



# TESINA DE LICENCIATURA

**Título:** Navegación web dirigida por comandos de voz

**Autor:** Javier Pérez

**Directores:** Prof. Javier Díaz, Esp. Ivana Harari

**Carrera:** Licenciatura en Sistemas

## Resumen

El predominio de interacción exclusivamente manual requerido a la hora interactuar tanto con el navegador como con los sitios web, se presenta como una condicionante restrictiva. Esta limitante es tomada como punto de partida y conforma la principal motivación para el desarrollo de un modelo de navegación web dirigido por comandos de voz.

Con el fin de conformar una herramienta que se presente como una alternativa a la navegación web realizada de forma exclusivamente manual, inicialmente se exploraron las diversas dificultades de acceso a la web que afectan a los usuarios con discapacidades motrices y las ayudas técnicas disponibles. Dentro de estas tecnologías de asistencia se profundiza en los diversos aspectos que proponen a los sistemas de reconocimiento de voz como una posible solución de interacción oral.

En concordancia con el objetivo general de este trabajo, y siguiendo el modelo navegacional dirigido por comandos de voz propuesto, se describe una herramienta capaz de permitirle al usuario navegar la web exclusivamente invocando comandos de voz.

## Palabras Claves

Handsfree for Web, navegación web, reconocimiento de voz, accesibilidad web, discapacidad motriz, interfaces de voz, interacción web, funcionalidad web, comandos de voz, extensión Google Chrome, Web Speech API, Chrome Web Store

## Trabajos Realizados

Un análisis acerca de cómo las discapacidades afectan el acceso a los contenidos web.

Una revisión bibliográfica actualizada acerca de los sistemas de reconocimiento de voz e interfaces de voz.

Se describió el conjunto de funcionalidades que hacen a la navegación web y una alternativa de interacción utilizando comandos de voz.

Se desarrollo un modelo de navegación web dirigido por comandos de voz y una extensión de Google Chrome que permite al usuario navegar la web exclusivamente utilizando comandos de voz.

## Conclusiones

Los diversos aplicativos capaces de proveer al usuario la capacidad de controlar la computadora a través del uso de la voz, no brindan la posibilidad de interactuar plenamente con los sitios web como lo puede hacer una persona manualmente.

Se desarrolló un modelo navegacional de interacción oral y una herramienta que permite al usuario navegar la web exclusivamente invocando comandos de voz, al mismo tiempo que propone y hace posible un cambio de paradigma de interacción.

## Trabajos Futuros

- ❖ Portaciones de la herramienta a otras plataformas
- ❖ Mejoras al reconocimiento de comandos de voz
- ❖ Pruebas automatizadas de una interfaz de voz
- ❖ Repositorio de módulos personalizados de comandos de voz
- ❖ Mejorar soporte de elementos presentes en los sitios web
- ❖ Handsfree for Web para personas no videntes

# Agradecimientos

Quisiera agradecer a mis directores, Javier e Ivana, por todo el trabajo y disposición brindado durante este tiempo.

A Belén De Grandis por acompañarme durante todos los años de cursada de la carrera y asistirme en el laborioso proceso de diseño y redacción de este trabajo.

A las personas que colaboraron en las pruebas realizadas, aportando información que resultó de gran interés para evaluar el trabajo realizado. Particularmente a Alicia, Maite, Carlos, Belén, e Iván por el tiempo dedicado.

A los miembros de mi familia que me dieron la oportunidad de iniciar mis estudios universitarios. Especialmente a mis padres, a Abel, Norma y Raúl.

A todos ellos, gracias!

# Resumen

El predominio de interacción exclusivamente manual requerido a la hora interactuar tanto con el navegador como con los sitios web, se presenta como una condicionante restrictiva. Esta limitante es tomada como punto de partida y conforma la principal motivación para el desarrollo de un modelo de navegación web dirigido por comandos de voz.

Con el fin de conformar una herramienta que se presente como una alternativa a la navegación web realizada de forma exclusivamente manual, inicialmente se exploraron las diversas dificultades de acceso a la web que afectan a los usuarios con discapacidades motrices y las ayudas técnicas disponibles. Dentro de estas tecnologías de asistencia se profundiza en los diversos aspectos que proponen a los sistemas de reconocimiento de voz como una posible solución de interacción oral.

En concordancia con el objetivo general de este trabajo, y siguiendo el modelo navegacional dirigido por comandos de voz propuesto, se describe y provee una herramienta capaz de permitirle al usuario navegar la web exclusivamente invocando comandos de voz.

# Abstract

The predominance of exclusively manual interaction required when interacting with the browser and websites is presented as a restrictive condition. This limitation is taken as a starting point and shapes the main motivation for the development of a web navigation model driven by voice commands.

In order to provide a tool that is presented as a handsfree web navigation solution, initially were explored the web accessibility issues that affect users with motor disabilities and the available assistive technologies. Upon examination of these tools, was introduced the aspects that point voice recognition systems as a possible solution to achieve a handsfree interaction.

In agreement with the general objective of this work, and following the navigational model driven by voice commands, is described and provided a handsfree tool able to allow the user to navigate the web exclusively invoking voice commands.

# Índice

<b>Capítulo 1</b>	<b>1</b>
Introducción	1
Motivación	1
Objetivos	2
Organización de la tesis	3
Contribuciones	4
<b>Capítulo 2</b>	<b>5</b>
Fundamentación: Problemáticas de accesibilidad web	5
Discapacidades	5
Ayudas Técnicas	7
<b>Capítulo 3</b>	<b>13</b>
Antecedentes: Interfaces de comandos de voz	13
Tipos de interfaces de voz	13
Sistemas de Reconocimiento de Voz	14
Dificultades en el proceso de reconocimiento de voz	16
Antecedentes de interfaces de voz	18
<b>Capítulo 4</b>	<b>21</b>
Propuesta: Interacción web a través de comandos de voz	21
Comandos de Voz	21
Contextos de ejecución de comandos	23
Modelo de ejecución de comandos voz	25
<b>Capítulo 5</b>	<b>30</b>
Desarrollo de la herramienta: Handsfree for Web	30
Tecnologías utilizadas	30
Modelo de arquitectura y flujos de datos	31
Diseño de la herramienta	32
Funcionalidad Soportada	44
Asistencia al usuario	68
Cuestiones de seguridad	71
<b>Capítulo 6</b>	<b>74</b>
Pruebas de usabilidad	74
Objetivos de la prueba	74
Evaluación	74
Selección de los participantes	75
Análisis de los resultados	76

<b>Capítulo 7</b>	<b>81</b>
Difusión de la Herramienta	81
Chrome Web Store	81
Sitio web oficial de la aplicación	82
<b>Capítulo 8</b>	<b>86</b>
Trabajos futuros	86
Portaciones de la herramienta a otras plataformas	86
Mejoras al reconocimiento de comandos de voz	87
Pruebas automatizadas de una interfaz de voz	89
Repositorio de módulos personalizados de comandos de voz	89
Mejorar soporte de elementos HTML	89
Handsfree for Web para personas no videntes	90
<b>Capítulo 9</b>	<b>91</b>
Conclusiones	91
<b>Anexo 1</b>	<b>93</b>
Soporte de comandos de voz a funcionalidad exclusiva del navegador	93
Gestión de Favoritos	93
Gestión de Descargas	97
Gestión de Historial de Navegación	99
Navegación	101
Buscar texto en página web	103
Imprimir	104
Gestión de Pestañas	104
Opciones en Menú Contextual	106
Desplazamientos dentro del sitio web	108
<b>Anexo 2</b>	<b>110</b>
Soporte de comandos de voz a funcionalidad genérica de los sitios web	110
Elementos <input>	110
Componente Textarea	128
Componente Button	129
Componente Select	129
Componente Select multiple	130
Componente Datalist	132
Componente Keygen	133
Componente Map	134
Componente Audio	135
Componente Video	136
Componente Vínculo	137
<b>Anexo 3</b>	<b>139</b>

Evaluación de usabilidad de la herramienta	139
Instrucciones iniciales para el evaluador	139
Ficha de Evaluación	139
<b>Bibliografía</b>	<b>143</b>

# Capítulo 1

## Introducción

### Motivación

Los sitios web actuales presentan, en su gran mayoría, representaciones exclusivamente visuales y multimediales de la información. Estas cuestiones de formato y diseño dificultan el acceso e interacción con el contenido, que por lo general es alcanzado exclusivamente a través de una interacción manual.

Las personas que poseen discapacidades son afectadas directamente por estas restricciones. Limitaciones como la ceguera y la disminución visual impiden directamente el acceso al contenido, mientras que las personas con problemas motrices poseen graves problemas a la hora de interactuar y hacer uso de las potencialidades de la web. En la actualidad, para que un usuario discapacitado pueda navegar un sitio web debe, en principio, acondicionar su PC con herramientas de adaptación que lo asista en sus necesidades funcionales. Algunos ejemplos de éstas herramientas son; lectores de páginas web, amplificadores de pantalla, teclados especiales, impresoras braille, entre otros. Sin embargo, la adquisición, uso y aprendizaje de estas herramientas informáticas no le garantizan el acceso a los contenidos de la web salvo que estos hayan sido desarrollados de forma accesible. [1]

Dentro del conjunto de personas con discapacidad se encuentran aquellas que están afectadas por algún tipo de disminución motriz, las cuales presentan numerosos obstáculos para manipular el mouse y el teclado convencional y con ello poder acceder e interactuar con los complejos sitios web actuales.

Atendiendo a las estadísticas de Argentina, la publicación de “Población con dificultad o limitación permanente” basada en datos del “Censo Nacional de Población, Hogares y Viviendas 2010”, advierte que el 3.8% de las personas con una discapacidad permanente posee discapacidad motora superior, es decir que presentan alguna de las siguientes afecciones: anomalías congénitas y/o adquiridas, óseas, articulares o amputaciones en hombros y/o brazos, codos o manos con dificultad permanente de movilidad. Se excluye cualquier persona con dificultad temporaria, como por ejemplo a causa de torceduras, fracturas, aplicación de yesos, estado posoperatorio, etcétera. Si se toman en cuenta estas dificultades no permanentes, el porcentaje de personas incapacitadas para manipular una computadora de forma convencional se incrementa significativamente. [2]

Este conjunto concreto de personas poseen el potencial para poder acceder e interactuar con los diferentes tipos de sitios web a través del uso de comandos de voz [3], pero se ven imposibilitadas exclusivamente a la hora de manipular el mouse y el teclado. Sin embargo pueden manifestarse verbalmente, visualizar los contenidos exhibidos por el navegador, y

comprender las complejas operaciones de interacción que proponen las aplicaciones web actuales.

En la actualidad las herramientas que proponen navegación e interacción web a través del uso de comandos de voz presentan significativas problemáticas. La primera y más relevante es la escasa diversidad de operaciones que permiten efectuar sobre una creciente complejidad de interacción que demandan las aplicaciones web actuales. En segundo lugar estas herramientas poseen una licencia de uso comercial o se encuentran restringidas a un sistema operativo específico, produciendo una limitación para aquellas personas con recursos monetarios escasos que no pueden abonar la licencia.

Recientemente se han difundido diferentes aplicaciones para dispositivos móviles que permiten a las personas interactuar a través de la invocación de comandos de voz. Lamentablemente la interacción está limitada al uso de ciertas aplicaciones específicas como el calendario, la alarma o hacer búsquedas en la web, pero no brindan la posibilidad de navegar efectivamente los sitios web.

Las utilidades de una herramienta de navegación web a través de comandos de voz exceden el uso exclusivo de personas con discapacidad. Las mismas pueden satisfacer las necesidades de personas que desean o necesitan por diversos motivos interactuar de una forma no manual.

## **Objetivos**

### **Objetivo general**

- Desarrollar una herramienta que permita la navegación web dirigida exclusivamente por comandos de voz.

### **Objetivos específicos**

- Evaluar el soporte actual de comandos de voz aplicados específicamente a navegación web.
- Analizar y clasificar las operaciones manuales que realiza un usuario con los dispositivos hardware para interactuar con los sitios web.
- Determinar los comandos de voz equivalentes a las operaciones manuales que puede realiza un usuario para interactuar con los sitios web.
- Desarrollar una herramienta extensible, escalable, flexible y modular que pueda ser internacionalizada y utilizada en múltiples lenguajes.
- Implementar un mecanismo por el cual los usuarios puedan extender la herramienta con comandos de voz personalizados.

