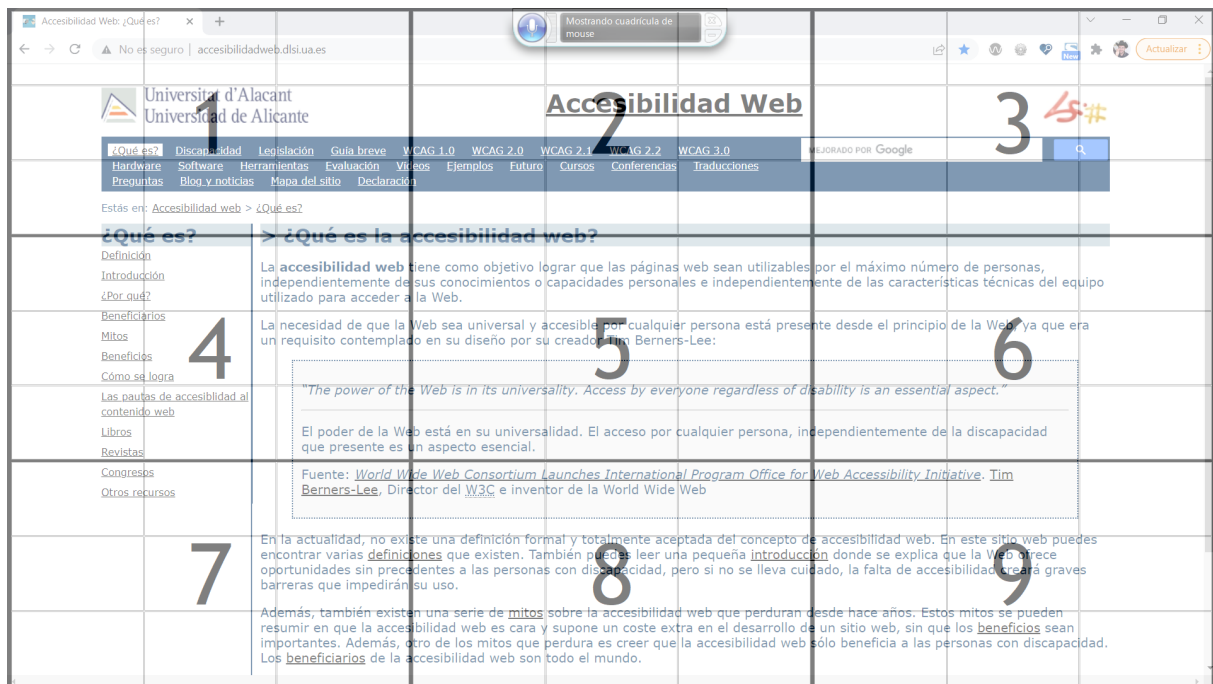


Figura 7: Comando “Cuadrícula de ratón” en una página web

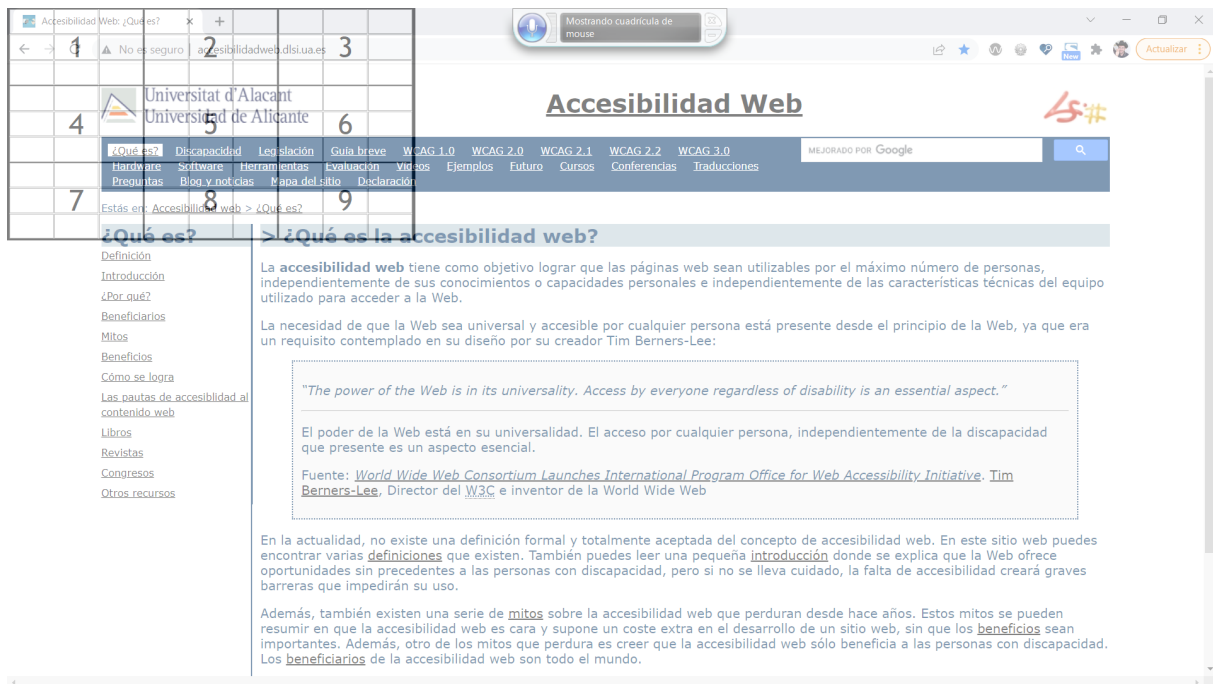


Figura 8: Comando “Cuadrícula de ratón” en una página web

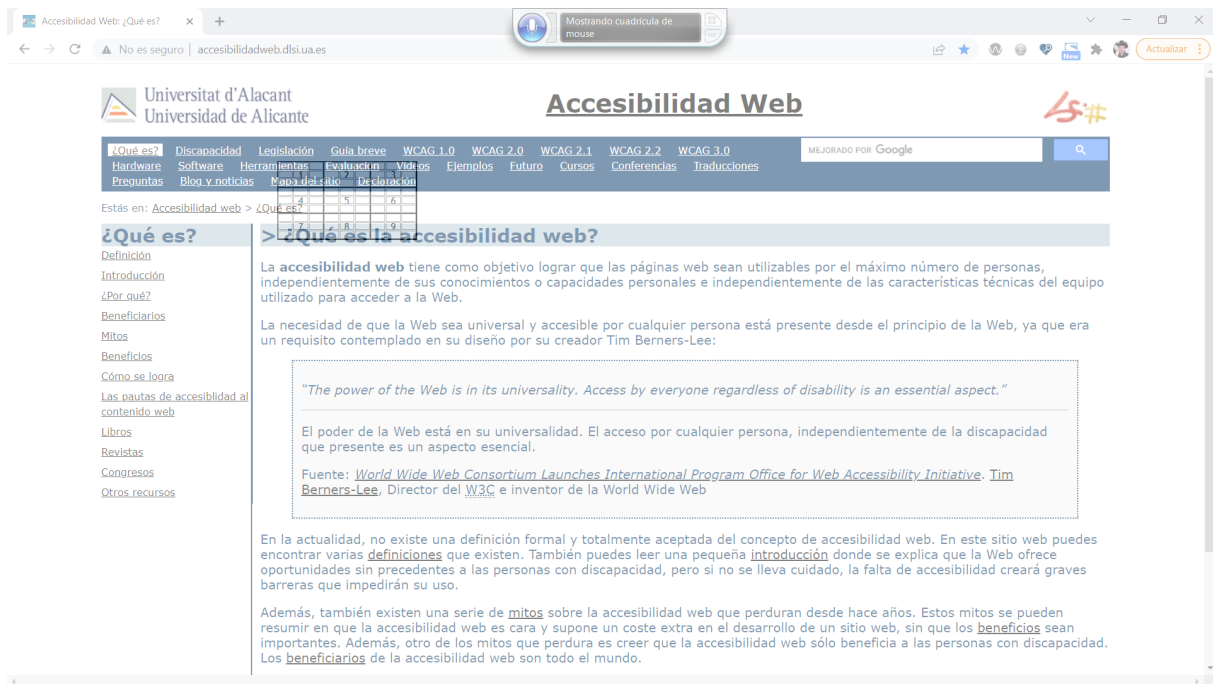


Figura 9: Comando “Cuadrícula de ratón” en una página web

de Web Content Accessibility Guidelines (WCAG) 2.1¹¹ que son importantes para los usuarios de sistemas de reconocimiento de voz. Principalmente, son los criterios relacionados con la accesibilidad mediante teclado y los relacionados con la semántica de los elementos y las limitaciones de interacción:

- Level A
 - 1.3.1 Info and Relationships
 - 2.1.1 Keyboard
 - 2.1.2 No Keyboard Trap
 - 2.2.1 Timing Adjustable
 - 2.4.1 Bypass Blocks
 - 2.4.2 Page Titled
 - 2.4.3 Focus Order
 - 2.4.4 Link Purpose (In Context)
 - 2.5.1 Pointer Gestures
 - 2.5.2 Pointer Cancellation
 - 2.5.3 Label in Name
 - 2.5.4 Motion Actuation
 - 3.2.1 On Focus
 - 3.2.2 On Input
 - 4.1.2 Name, Role, Value
- Level AA
 - 1.3.5 Identify Input Purpose
 - 1.4.13 Content on Hover or Focus
 - 2.4.7 Focus Visible

¹¹<https://www.w3.org/TR/WCAG21/>

- 3.2.4 Consistent Identification
- 3.3.3 Error Suggestion
- 3.3.4 Error Prevention (Legal, Financial, Data)
- 4.1.3 Status Messages
- Level AAA
 - 2.1.3 Keyboard (No Exception)
 - 2.2.3 No Timing
 - 2.5.6 Concurrent Input Mechanisms
 - 3.3.5 Help

5. ¿Qué tengo que entregar?

1. El informe de los problemas de accesibilidad o situaciones curiosas que has encontrado en tu sitio web.
2. Un vídeo tipo *screencasting* que muestre como navegas por tu sitio web y puedes acceder a todo su contenido, especialmente a los formularios, mediante el uso del sistema de reconocimiento de voz.