- Definir un estándar que permita a las aplicaciones web brindar su propio conjunto de comandos de voz soportados.

## **Organización de la tesis**

Atendiendo a la objetivos planteados en este trabajo final, el presente documento se encuentra dividido en nueve capítulos. Estas secciones plantean de forma progresiva las etapas requeridas para el correcto desarrollo y cumplimiento de los objetivos propuestos.

Complementariamente a lo expresado en la motivación de este trabajo final, en el capítulo dos se describen las problemáticas de accesibilidad web que afectan a personas con discapacidades, en particular aquellas que sufren algún tipo de afección motriz que le impida la libre interacción con los contenidos web. Complementariamente se explora y detalla las posibles ayudas técnicas que facilitan a personas con discapacidad el control de la PC.

Una de las ayudas técnicas más prometedoras con la que cuentan en la actualidad las personas con discapacidad motriz, es la posibilidad de controlar verbalmente dispositivos electrónicos. Para ello es indispensable la integración de estos con software de reconocimiento de voz. En el capítulo tres se exploran los distintos tipos de interfaces de voz que presentan sistemas actuales con software de reconocimiento. Complementariamente se describen; las distintas variantes de sistemas de reconocimiento de voz, las dificultades en el procesos de reconocimiento y se enumeran antecedentes de aplicaciones de uso habitual que brindan servicios de interacción a través del uso de la voz.

Con el fin de proveer una herramienta que le permita al usuario interactuar con los sitios web, sin limitaciones y exclusivamente a través de comandos de voz, en el capítulo cuatro, se analizan y describen por extensión y comprensión las acciones que realizan los usuarios al navegar la web. Basado en este análisis de interacción, se describe un modelo de comandos de voz que puede ser implementado con el fin de brindar un mecanismo de interacción verbal en contra oposición a la interacción manual.

A lo largo del capítulo cinco se describen en profundidad los diversos aspectos técnicos y conceptuales correspondientes a la herramienta desarrollada. Inicialmente, se detalla la arquitectura y las tecnologías utilizadas junto con la justificación de las decisiones de diseño tomadas. Seguidamente se profundiza en el soporte del modelo de navegación web dirigido por comandos de voz detallado en el capítulo cuatro. Posteriormente se mencionan los recursos que tienen como fin asistir al usuario mientras el mismo navega la web. Finalmente se analizan los aspectos de seguridad vinculados a la aplicación.

Con el objetivo de validar la utilidad de la herramienta junto con otras características de la misma que se consideran de sumo interés, en el capítulo seis se describe en profundidad

las pruebas de usabilidad realizadas, sus objetivos, participantes y finalmente el análisis y conclusiones sobre los resultados obtenidos.

Debido a la naturaleza del uso social de la aplicación desarrollada como objetivo general de este trabajo, se considera de suma importancia la difusión de la misma. Para ello en el capítulo siete se describen los medios que se utilizarán para difundir públicamente la herramienta junto con la información necesaria para que los posibles usuarios puedan interiorizarse respecto a las características de la misma, su soporte y posibles extensiones.

En el capítulo ocho se exploran las distintas líneas de trabajos futuros que se desprenden de las diversas aristas que hacen a este trabajo final. En este capítulo se fundamentan y describen brevemente posibles extensiones, portaciones y mejoras a los aportes que se brindan en este trabajo.

Finalmente en el capítulo nueve se presentan las conclusiones a las que se llegaron durante el desarrollo de este trabajo.

## **Contribuciones**

Las principales contribuciones que presenta este trabajo son:

- Un análisis acerca de cómo las discapacidades, en particular las producidas por afecciones motrices, afectan el acceso a los contenidos web. Se describen los trastornos físicos más usuales y se exploran las ayudas técnicas disponibles.
- Una revisión bibliográfica actualizada acerca de los sistemas de reconocimiento de voz integrados en aplicaciones que cuentan con interfaces de voz, junto con un análisis acerca de cómo estos están siendo utilizados en la actualidad por dispositivos electrónicos de uso masivo.
- Un relevamiento del conjunto de funcionalidades de usuario soportadas por el navegador Google Chrome y una alternativa de interacción utilizando comandos de voz.
- Un relevamiento del conjunto de operaciones manuales requeridas para interactuar con las funcionalidades soportadas por componentes web que forman parte del estándar HTML 5 y una alternativa de interacción utilizando comandos de voz.
- Un modelo de navegación web dirigido exclusivamente a través de comandos de voz sin pérdida de funcionalidad por parte del usuario a la hora de navegar la web.
- Una extensión de Google Chrome que implementa completamente el modelo especificado de navegación web a través de comandos de voz.

# Capítulo 2

## Fundamentación: Problemáticas de accesibilidad web

Actualmente cerca de la mitad de la población mundial tiene conexión a internet [4]. Las personas diariamente navegan durante horas en la web. Internet hace posible que una persona, con solo presionar un botón, tenga a disposición el acceso a un universo de información y comunicaciones jamás antes visto.

Lamentablemente, esta nueva realidad conectada, se concreta sólo para aquellos que pueden hacer click con el mouse, leer en una pantalla o escuchar un audio. En otras palabras, para personas que no tienen discapacidades.

A pesar de las limitaciones de interacción que presenta la web actual, las personas con ciertas discapacidades pueden acceder, a menudo con dificultades, a ciertos contenidos.

Generalmente el grado de acceso a los contenidos presentes en la web depende del tipo y severidad de la discapacidad que presenta la persona. Afortunadamente no todas las discapacidades impiden el acceso a los contenidos web.

### Discapacidades

En esta sección se presentan brevemente las diferentes discapacidades que pueden afectar el acceso a la web clasificadas funcionalmente [5]. Debido a la naturaleza de este trabajo es de mayor interés describir principalmente las discapacidades de tipo motriz por lo que se va a profundizar en ellas.

#### Discapacidades Sensoriales

##### Discapacidades Visuales

Las discapacidades visuales se presentan en diversos niveles, los cuales van desde la ceguera para cuyo acceso es necesario utilizar lectores de texto o dispositivos braille, hasta baja visión, cuyo acceso a la web puede ser mejorado utilizando magnificadores. Por otra parte, el acceso a los contenidos puede ser dificultado ante la presencia de trastornos visuales como el daltonismo o la falta de sensibilidad a determinados colores.

##### Discapacidades Auditivas

La sordera dificulta la recepción de sonidos presentes en el contenido multimedia. Al igual que las discapacidades visuales esta discapacidad se presenta en niveles, muchos de los

cuales pueden ser solucionados al utilizar audífonos o alternativamente al leer los subtítulos en caso de estar disponibles.

## **Discapacidades Cognitivas**

Dentro de esta categoría se encuentran aquellas enfermedades o trastornos que dificultan la correcta interpretación del contenido. Por ejemplo; la dislexia produce dificultades de comprensión del lenguaje, el trastorno de déficit de atención impide concentrarse en la información y la falta de memoria a corto o largo plazo impide la asimilación de los contenidos. Para mitigar estos problemas, los contenidos web deberían ser presentados de forma clara y sencilla para así de esta forma evitar distraer la atención del usuario.

## **Discapacidades Motoras**

Las discapacidades físicas pueden producir diversos síntomas que son perjudiciales a la hora de interactuar con la PC. Algunos de ellos son; debilidad y limitaciones en el control muscular, insensibilidad, dolor en las articulaciones, movimientos involuntarios, falta de coordinación, ausencia de un miembro y parálisis entre otras. Estas condiciones pueden afectar a las manos y a los brazos entre otras partes del cuerpo.

## **Lesiones Traumáticas**

### *Lesiones en médula espinal*

Este tipo de afección puede dar lugar a un estado de parálisis de los miembros superiores, llamada cuadriplejía. Las personas que sufren esta afección suelen tener severas dificultades a la hora de usar la computadora debido a que no pueden disponer de las manos.

### *Pérdida o daño en los miembros*

A pesar de la pérdida de una mano, es posible acceder a internet sin demasiada dificultad a través del uso de teclados específicos para una mano. Por el contrario la pérdida de ambos miembros puede requerir el uso de otro tipos de dispositivos, los cuales se describen en la siguiente sección.



Figura 2.1:Teclado adaptado para una sola mano

## **Enfermedades congénitas**

### *Parálisis cerebral*

La parálisis cerebral se trata de una lesión en el cerebro cuyo resultado es la pérdida del control muscular. Esta afección puede producir; tirantez en los músculos, espasmos,

movimientos involuntarios y discapacidad en el habla. Los casos agravados pueden conducir a la parálisis. Los pacientes que sufren este tipo de afección suelen tener problemas a la hora de utilizar el mouse.

#### *Distrofia muscular*

Debido a un desorden genético este trastorno produce una degeneración progresiva de los músculos, impidiendo así el uso regular de los dispositivos de acceso a la PC.

#### *Esclerosis múltiples*

Esta enfermedad impide que las fibras nerviosas sean capaces de enviar señales del sistema nervioso central a los músculos del cuerpo. Personas que sufren esta afección poseen una dificultad de moderada a severa a la hora de acceder a internet.

#### *Espina Bífida*

Debido a una malformación en la médula espinal producida en los primeros meses de embarazo, las personas con este trastorno experimentan dificultades motoras y probablemente parálisis.

#### *Esclerosis Lateral Amiotrófica*

La esclerosis lateral amiotrófica o ELA, es una enfermedad de las neuronas en el cerebro, el tronco cerebral y la médula espinal que controlan el movimiento de los músculos voluntarios. La ELA también es conocida como la enfermedad de Lou Gehrig. Los síntomas generalmente no se presentan sino hasta después de los 50 años, pero pueden empezar en personas más jóvenes. Las personas que padecen esta afección tienen pérdida de la fuerza muscular y la coordinación que con el tiempo empeora y les hace imposible la realización de actividades rutinarias.

#### *Artritis*

Es una enfermedad crónica que origina dolor, rigidez, hinchazón, pérdida de movilidad y deformaciones en las articulaciones. Muchas personas con artritis son capaces de, con dificultad, utilizar un teclado y un mouse.

#### *Enfermedad de Parkinson*

Es un desorden del sistema nervioso central que produce temblores involuntarios y rigidez en los músculos. Las personas con un estado avanzado poseen severos problemas a la hora de intentar utilizar el teclado o el mouse. En ciertas circunstancias los pacientes que sufren esta enfermedad ven también afectada su voz.

#### *Temblor esencial*

Es un tipo de movimiento tembloroso involuntario. Este tipo de temblor se da con mayor frecuencia en personas mayores a 65 años y afecta por lo general la parte superior del cuerpo tales como las manos, brazos, cabeza y laringe.

## **Ayudas Técnicas**

En la actualidad existe una gran cantidad de dispositivos no convencionales que tienen como objetivo asistir a las personas en el acceso a la computadora [5]. A continuación, se brinda una lista de los más relevantes.

## Dispositivos Braille

El Braille es un sistema que emplea seis u ocho puntos en relieve en diferentes posiciones para representar letras y números que pueden ser leídos con los dedos.

Los dispositivos de líneas Braille dinámicas posibilitan leer líneas de caracteres y emplean un dispositivo mecánico con clavijas que suben y bajan automáticamente para permitir la presentación de caracteres braille. Además de estas líneas, existen impresoras Braille que producen documentos impresos y teclados Braille para facilitar la entrada de datos.



Figura 2.2: Línea Braille

## Magnificadores de pantalla

La magnificación de pantalla es un software empleado principalmente por personas con baja visión, la misma amplía una zona de la pantalla para facilitar su visualización.

## Lectores de pantalla

Se trata de un software empleado por las personas ciegas o con dislexia, que interpreta el texto en pantalla y lo dirige a un dispositivo de síntesis de voz para ser reproducido en forma sonora o representado en Braille por una Línea Braille.

## Navegadores de texto

Los navegadores de texto, por ejemplo Lynx, son una alternativa a los navegadores con interfaz gráfica. Pueden ser utilizados con lectores de pantalla por las personas ciegas.

## Navegadores de voz

Los navegadores de voz son sistemas que permiten la navegación guiada por la voz. Algunos permiten tanto la entrada como la salida por voz.

## Notificaciones visuales

Una notificación visual es una característica que ofrecen algunos sistemas operativos al permitir que los usuarios con problemas de audición, puedan recibir una alerta visual acerca de algún mensaje de error o advertencia, que generalmente es notificada mediante un sonido.

## Varita en la boca

Debido a la simplicidad y a su bajo costo, la varita dirigida por la boca es una de las asistencias más populares. Por lo general se presenta como una varilla de plástico con una pieza de goma en un extremo y un arco en el otro. Una persona que está incapacitada de usar el mouse y el teclado, puede colocar esta varilla en la boca y dependiendo del grado

de control y paciencia que posea puede controlar un teclado convencional, un mouse o una superficie táctil.



Figura 2.3: Varita en la boca

### **Varita en la cabeza**

De forma similar a la varita dirigida por la boca, una varilla con distinto formato y extensión puede ser adherida a un sombrero o casco, el cual es utilizado por el paciente a la hora de querer interactuar con la computadora.



Figura 2.4: Varita en la cabeza

### **Conmutadores sencillos**

Las personas con una movilidad muy limitada son las que utilizan este tipo de dispositivo. Pueden ser accionados por cualquier parte del cuerpo sobre el que se tenga control muscular. Así, por ejemplo, si una persona puede mover únicamente la cabeza, se podría poner uno de estos interruptores en el lado de la cabeza que permitiera ser pulsado con los movimientos de la misma. Interpretando las pulsaciones por un software específico, es posible controlar el sistema operativo, sitios web u otros entornos.



Figura 2.5: Conmutadores sencillos

### **Conmutadores de aspiración y soplado**

Con una funcionalidad muy parecida a los conmutadores sencillos, estos interruptores de aspiración-soplado son capaces de interpretar la respiración del usuario y traducirlas a señales que tienen dos estados.



Figura 2.6: Conmutadores de aspiración y soplado

### **Mouse Trackball de gran tamaño**

Este tipo de mouse no necesariamente debería ser una asistencia técnica, algunas personas sin discapacidad simplemente los prefieren sobre el mouse convencional. Regularmente este tipo de mouse es más fácil de operar para una persona con discapacidades. Por ejemplo, se podría utilizar en conjunción con un conmutador o una varita de boca. Algunas personas con temblores en las manos pueden encontrar este mouse más conveniente debido a que existen menos riesgos de mover el cursor al hacer un click. Las personas que se ven impedidas de utilizar sus manos podrían llegar a ser capaces de utilizar este dispositivo con los pies.



Figura 2.7: Mouse Trackball

### Teclado adaptado

Este tipo de teclado es muy útil para aquellos casos donde la persona no posee un control muscular confiable o preciso. Algunas teclados adaptados presentan áreas elevadas entre las teclas con el fin de que la persona pueda apoyar la mano entera sobre el teclado sin haber riesgo de oprimir una tecla accidentalmente. Un persona con temblores o movimientos repentinos puede beneficiarse de este tipo de teclados. A menudo los teclados adaptados poseen teclas especializadas vinculadas con software que facilita la entrada de datos. Por ejemplo, hay programas que permiten completar palabras, de esta forma el usuario requiere oprimir menos teclas a la hora de ingresar una palabra.



Figura 2.8: Teclado adaptado

### Seguimiento Ocular

Este dispositivo sigue el movimiento de los ojos y lo coordina con el cursor en la pantalla. Puede ser una muy buena alternativa para aquellos que se ven incapaces de controlar otros dispositivos. Los dispositivos de seguimientos ocular pueden ser muy costosos y por esta razón no son muy comunes.



Figura 2.9: Seguimiento ocular

### **Reconocimiento de voz**

En contraposición a la mayoría de las ayudas técnicas mencionadas hasta el momento, los dispositivos de reconocimiento de voz permiten a la persona controlar verbalmente la computadora. Si bien es una solución prometedora para diversos tipos de deficiencia motoras, puede no serlo para aquellas personas que no pueden pronunciar claramente las palabras, particularmente aquellas personas con parálisis cerebral. Las mismas tienen dificultades a la hora de expresarse oralmente debido a que los músculos que controlan la voz responden con lentitud y las palabras a menudo son mal pronunciadas, a pesar de no tener mermadas sus facultades mentales. [5]

# Capítulo 3

## Antecedentes: Interfaces de comandos de voz

Una interfaz de comando de voz propone un medio de interacción oral que permite a las personas comunicarse con los dispositivos de forma natural utilizando exclusivamente la voz.

Al contrario de los otros mecanismos de interacción, donde las personas deben adaptarse al uso de dispositivos artificiales, en la interacción por medio de la voz se invierte los roles, es la computadora quien se ocupa de entender al ser humano.

Al hacer una analogía con el teclado convencional de una computadora, en una interfaz de reconocimiento de voz, una persona accede a un teclado virtual compuesto de miles de teclas de acceso directo, donde cada una de ellas representa una acción a ejecutar.

Las ventajas de interacción a través del uso de la voz producen un incremento significativo en la calidad de comunicación que se da entre el dispositivo y la persona, al mismo tiempo que reduce la cantidad de componentes visuales que deben ser exhibidos para lograr una buena experiencia de usuario.

En la actualidad se encuentran disponibles diversas aplicaciones y dispositivos con los cuales podemos interactuar utilizando la voz. Cada uno de ellos propone una relación diferente entre la computadora y la persona. A continuación, se brinda una clasificación de los tipos de interfaces de voz.

## Tipos de interfaces de voz

### Interfaces de comandos de voz

La interacción a través de comandos de voz es la forma más simple de reconocimiento de voz. Desde su origen permitió la inserción de texto expresado oralmente [6].

Técnicamente las interacciones basadas en comandos son las acciones más simples de procesar debido a que se pueden comparar con un conjunto pequeño de comandos disponibles.

Los comandos de voz son siempre construcciones sintácticas simples que denotan una acción específica. Un ejemplo de comando de voz es *“reproducir música”*. Un uso más avanzado de comandos puede ser procesado si el usuario indica más detalles de la forma esperada por el sistema. Por ejemplo *“reproducir séptima sinfonía de beethoven”*.

Las interfaces de comandos de voz fueron muy comunes en las versiones iniciales de los sistemas que adoptaron reconocimiento de voz, tales como la primera generación del control del iPhone de Apple y Ford Sync de Microsoft.

En la actualidad se provee una forma más avanzada de interacción, donde el lenguaje natural expresado de forma libre permite una mejor comunicación con los dispositivos. A pesar de esto, nos encontramos con situaciones donde simples y estructurados comandos son más apropiados y valerosos. Escenarios donde se requiere un alto nivel de precisión y el acceso físico es limitado o incluso no existe. Estas circunstancias requieren por parte de los sistemas comandos directos y concisos.

## **Interfaces conversacionales**

Habitualmente una interfaz conversacional es parte de un agente capaz de procesar de forma inteligente una comunicación oral expresada en forma de lenguaje natural. Posteriormente el sistema puede efectuar una acción y responder de forma apropiada a través de una pantalla o verbalmente mediante los parlantes.

Las interfaces conversacionales populares Siri de Apple y Google Now. Tienen como objetivo asistir al usuario en el control de las aplicaciones de los dispositivos móviles. Por ejemplo al interactuar con el asistente de Apple se puede expresar: “*Siri, ¿voy a necesitar hoy un paraguas?*”, el sistema es capaz de entender la pregunta y, luego de consultar el estado del tiempo, responder correctamente.

A pesar de que este tipo de interfaces están en sus inicios, es notable la rápida adopción en los últimos años por parte de los llamados “dispositivos conectados”<sup>1</sup> tales como autos, televisores, relojes, heladeras, cafeteras, etc.

## **Sistemas de Reconocimiento de Voz**

El campo de la Ciencia de la Computación que se ocupa del diseño de sistemas que reconocen palabras expresadas oralmente es llamado Reconocimiento de Voz. Es importante mencionar que el reconocimiento implica la identificación literal de las palabras presentes en el sonido analizado y no el entendimiento acerca del significado de ellas. [7]

En este apartado se describen las características que presentan algunos de los sistemas de reconocimiento más populares que se encuentran en el mercado junto con las dificultades de reconocimiento de voz que es necesario superar al momento de utilizarlos. Queda fuera del alcance de este trabajo la descripción de los fundamentos que hacen posible el reconocimiento de voz.

En la actualidad un gran número de sistemas de reconocimiento están disponibles en el mercado, cada uno de ellos presenta distintas características y muestra un diferente grado de sofisticación. Por ejemplo, los más potentes pueden reconocer miles de palabras de

---

<sup>1</sup> Dispositivos Conectados: Son dispositivos inteligentes electrónicos, por lo general conectados a otros dispositivos o redes a través de diferentes protocolos como Bluetooth, NFC, Wi-Fi, 3G, X10, etc, que pueden funcionar hasta cierto punto de forma interactiva y autónoma.

decenas de lenguajes. Algunos de ellos, previo a poder ser utilizados, requieren de sesiones de entrenamiento por parte del usuario. Otros necesitan que se les hable de forma lenta y pausada.

Debido a las limitaciones tecnológicas y a altos costos, los sistemas de reconocimiento de voz tradicionales han sido utilizados solo en situaciones especiales. Estos sistemas requieren dar soporte a un usuario que por diversas razones no puede manipular un teclado o un mouse. En estos casos la interacción manual es reemplazada por la incorporación de un sistema reconocimiento de comandos de voz. A medida que los costos se reducen y la precisión aumenta, los sistemas de reconocimiento son progresivamente incorporados en nuevos productos tecnológicos con el fin brindar una interacción oral y no solo manual.

### **Sistemas de reconocimiento dependientes e independientes del usuario**

Un sistema de reconocimiento, para ser efectivo, contempla una gran cantidad de atributos que presenta la voz del usuario. Entre estos atributos encontramos:

- Idioma
- Dialecto
- Velocidad del habla
- Tono de la voz
- Volumen de la voz

Podemos clasificar los sistemas en dos tipos, dependientes e independientes del usuario. Los primeros requieren de sesiones de entrenamiento antes de poder procesar la voz del orador. En estas sesiones es en donde el sistema de reconocimiento detecta las características de la voz del usuario y a partir del conocimiento de ellas podrá brindar resultados precisos. En caso de desear utilizar el sistema para reconocer la voz de otro usuario será necesario iniciar el entrenamiento nuevamente con este nuevo orador.

Ejemplos de sistemas dependientes son:

- XVoice
- Sphinx
- Dragon Naturally Speaking
- Dragon Dictate

Por el contrario, luego de indicar el idioma y el dialecto, los sistemas independientes del usuario no requieren de otro tipo de configuración inicial que haga referencia a los atributos de la voz. Es probable que con el uso, algunos de los sistemas internamente recolecten información acerca de los atributos de la voz del orador con el fin de ser capaz de brindar resultados progresivamente más precisos. En caso de desear reconocer la voz de nuevos oradores, no es necesario brindar una configuración inicial.

Ejemplos de sistemas independientes son:

- Loquendo

- NICO Toolkit
- Web Speech API

## **Sistemas de reconocimiento de palabras o palabras conectadas**

Ciertos sistemas de reconocimiento requieren que el orador exprese palabra por palabra de forma aislada. Indicar una pausa entre cada palabra es un requisito fundamental para que la misma sea identificada correctamente.

Dragon Dictate es un ejemplo de sistema de reconocimiento de palabras.

De forma opuesta encontramos sistemas de reconocimiento que son capaces de procesar el habla de forma continua expresada de manera natural. Marcar las pausas entre cada palabra deja de ser necesario. A pesar de esta situación, pronunciar claramente las palabras siempre es una requisito fundamental para obtener resultados precisos.

Ejemplos de sistemas de reconocimiento de palabras conectadas:

- Loquendo
- XVoice
- NICO Toolkit
- Sphinx
- Dragon Naturally Speaking
- Web Speech API

## **Sistemas de reconocimiento de propósito general o específico**

Los reconocedores de voz pueden ser clasificados de acuerdo al fin al que están destinados, pudiendo ser de propósito general, cuando se reconocen palabras de cualquier dominio [8]. Todos los reconocedores mencionados hasta el momento corresponden a esta clasificación.

De forma opuesta, se encuentran los sistemas de propósito específico, estos además de reconocer palabras propias del vocabulario general, reconocen palabras de un dominio en particular. Por lo general los reconocedores de propósito específico son productos comerciales, en esta categoría encontramos por ejemplo los sistemas provistos por Nuance, Dragon Professional y Dragon Legal.

## **Dificultades en el proceso de reconocimiento de voz**

El problema de reconocer automáticamente el habla de una persona es un problema complejo [9]. A continuación, se describen las problemáticas ambientales que los sistemas de reconocimiento deben sortear para poder brindar resultados precisos.

### **Ruido**

El habla siempre se da en un entorno lleno de sonidos, el tick tack de un reloj, el zumbido de una computadora, una radio sonando al otro lado de la habitación, personas hablando a cierta distancia. Estos sonidos son regularmente denominados “ruido”, y es información no

deseada a la hora de reconocer el habla. Los sistemas de reconocimiento deben identificar y filtrar el ruido de la señal a la hora de realizar el procesamiento. Otro tipo de ruido es el efecto del eco, producido por la voz del orador, el cual es percibido por el micrófono algunos milisegundos más tarde.

## **Voz continua**

Un sistema de reconocimiento del habla continua recibe una señal sonora compuesta de palabras de forma sostenida, las cuales a menudo, al ser expresadas de forma natural, carecen de pausas entre ellas. El sistema debe hacer un esfuerzo extra para descomponer en palabras el sonido continuo.

## **Diversidad de oradores**

Los usuarios poseen una voz especial, debido a su única fisionomía y personalidad. La voz no es solamente distinta entre diferentes oradores, también lo es respecto al mismo orador a lo largo del tiempo. Incluso las palabras no son exactamente pronunciadas si una misma persona intenta pronunciarlas una y otra vez esforzándose por que suenen exactamente igual.

Todas las personas hablan de manera diferente, es una de las formas que poseen para expresar su propia personalidad. No solamente utilizan su vocabulario personal, también pronuncian y enfatizan de una forma particular. El estilo del habla varía según la situación, las personas no hablan de la misma forma en un banco que con sus padres o con sus amigos. Es normal que las personas expresen sus emociones verbalmente.

Así mismo, los hombres, mujeres y niños poseen significativas diferencias en el tono, timbre e intensidad de la voz. Estas diferencias son notables aún entre personas del mismo género y edad debido a las características físicas propias de cada individuo.

Habitualmente las personas hablan a distintas velocidades, dependiendo de una gran cantidad de factores. Las personas estresadas tienden a hablar menos pausado que aquellas que están tranquilas o cansadas. Por lo general cuando se habla de algo desconocido el discurso se comunica más lentamente que cuando uno habla de un tema habitual o coloquial.

Los dialectos son variaciones dentro de un lenguaje, estos derivan de distintas regiones o grupos sociales y conllevan cambios de vocabulario, gramática e incluso en la pronunciación de las palabras.

## **Cantidad de información a procesar**

Los sistemas de reconocimiento deben ser capaces de analizar un flujo de datos constante de información y brindar una respuesta en tiempo real, esto demanda una capacidad de procesamiento y velocidad de comunicación superlativa.

## **Ambigüedad del lenguaje**

Al analizar la pronunciación de las palabras es normal encontrar un gran número de ambigüedades llamadas homófonos, palabras que tienen un distinto significado pero que se pronuncian de la misma forma. Por ejemplo, las palabras “casa” y “caza” no significan lo mismo, pero oradores de diversos dialectos las pronuncian igual.

Todas estas circunstancias dificultan el proceso de reconocimiento de voz, y atentan directamente contra el objetivo de brindar resultados precisos.

## **Antecedentes de interfaces de voz**

Hasta el momento, la idea de mantener una conversación con una computadora aparenta ser un tema de ciencia ficción. Pero esta circunstancia está cambiando muy rápidamente.

Un gran número de personas le hablan, en este momento a su dispositivo móvil. Le solicitan, por ejemplo, enviar un email, mensajes de textos, crear notificaciones en el calendario, buscar personas en la lista de contactos, detener una alarma o buscar información en internet [10].

## **Dispositivos móviles**

Actualmente una de las interfaces de voz más populares es Siri, un asistente personal que es incluido en los teléfonos móviles de Apple. De forma similar, las plataformas móviles Android y Windows Phone provee aplicaciones de voz conversaciones que tienen como objetivo asistir al usuario en el uso genérico del dispositivo y de las aplicaciones instaladas.

Dado que las plataformas de los dispositivos móviles son abiertas, un gran número de empresas han provisto aplicaciones capaces de interactuar con el usuario de forma verbal. Un ejemplo muy popular en la actualidad es la capacidad de dictar una nota a una aplicación y que esta se transcriba automáticamente.

## **Industria automotriz**

Una revolución similar está sucediendo en la industria automotriz. Durante mucho tiempo se ha intentado asistir al conductor de cierta forma que el mismo pierda la menor atención posible respecto a la conducción del automóvil, para esto es vital que no quite las manos del volante y la vista sobre el camino. Esta particular circunstancia genera un contexto perfecto para el uso de las interfaces de voz como solución al problema de interacción. En la actualidad se comercializan vehículos que interactúan con el conductor oralmente, respondiendo preguntas respecto al estado del automóvil, posibles rutas a tomar dependiendo del destino mencionado, hasta incluso el control del reproductor multimedia [11].

## **Computadoras de escritorio**

Una gran diversidad de interfaces de reconocimiento de voz han sido desarrolladas para los sistemas operativos Windows, OSX y Linux [12] [13]. De forma similar a los asistentes de voz que vienen incluidos en los dispositivos móviles, los sistemas operativos de

computadoras de escritorio atienden a los comandos de voz efectuados por el usuario. Abrir la papelera de reciclaje, minimizar una aplicación, e iniciar el modo de dictado, son algunas de las operaciones regulares soportadas. Al igual que los asistentes de los dispositivos móviles, el modo de reconocimiento de voz de los sistemas operativos permiten al usuario interactuar solo con las aplicaciones soportadas.

Complementariamente a las herramientas de accesibilidad, integradas con sistemas de reconocimiento de voz, destinadas para sistemas operativos de escritorio, se encuentran productos comerciales como Natural Dragon Speaking, los cuales una vez instalados se ejecuta en segundo plano y permiten al usuario interactuar con la aplicación abierta en primer plano utilizando exclusivamente comandos de voz [14]. Simular clicks y entradas de teclado son algunas de las acciones básicas que estos programas son capaces de realizar bajo petición del usuario. De forma complementaria brindan soporte a operaciones más sofisticadas como gestión de correo, reproducción multimedia y control del editor de textos.

### **Navegadores Web**

Luego del año 2000 las interfaces de comandos de voz llegaron a los navegadores web. En las ediciones de Windows como Windows Vista, el usuario tenía la posibilidad de ejecutar comandos de voz para operar algunas acciones básicas de Internet Explorer, tales como abrir una pestaña, hacer click en un link o abrir un sitio web. De forma similar Opera hasta su versión 12 si era ejecutado en Windows XP o 2000, permitía interactuar oralmente con el navegador a través del uso de comandos de voz [15].

En la actualidad Natural Dragon Speaking da soporte para interactuar con los navegadores web más populares permitiendo realizar un dictado de texto o acceder a las opciones de menú del programa.

Luego del lanzamiento y posterior soporte por parte de algunos de los navegadores más importantes, la api de reconocimiento de voz definida en el estándar de HTML 5 ha hecho posible el desarrollo de múltiples aplicaciones web que proponen una interfaz de voz. Un ejemplo muy bien logrado de ello es la búsqueda a través de voz que soporta el buscador Google.



Figura 3.1: Búsqueda por voz en Google

Otro ejemplo más complejo y totalmente funcional es la API de comandos de voz para la edición de documentos de Google Docs y Google Slides [16]. Además de la posibilidad de dictar un texto, comandos tales como “Copiar”, “Pegar”, “Insertar tabla”, “Alinear a la izquierda” y muchos más permiten al usuario construir un documento de la forma deseada exclusivamente utilizando la voz.

# Capítulo 4

## Propuesta: Interacción web a través de comandos de voz

Debido al predominio de interacción manual requerido a la hora interactuar tanto con el navegador como con los sitios web, es de interés formular una alternativa de interacción sin resignar la funcionalidad soportada.

A lo largo de este capítulo, se presenta un modelo cuyo objetivo es brindar una alternativa de navegación web prescindiendo totalmente de la interacción manual mediante el uso exclusivo de comandos de voz.

A continuación, se describen los componentes de una interfaz de comandos de voz capaz de atender a las instrucciones verbales provistas por el usuario, para luego poder hacerlas efectivas sobre el navegador de forma equivalente a si estas hubiesen sido desempeñadas manualmente.

### Comandos de Voz

Como se describió en el capítulo anterior, en una interfaz de voz no conversacional, un comando de voz es una construcción sintáctica asociada a una acción que puede ser ejecutada en el sistema.

Con la intención de hacer referencia a ciertos tipos de comandos en los apartados posteriores de este trabajo, a continuación, se clasifica, distingue, describe y ejemplifica cuatro tipos de comandos de voz.

### Comandos de voz atómicos y parciales

Se definen los **comandos de voz atómicos** como aquellos que indican una acción concreta e indivisible. Por ejemplo, un comando de voz atómico soportado en un sistema de navegación web sería “Recargar página” o “Cerrar pestaña”. Ambos comandos son concretos y al ser atendidos no requieren de subsiguientes comandos para invocar unívocamente la acción a realizar.

Por el contrario, un **comando de voz parcial** es aquel que, junto con otros comandos parciales, refieren a una acción concreta. Por ejemplo, en un editor de texto, un comando de voz que podría ser de interés implementar, es la opción de copiar un fragmento de un texto determinado. Para ello probablemente sea necesario invocar los comandos “*seleccionar texto*”, “*copiar {referencia}*” donde {referencia} es el indicador que posee el texto de interés.

En este ejemplo la acción era clara, concreta e indivisible, y consiste en copiar ciertas líneas de texto, sin embargo fue necesario ejecutar dos comandos para poder hacer referencia unívoca a la acción requerida sobre el elemento deseado.

## Comandos de voz genéricos y específicos

Un **comando de voz genérico** es definido como aquel que efectúa una acción sobre un aspecto común en el sistema, el cual puede derivar en que el mismo sea ejecutado múltiples veces sobre distintas componentes con el fin de alcanzar distintos objetivos. Un ejemplo simple de un comando genérico en un sistema de navegación web sería “*click enlace*”. Dado que un enlace (o link) es un componente habitual en los sitios web, un usuario podría invocar sucesivas veces este comando con la intención de navegar desde un sitio web A hasta uno B sin importar la particularidades específicas de contenidos o de implementación que estos sitios contienen.

Por el contrario, un **comando de voz específico** sería aquel que refiere unívocamente a un aspecto del sistema. Un ejemplo de estos sería “*enviar email*”, el mismo presupone la existencia de un email junto con la capacidad de poder enviarlo. Es necesario notar que esta acción es de carácter específico de un sistema particular y muy probablemente pueda ser ejecutado solamente en un momento preciso.

Una particularidad de los comandos de voz específicos es que por lo general son equivalentes a una sucesión de comandos genéricos. Por ejemplo haciendo uso del comando “*click botón*” y otros de manejo de interfaces visuales como “*desplazar hacia abajo*” podríamos localizar el botón que efectúa el envío del email y presionarlo.

Finalmente es importante destacar que un comando de voz siempre corresponde a las dos categorías en simultáneo, por ejemplo:

- “*enviar email*”: atómico y específico
- “*ayuda*”: atómico y genérico
- “*seleccionar texto*” y “*texto {referencia}*”: parcial y genérico
- “*agregar email del destinatario*” “*{email\_destinatario}*”: parcial y específico

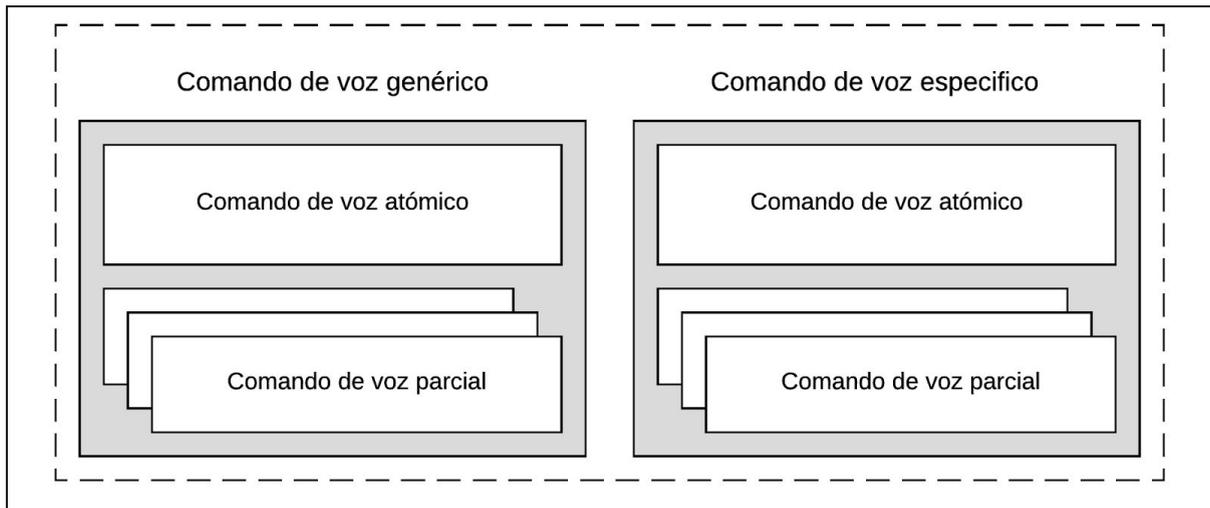


Figura 4.1: Clasificación de comandos de voz

## Contextos de ejecución de comandos

En ciertas ocasiones resulta necesario determinar el contexto del usuario antes de ejecutar el comando, de lo contrario se puede incurrir en realizar una acción no deseada. Un ejemplo claro de esta situación es la del comando “salir” implementado por un sistema de navegación web. Antes de proceder a abortar la ejecución de algún componente del sistema es importante detectar el contexto del usuario, debido a que resulta radicalmente distinta la consecuencia de la acción ejecutada si el usuario se encuentra en algunas de las siguientes situaciones:

1. viendo un video en pantalla completa
2. tiene abiertas múltiples pestañas
3. tiene abierta solo una pestaña

Depende del contexto será la acción que este comando genérico debe ejecutar.

Como complemento a los contextos de ejecución, es de interés mencionar que un usuario potencialmente debería poder estar vinculado a múltiples contextos de ejecución en simultáneo, donde cada uno de estos posea cierto orden de precedencia según la acción a realizar. Contar con una correcta jerarquización de los contextos ayudará a poder determinar cuál comando de voz ejecutar en caso de que el mismo es soportado por múltiples contextos y todos ellos son congruentes al estado del usuario.

Volviendo al ejemplo enunciado anteriormente donde el usuario se encuentra visualizando un video en pantalla completa, los siguientes tres hipotéticos contextos estaría activos:

1. Contexto de reproducción de video
2. Contexto de navegación web
3. Contexto de administración del navegador

Si durante la reproducción del video el usuario invoca el comando “*Recargar*”, con total seguridad debería ser invocado atendiendo al “Contexto de navegación web” y por ende volver a cargar el sitio web actual. Distinta es la situación del comando “*salir*” la cual debería ser resuelta por el “Contexto de administración del navegador” y por lo tanto cerrar la pestaña. Esta resolución se desarrollará si y sólo si el contexto de reproducción de video no implementa este mismo comando. Puede ser requerimiento del sistema que al invocar “*salir*”, se detenga la reproducción y se cierre el panel donde se estaba reproduciendo el video. Queda a consideración del lector, preguntarse qué debería suceder si el usuario menciona “ayuda” y cómo puede este comando producir un conflicto.

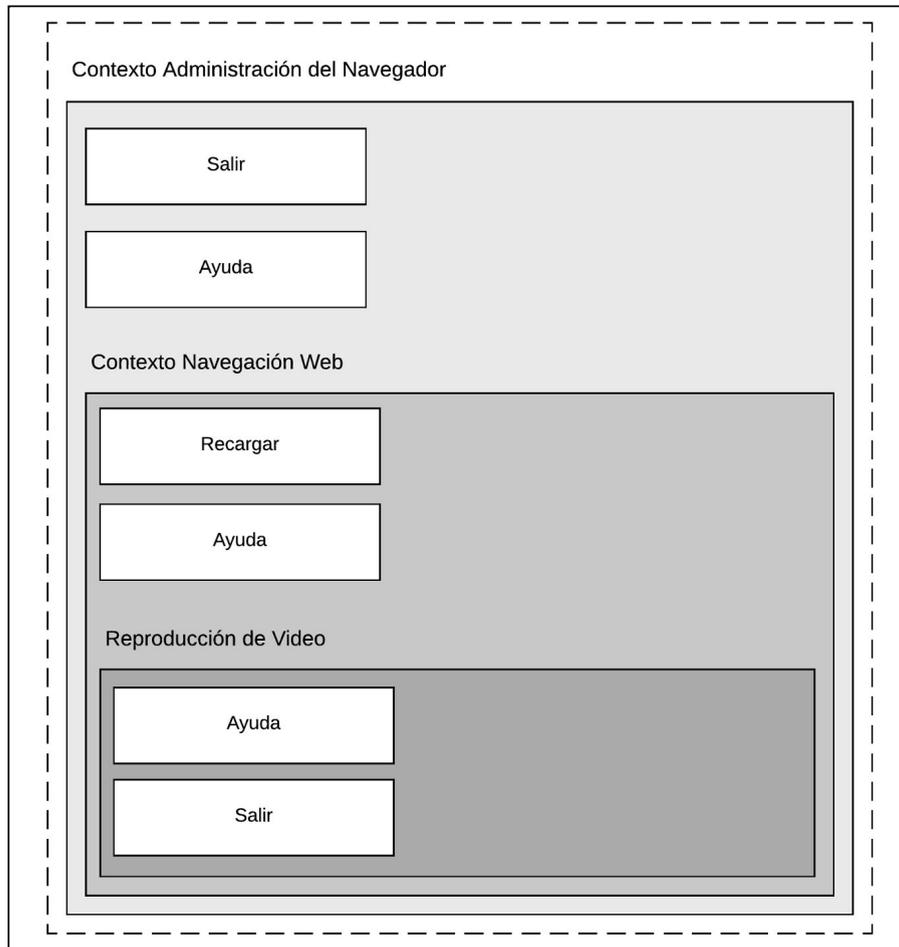


Figura 4.2: Jerarquías de contextos

Junto con la noción de multiplicidad de contextos de ejecución y jerarquías entre los mismos, emerge la característica de incompatibilidad entre contextos. Un ejemplo de ello es la imposibilidad de un usuario de encontrarse vinculado simultáneamente a los siguientes contextos.

- Contexto de reproducción de video en pantalla completa
- Contexto de inserción de texto

De producirse esta situación, cuando el usuario invoca el comando “*detener*” el resultado puede ser que la palabra mencionada se escriba en algún campo de texto en vez de interrumpir la reproducción del video.

Con el fin de presentar un modelo de interacción web a través de comandos de voz seguro y confiable, es necesario definir los contextos para los cuales aplica cada comando de voz soportado por el modelo.

## Modelo de ejecución de comandos voz

El siguiente modelo de ejecución de comandos de voz tiene como objetivo brindarle al usuario, la posibilidad de navegar todos los aspectos de los sitios web modernos sin limitaciones en cuanto a la funcionalidad soportada. Con el fin de indicar cómo puede ser esto logrado, inicialmente se presenta la funcionalidad a soportar clasificada en tres grupos y luego se describe los mecanismos por los cuales se soporta la interacción que presenta cada grupo.

- Funcionalidad exclusiva del navegador
- Funcionalidad genérica de los sitios web
- Funcionalidad específica de los sitios web

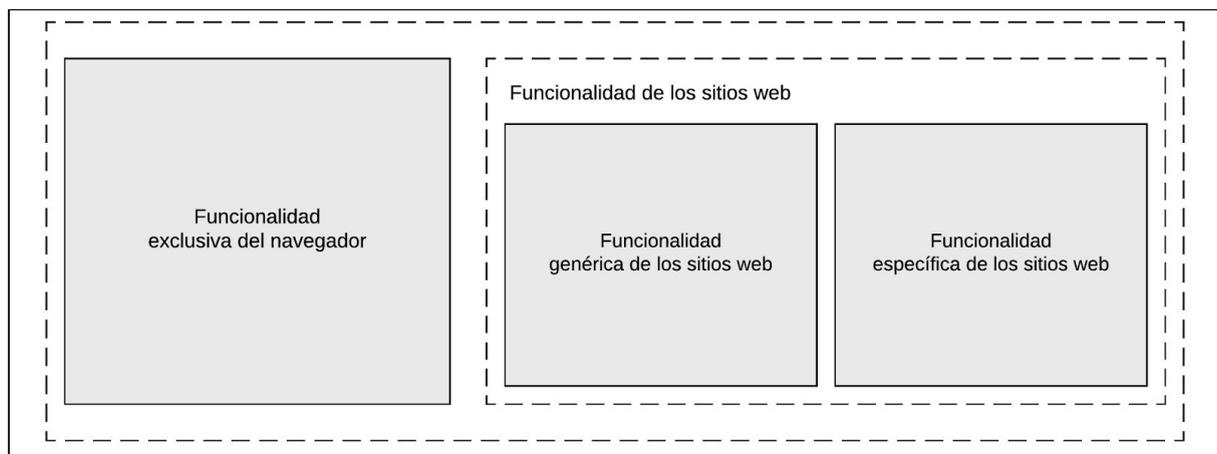


Figura 4:3 Clasificación de la funcionalidad presente en la navegación

### Funcionalidad exclusiva del navegador

En esta clasificación se encuentran todas aquellas acciones que son efectuadas directamente por el navegador sin intervención de lógica de un sitio web. Por ejemplo, gestión de marcadores, administración de pestañas, gestión de descargas, uso de las barras de desplazamiento, gestión del historial y tantas otras que los navegadores actuales soportan.

## **Funcionalidad genérica de los sitios web**

Esta categoría refiere a aquellas acciones que forman parte de la interacción propuesta por componentes web estándares presentes dentro del sitio web visualizado. Por ejemplo, click en un vínculo, seleccionar un texto, escribir en un campo de texto, reproducir un video, copiar una dirección de un enlace, presionar un botón, y todas aquellas que son parte de la interacción soportada por los diferentes componentes web del estándar HTML 5.

## **Funcionalidad específica de los sitios web**

Finalmente en esta categoría se incluyen todas aquellas acciones derivadas de la interacción con un componente web no estándar.

Cuando se observa en detalle las acciones que se pueden realizar en sitios web concretos, encontramos que muchos de ellos poseen funcionalidad hecha a medida. Por ejemplo en la interacción con un mapa, por lo general se pueden hacer paneos en todas las direcciones, acercarse y alejarse, al no haber un componente estándar definido para los mapas, cada sitio web implementa el propio. Las acciones que se pueden realizar sobre este componente se la considera funcionalidad específica, debido a la diversidad de implementaciones de mapas en la web que brindan funcionalidades personalizadas.

Luego de analizar estos tres grupos de funcionalidades se concluye que es inviable brindar un conjunto de comandos de voz que sea capaz de cubrir el vasto universo de funcionalidades soportadas en la web. La existencia de comportamiento personalizado en sitios web atenta directamente contra la iniciativa de proveer un conjunto de comandos que brinde soporte a un conjunto infinito de funciones.

Por las razones presentadas, a continuación se describen mecanismos por los cuales se logra soportar cada subtipo de funcionalidades clasificadas por separado, con el objetivo de lograr, en su conjunto, el soporte total.

## **Soporte a funcionalidades exclusivas del navegador**

Con el objeto de soportar las funciones que un navegador brinda al usuario, es necesario inicialmente analizar, explorar y enumerar todas las acciones que están a disposición del usuario para luego poder atribuirle a cada una de ellas un comando de voz que la represente.

En el Anexo 1 se define por extensión el conjunto general de funcionalidades soportadas por el navegador Google Chrome y a cada una de ellas se le atribuye un comando de voz.

## **Soporte a funcionalidades genéricas de los sitios web**

La estructura de todos los sitios web modernos poseen un importante rasgo en común, han sido desarrollados a través de la composición de componentes web del estándar HTML. Debido a esta característica en común compartida, al dar soporte a través de comandos de

voz a todos los componentes web del estándar, se estaría brindando soporte a todas las funcionalidades genéricas de los sitios web.

En forma similar al método descrito en el punto anterior, se puede analizar, explorar y enumerar todos los componentes web estándares de la versión actual de HTML junto con sus acciones soportadas. Posteriormente es necesario asignarle a cada acción sobre un componente, un comando de voz y un contexto de ejecución en el cual el comando es válido.

En el Anexo 2 se define por extensión el conjunto general de funcionalidades de los componentes HTML 5 soportados por el navegador Google Chrome y a cada acción se le atribuye un comando de voz.

### **Soporte a funcionalidades específicas de los sitios web**

Si bien se presenta como un desafío el analizar, describir y enumerar las funcionalidades específicas de cada sitio web presente en la red, resulta de mayor complejidad el poder definir los comandos de voz para aquellas funcionalidades que van a estar presentes en los sitios web que no se han implementado aún.

Estas razones hacen que sea virtualmente imposible dar soporte al universo de comandos de voz específicos presentes en la web actual o futura.

Sin embargo, como se mencionó anteriormente, los comandos de voz específicos en su gran mayoría son equivalentes a un conjunto de comandos de voz parciales. Por ejemplo podemos citar el comando de voz específico de una red social “*Ir a mi perfil*”, la misma acción puede ser efectuada si el usuario ejecuta los comandos de voz “*click vínculo*” y luego “*{referencia\_al\_vínculo\_del\_perfil}*”. En este caso el usuario no va a sentir la ausencia del comando “*Ir a mi perfil*” debido a que tiene al alcance un conjunto de comandos no específicos que pueden ser invocados con el fin de poder lograr el resultado esperado de una forma alternativa.

Sin embargo, no todos los comandos de voz específicos pueden ser reemplazados por una secuencia de comandos de voz atómicos o parciales, por ejemplo un comando de voz específico de un mapa interactivo puede ser “*mostrar caminos*”, salvo que casualmente el mapa se comporte de forma similar a un componente genérico no habrá un conjunto de comandos de voz genérico que logren efectuar la acción.

Para suplir esta necesidad, herramientas como Dragon Naturally Speaking [14] permiten al usuario operar el cursor con el uso de la voz. Si bien es una solución viable al problema, usualmente es poco práctica.

En este apartado se proponen dos formas complementarias de proveer comandos de voz que den soporte a las funcionalidades específicas de los sitios web.

- Módulos personalizados de comandos de voz
- API de comandos de voz nativos del sitio web

### **Módulos personalizados de comandos de voz**

La incorporación de módulos personalizados emerge como un mecanismo mediante por el cual un usuario puede dar soporte de forma personalizada a un conjunto de funcionalidad deseada.

Cada módulo incluido debería detallar por cada comando de voz la evaluación del contexto de ejecución y el código a ejecutar como consecuencia de la invocación del comando.

Volviendo al ejemplo del mapa interactivo, puede ser deseable para cierta parte de la comunidad, la implementación de un módulo que describa los comandos de voz necesarios para la correcta interacción con el mapa, junto con la evaluación del contexto que indique si el usuario está en posición de interactuar con el componente.

De esta forma la comunidad o el conjunto de usuarios de forma particular podría extender el modelo hasta alcanzar el nivel de soporte deseado.

### **API de comandos de voz nativos del sitio web**

Hasta el momento los sitios web presentan, en su gran mayoría, representaciones exclusivamente visuales y multimediales de la información, al mismo tiempo que solo permiten el acceso e interacción por medios exclusivamente manuales. Esta situación deriva del uso masivo del teclado y el mouse a la hora de interactuar con la PC.

En este apartado se propone un cambio de paradigma de interacción a la hora de navegar la web. Así como cada sitio web expone funcionalidad a través de componentes visuales que requieren interacción manual, estos mismos componentes podrían ser susceptibles a interacción a través de comandos de voz, al mismo tiempo que el sitio web podría exponer comandos de voz globales con la intención de ampliar la funcionalidad soportada a un mayor nivel de abstracción.

Con el objetivo de soportar la funcionalidad específica de un sitio web haciendo uso de comandos de voz, se puede brindar un mecanismo mediante el cual el sitio en forma nativa defina las instrucciones de voz que desea exponer al usuario, de forma equivalente a como se brinda un botón para que se pueda desempeñar una acción en el sistema.

De esta forma, un portal web, luego de detectar que se encuentra ante la presencia de un usuario que interactúa a través de comandos de voz, expondría los comandos soportados al mismo tiempo que podría cambiar la interfaz para brindar una mejor experiencia de usuario.

Tomando de ejemplo un sitio web de mensajería, podría ser de interés soportar los siguientes comandos de voz:

- *Iniciar sesión*
- *Cerrar sesión*
- *Nuevo mensaje*

- *Enviar a [nombre de contactos...]*
- *Guardar mensaje*
- *Eliminar mensaje*
- *Listar mensajes*
- *Abrir mensaje número [referencia]*
- *Nuevo contacto*
- *Guardar contacto*

Para dar soporte a estos comandos resulta necesario que el sitio, por cada instrucción, exponga el código de ejecución de la acciones soportadas junto con la evaluación del contexto para determinar si el comando de voz es de posible ejecución o brindar constantemente una lista actualizada de comandos de voz de posible ejecución.

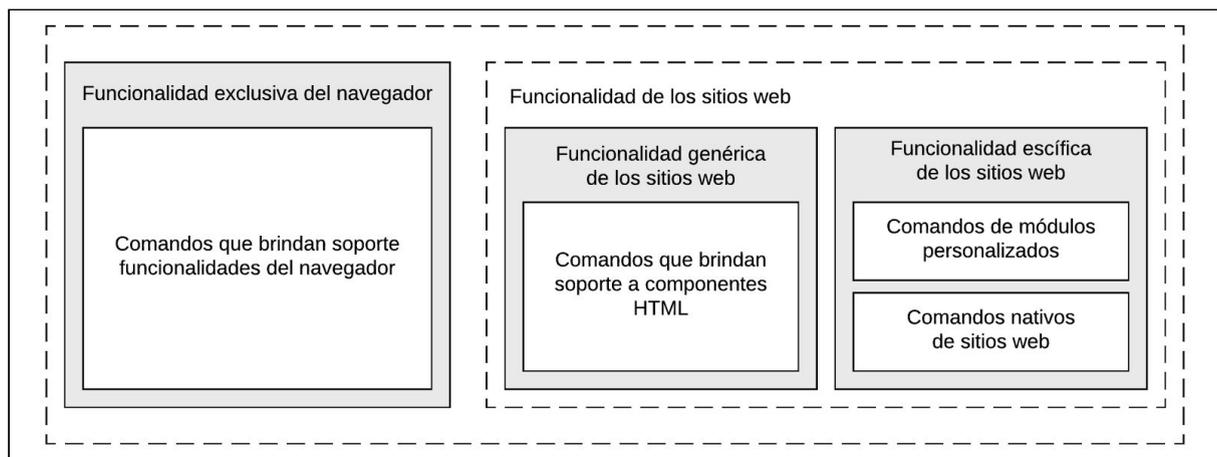


Figura 4.4: Soporte de funcionalidad presente en la navegación web

## Funcionalidades soportadas por la interacción oral

Es interesante destacar que si bien la motivación inicial de este capítulo es la de proveer un modelo de navegación web dirigido exclusivamente por comandos de voz sin pérdida de soporte de funcionalidad, el cambio de paradigma de interacción propuesto define nuevas posibilidades de interacción, en concreto aquellas que son cercanas a las de las interfaces de voz. Por ejemplo, un sitio web podría implementar un asistente con el que sea posible interactuar oralmente de forma tal que nos oriente hacia el objetivo deseado.

El nuevo paradigma de interacción oral desarrollado proveería de un cuarto grupo de funcionalidad soportada correspondiente a la navegación web, en este caso, un grupo que este modelo propuesto por definición ya estaría soportando.

El surgimiento de este nuevo conjunto de funcionalidades soportadas por los sitios web, llevaría a hacerse la siguiente pregunta. ¿Cómo dar soporte alternativo a las acciones orales a través de una interacción manual?

# Capítulo 5

## Desarrollo de la herramienta: Handsfree for Web

La implementación de la herramienta Handsfree for Web surge de la necesidad de proveer un mecanismo por el cual la navegación web sea posible de realizar de una forma alternativa a la interacción manual.

La aplicación desarrollada es una implementación del modelo propuesto en el capítulo anterior, donde se describe como se puede navegar la web a través del uso exclusivo de comandos de voz.

Al analizar los antecedentes de interfaces de voz en la web mencionados en el capítulo tres, se puede concluir que es deseable proveer una herramienta multiplataforma que cuente con mayor soporte de interacción que las que encontramos hoy en los sitios web y de ser posible que no cuente con una licencia comercial paga para que pueda ser utilizado masivamente.

A continuación se describen las diversas características que conforman la herramienta desarrollada junto con la justificación de las elecciones de diseño tomadas.

### Tecnologías utilizadas

#### Acerca de la plataforma

Handsfree for Web es una Extensión de Google Chrome [17] que puede a ser instalada en cualquier versión moderna del navegador Google Chrome en su edición para computadoras de escritorio, disponible para todos los sistemas operativos de uso masivo.

Las extensiones de Google Chrome poseen una gran cantidad y diversidad de ventajas. A continuación se enumeran aquellas que motivaron la elección de dicha plataforma para realizar el desarrollo.

- Son soportadas por un navegador de última generación, gratuito y multiplataforma
- Permiten la extensión del navegador a través del uso de apis
- Brindan acceso a los detalles de navegación de la sesión del usuario
- Permiten el acceso y la modificación de los sitios activos
- Son de fácil distribución e instalación
- Son estables, seguras y confiables

#### Acerca de la api de reconocimiento de voz

Una de las ventajas mencionadas de las extensiones de Google Chrome es el acceso a los sitios web activos en los cuales el usuario está navegando. Este acceso permite el uso de la

Api de Reconocimiento de Voz [18] definida en el estándar HTML 5 y soportado por Google Chrome.

A continuación se describen las ventajas que motivaron la elección del servicio de reconocimiento de voz.

- Uso gratuito y sin límites
- Sistema de reconocimiento independiente del usuario, de propósito general y capaz de procesar frases con palabras conectadas
- Soporte de múltiples idiomas y dialectos
- Procesamiento en tiempo real
- El servicio se encuentra en mejora continua
- Posibilidad de definir un servicio de reconocimiento alternativo al que viene asignado por defecto

### Acerca de la aplicación

La herramienta integra un conjunto de tecnologías destinadas al desarrollo de aplicaciones web de última generación. A continuación, se mencionan las tecnologías utilizadas más relevantes y su campo de acción.

- Javascript ES6: Lenguaje de programación
- CSS 3: Lenguaje de estilos
- HTML 5: Lenguaje de marcas de hipertexto
- React: Gestión de los componentes visuales
- Redux: Administración del estado
- JQuery: Manipulación de los componentes visuales del sitio web activo
- Webpack: Gestión de módulos y building de la aplicación

### Modelo de arquitectura y flujos de datos

El proceso de ejecución de comandos de voz en el navegador web necesita de la interacción activa de diversos componentes, los cuales deben actuar de forma sincronizada.

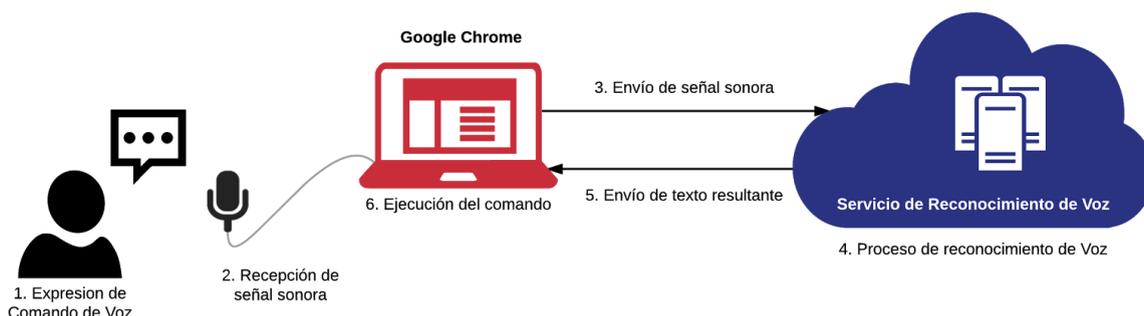


Figura 5.1: Proceso de ejecución de comando de voz

A continuación se detallan los componentes involucrados que forman parte activa en el proceso de reconocimiento de voz y posterior ejecución del comando resultante, junto con la mención al flujo de información comunicado en cada etapa.

1. El usuario expresa verbalmente el comando de voz
2. La aplicación Handsfree for Web instalada en el navegador Google Chrome recibe el inicio de la señal sonora y abre un canal de comunicación con el Servicio de Reconocimiento de Voz
3. El flujo de datos correspondiente a la señal sonora es enviada al servicio de Reconocimiento de Voz
4. A medida que es recibida la señal sonora, el servicio de reconocimiento de voz procesa el flujo de datos y retorna resultados parciales correspondientes a la transcripción del sonido recibido
5. Los resultados del procesamiento de voz son enviados al navegador web
6. La herramienta recibe el texto correspondiente al comando de voz expresado por el usuario y ejecuta el comando indicado

Es importante destacar que si el usuario expresa un comando de voz extenso o múltiples comandos de voz de forma continua, todos los pasos indicados son ejecutados en forma cíclica.

## **Diseño de la herramienta**

Diversas decisiones de diseño fueron tomadas al momento de desarrollar la aplicación; desde cuestiones visuales como el aspecto de la interfaz o la información a presentar, así como la organización referente a los comandos. En esta sección se describen los aspectos más importantes de la herramienta y las razones que fundamentan la existencia de los mismos.

### **Interfaz visual**

Dada la naturaleza oral de interacción de la aplicación, se podría suponer que la misma además de ser reactiva a comandos de voz, podría brindar información al usuario a través de un medio verbal. Si bien las posibilidades técnicas para informar al usuario a través de mensajes de voz están al alcance mediante el uso de síntesis de voz [19], se decidió que toda respuesta por parte de la aplicación sea de naturaleza visual y no oral.

Las razones por las cuales se optó por una interfaz de respuesta visual son las siguientes:

- Respuesta inmediata: al contrario de la expresión oral de un mensaje, donde su tiempo de reproducción está en el orden de los segundos, mostrar un mensaje por pantalla está en el orden de los milisegundos.
- Respuesta continua: al mostrar respuestas visuales por pantalla no invasivas, las mismas pueden ser exhibidas constantemente sin interferir con la atención del usuario, de esta forma es el mismo quien decide cuándo prestar atención a las señales emitidas.

- Información descriptiva: en la mayoría de los casos la información a brindar como resultados de comandos de voz o como disparadores de los mismos es de naturaleza visual o de difícil descripción en forma oral.
- Facilidad de contextualización de la información: en un esquema visual se pueden mostrar convenciones para reafirmar el contexto del mensaje a brindar. Por ejemplo utilizar el color rojo para un error acompañado de un ícono con una cruz.
- Evita la interferencia sonora con los comandos de voz: probablemente la razón más importante por la cual se eligió una interfaz visual de respuesta sobre una verbal es el hecho de que si se emiten respuestas orales, potencialmente el sonido correspondiente a esta respuesta puede interferir con la recepción del comando de voz emitido por el usuario.
- Congruencia de medio: el tipo de usuario que utiliza la herramienta percibe la web de una forma visual, esta capacidad es requerida por la aplicación. El usuario invoca comandos de voz con el fin de interactuar con los componentes web que visualiza. Esta situación es congruente con emitir señales visuales en la pantalla al mismo tiempo que se visualizan los efectos de los comandos invocados. En caso de emitir respuestas orales, la herramienta supondría que la persona es capaz o está disponible para escuchar los mismos, cuestiones de entorno o discapacidad pueden atentar contra ello.

### **Barra de estado**

Dado que la aplicación es reactiva a todo sonido capturado por el micrófono de la computadora, se consideró necesario brindar información en tiempo real acerca del estado de la herramienta.

La presencia de la barra de estado en la pantalla indica inicialmente que la aplicación fue instalada y puede ser utilizada. Este componente tiene como objetivo además indicar los siguientes aspectos que conforman el estado de la actividad en tiempo real.

- Estado de la recepción del sonido: a través de la figura icónica de un micrófono se muestra la disponibilidad a capturar el sonido y realizar el proceso de reconocimiento de voz.
- Texto correspondiente al comando de voz expresado: durante el proceso de reconocimiento de voz, en tiempo real se indica el texto correspondiente a lo expresado por el usuario.
- Contexto actual: como se verá más adelante todos los comandos de voz son ejecutados dentro de un contexto. Inicialmente la aplicación se encuentra en el contexto "Principal" pero dependiendo de la actividad a desarrollar se puede estar en otros contextos, por ejemplo a la hora de interactuar con un reproductor de video el

contexto es “Opciones de multimedia”. En gran medida el contexto determina la disponibilidad de comandos de voz que pueden ser ejecutados.

- Confirmación de la ejecución del comando: una vez recibido el comando de voz expresado por el usuario, se muestra una notificación que describe si el mismo pudo ejecutarse.



Figura 5.2: Barra de estado - Captura de sonido iniciada

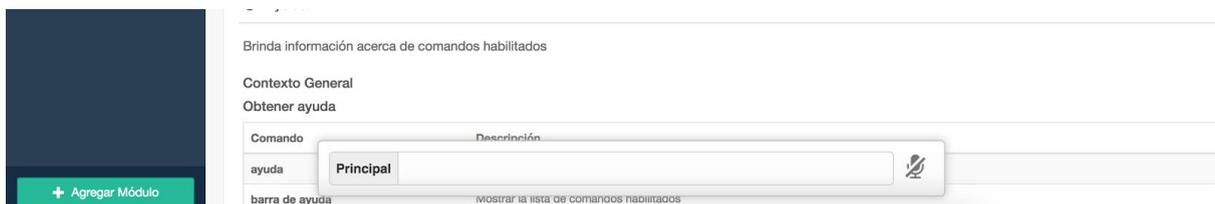


Figura 5.3: Barra de estado - Captura de sonido detenida



Figura 5.4: Barra de estado - Confirmación de ejecución de comando indicado

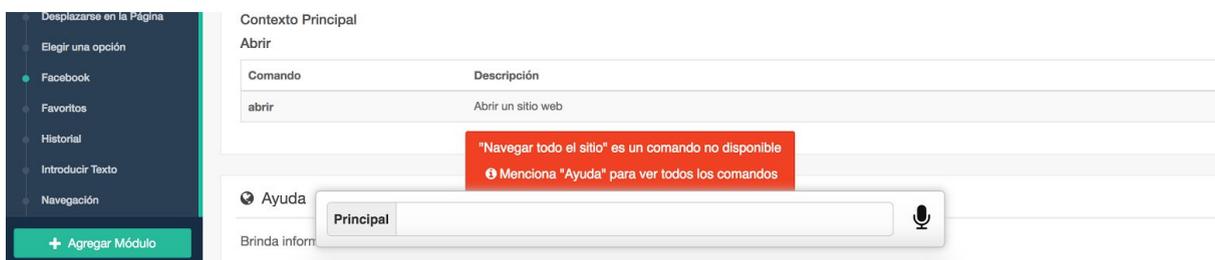


Figura 5.5: Barra de estado - Comando indicado no soportado

## Marcadores

Generalmente los usuarios de la herramienta interactúan con componentes web presentes en el sitio web, para ello es necesario de alguna forma indicar a la persona los comandos de voz disponibles que se puede ejecutar. Por ejemplo, si el usuario desea hacer “click” sobre un vínculo, la herramienta además de recibir el comando de voz “click” expresado por el usuario necesita saber cual es el vínculo en cuestión. Para ello luego de la ejecución del comando de voz “click”, se muestra por pantalla marcadores sobre cada vínculo con una

referencia que puede ser invocada por el usuario para especificar cual es el elemento que debe ser presionado.



Figura 5.6: Marcadores correspondientes a elementos presentes en un sitio web

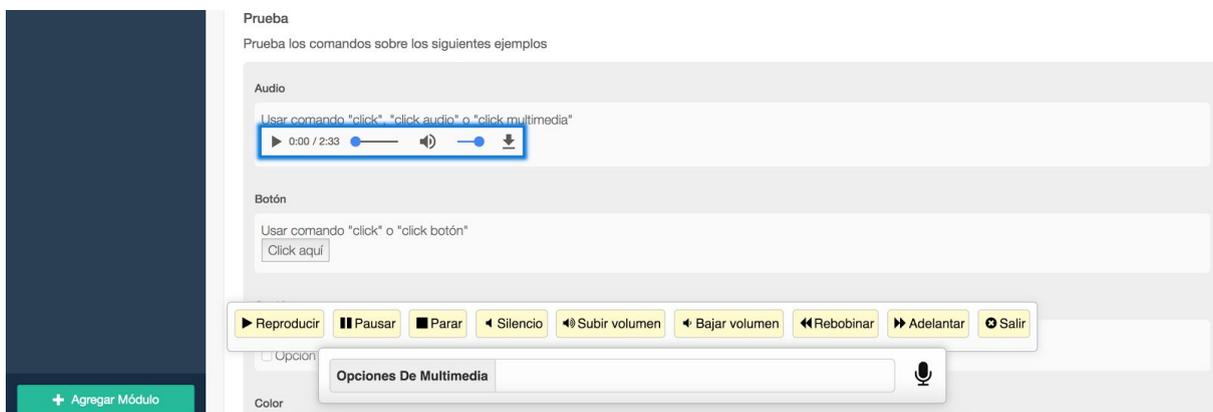


Figura 5.7: Marcadores correspondientes a opciones del reproductor multimedia

## Cuadros de Diálogos

A menudo cuando el usuario interactúa con ciertos aspectos del navegador o de un sitio web, es necesario presentar más información que la que el sitio web actual describe. Esta información accesoria solicitada por el usuario se muestra contenida en un cuadro de diálogo de pantalla completa. Un ejemplo de cuadro de diálogo de pantalla completa es visualizado al requerir ver la lista de favoritos.

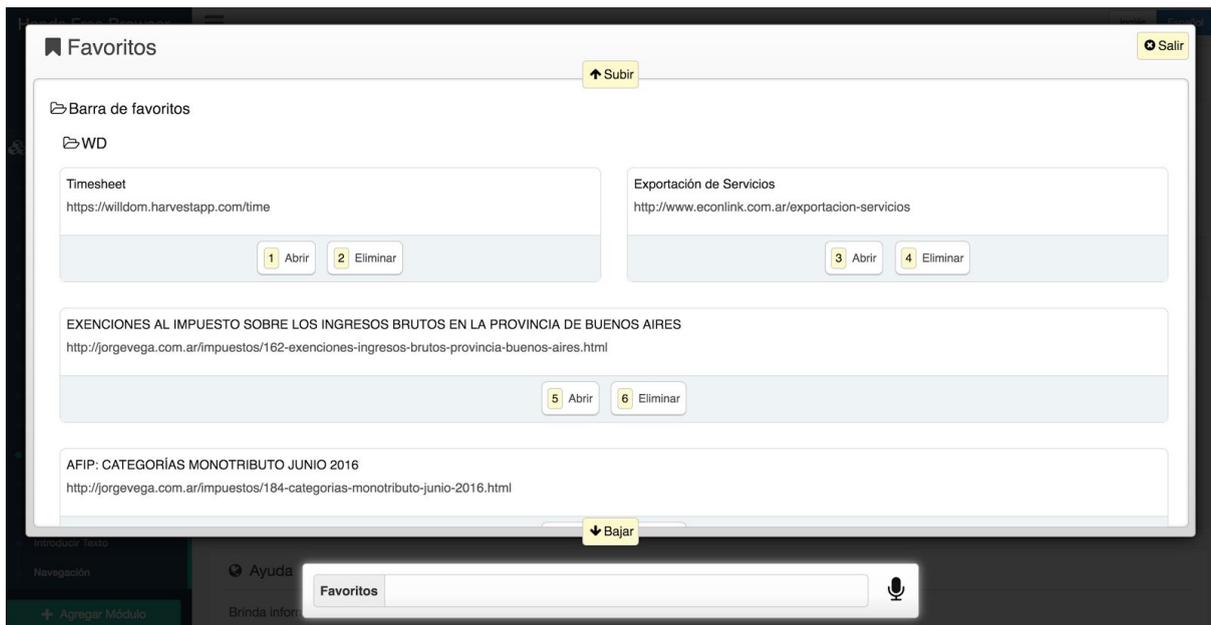


Figura 5.8: Cuadro de diálogo correspondiente a la administración de favoritos

Complementariamente al cuadro de diálogo de pantalla completa, se provee un cuadro de diálogo de tamaño reducido sobre la barra de estado.

De forma habitual se necesita sugerir comandos de voz para asistir al usuario en la realización de una actividad que se lleva a cabo dentro del sitio web. Para ello se presenta un elemento visual que no cubra toda el cuerpo del sitio web.

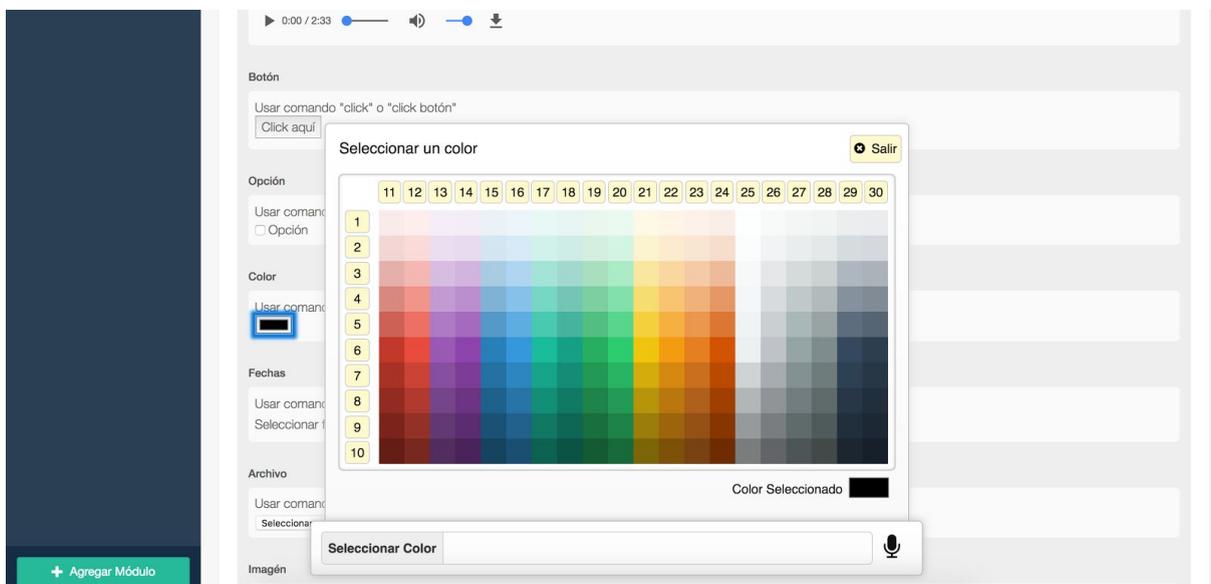


Figura 5.9: Cuadro de diálogo correspondiente a la selección de un color

### Identificación de elemento seleccionado

En ciertas ocasiones cuando un usuario interactúa con un componente web es necesario denotar visualmente cual es el elemento sobre el que los futuros comandos van a ejecutar las acciones.



Figura 5.10: Selección de un vínculo

## Módulos de comandos de voz

Resulta de interés organizar los comandos de voz soportados por la aplicación a través de una estructura modular, la cual permita la carga y descarga de módulos de comandos sin afectar la estabilidad de la aplicación, al mismo tiempo que permita la incorporación de nuevos módulos de comandos de una forma simple.

Si bien existen múltiples ventajas al arquitecturar una aplicación de forma modular, resulta de particular interés en esta sección el hecho de que todos los módulos deben respetar la misma interfaz para poder ser compatibles con el sistema que los incorpora.

A continuación las componentes de un módulo de comandos de voz.

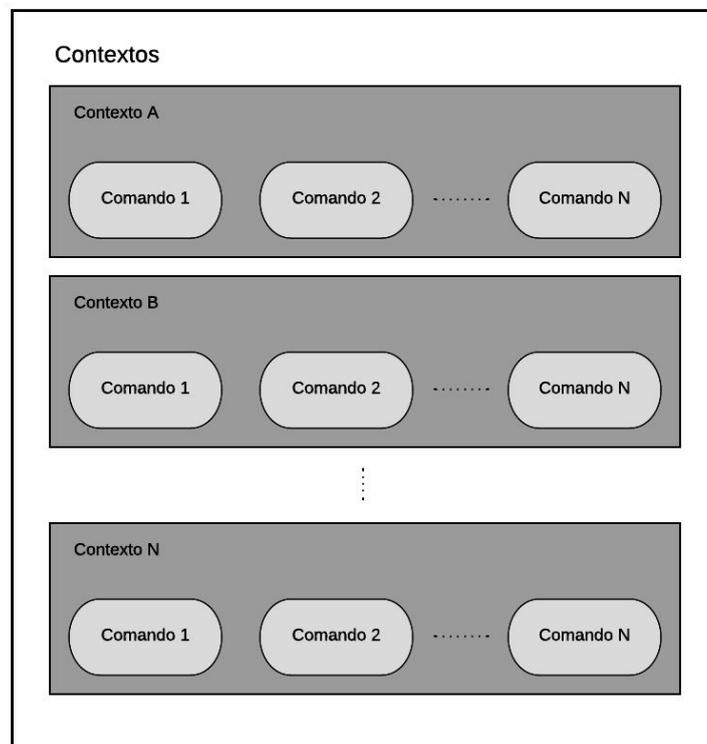


Figura 5.11: Estructura de un módulo de comandos de voz

## **Contextos**

Como se describió en el capítulo anterior un contexto brinda información acerca del entorno de ejecución en el cual se van a invocar los comandos. En particular para la herramienta desarrollada los contextos poseen las siguientes características:

- Identifica un conjunto de comandos de voz que pueden ser ejecutados
- Posee un identificador único
- La herramienta siempre tiene cargado solo un contexto a la vez, la ejecución de un comando de voz puede resultar en un cambio de contexto.

## **Comandos**

Cada contexto define un conjunto de comandos, los cuales pueden ser sujeto de ejecución por parte de la herramienta una vez el contexto es cargado.

Los comandos poseen las siguientes características:

- Presentan una identificación única dentro de un contexto
- Definen el código a ejecutar una vez que son invocados
- Tienen acceso al estado de la aplicación y del contexto
- Pueden realizar un cambio de contexto como resultado de su ejecución
- Pueden acceder al sitio web y a apis del navegador

## **Módulos**

El módulo es la entidad que engloba los contextos y sus comandos, cada módulo puede definir y/o extender contextos con nuevos comandos de voz.

Los módulos poseen las siguientes características:

- Definen y/o extienden contextos con comandos de voz
- Poseen una identificación única
- Pueden ser provistos por la herramienta o por el usuario

## **Eventos**

La aplicación desarrollada es activada luego de la carga del sitio web, durante este proceso se evalúan los módulos disponibles y se procede a cargar los contextos de los mismos.

Cada módulo opcionalmente puede indicar una función de evaluación, cuyo resultado luego de ser ejecutada indica si el módulo está disponible para ser cargado. Esta evaluación es muy útil por ejemplo si se desea que cierto módulo solamente esté disponible para ciertos tipos de páginas web.

De forma similar, los contextos opcionalmente pueden definir tres tipos de funciones que serán invocadas en distintos momentos de la ejecución de la herramienta.

- Cada vez que se selecciona un elemento: los contextos definidos pueden atender a este evento y decidir si ser cargados para exponer sus comandos de voz que interactúan con el elemento.
- Antes de cargar el contexto: Una vez que la herramienta decide cargar el contexto se da la oportunidad al mismo para que este puede inicializar los valores útiles para la posterior ejecución de los comandos.
- Antes de descargar el contexto: Cada vez que un contexto es descargado, por cuestiones de seguridad su información de estado es eliminada, de esta forma el nuevo contexto no puede acceder al estado del contexto anterior. Un contexto puede querer resguardar esa información antes de ser eliminada o puede necesitar que ciertos valores de su estado permanezcan en el estado de la aplicación una vez es descargado para así poder ser compartidos con el próximo contexto.

Finalmente los comandos disponen de un mecanismo por el cual puede indicar si están disponibles para ser ejecutados, una vez que el contexto es cargado u otro comando de voz correspondiente al mismo contexto es ejecutado. Esta posibilidad es muy útil por ejemplo cuando hay comandos de voz que dejan de ser viables luego de la ejecución de otro comando de voz. Por ejemplo luego de dictar una palabra no debería ser posible ejecutar el comando de voz “deshacer” dos veces. Para ello el comando de voz “deshacer”, luego de evaluar el historial de actividad de inserción de texto, podría decidir si es un comando disponible de ser ejecutado.

### **Información de asistencia al usuario**

Dentro de cada módulo definido opcionalmente se puede indicar la información que es utilizada para asistir al usuario.

A nivel de módulo se puede incluir una descripción del mismo y un ícono. Cada contexto puede describir un ejemplo de uso y, de forma similar, los comandos pueden definir una descripción y la identificación de un grupo al que pertenecen.

A menudo subdividir los grupos de comandos de voz dentro de un contexto es útil para el usuario. Por ejemplo, en el contexto de reproducción de un video, puede ser conveniente dividir los comandos de voz en dos categorías:

- Control de reproducción: Reproducir, detener, pausar, rebobinar [tiempo] y adelantar [tiempo]
- Control de volumen: Subir, bajar y silenciar

### **Internacionalización**

Toda información indicada dentro de un módulo, a excepción de las funciones de ejecución, puede ser internacionalizada. Para ello, todos los textos declarados sujetos a cambios dependiendo el idioma deben poseer el prefijo “i18n-” y dentro una sección del módulo se deben indicar las traducciones correspondientes para los idiomas soportados.

La herramienta desarrollada no posee un idioma principal, sino que se adapta automáticamente por defecto al idioma del navegador.

Todos los módulos incluidos en la aplicación hasta el momento están parametrizados con textos en inglés y español. Para soportar un nuevo idioma basta con replicar las traducciones de alguno de los idiomas existentes a su equivalente en el idioma deseado.

La aplicación desarrollada permite definir módulos que sólo indiquen traducciones, por lo que la falta de soporte de un idioma por parte de un módulo no es un impedimento para que otro módulo indique las traducciones faltantes.

A pesar de que la herramienta virtualmente podría dar soporte a todos los idiomas, existe un límite determinado por el servicio de reconocimiento de voz. En la actualidad la api de reconocimiento de Google utilizada por la herramienta soporta noventa dialectos correspondientes a cuarenta y seis idiomas [20].

## **Proceso de ejecución de un comando de voz**

Múltiples son los componentes que deben interactuar a lo largo de diversas etapas para lograr que un sonido capturado por el micrófono de la computadora derive en una acción deseada dentro del navegador.

En esta sección se describen los componentes y las etapas que forman parte del proceso de ejecución de los comandos de voz.

### **Recepción y envío de señal sonora**

Inicialmente cuando la herramienta es instalada, solicita permisos de acceso a los sonidos capturados por el micrófono. De esta forma la aplicación puede iniciar la captura del sonido en el momento que lo requiera. Para ello se brindan dos sistemas de inicio de reconocimiento de voz.

#### *Reconocimiento de voz continuo*

Cuando este modo está activo, mientras el navegador esté abierto en primer plano, la herramienta está escuchando, analizando y procesando el sonido en búsqueda de comandos de voz a ejecutar. Este modo es ideal para ambientes silenciosos donde el único sonido que va a capturar la computadora es emitido por el usuario y con la intención de ejecutar un comando.

#### *Reconocimiento de voz luego de presionar una tecla*

Este modo de reconocimiento requiere que el usuario presione la tecla `Ctrl` antes de invocar el comando: De esta forma, una vez presionada la tecla, se inicia el proceso de reconocimiento de sonido, el cual está activo hasta obtener la transcripción correspondiente al comando de voz, luego de ello se detiene la captura del sonido y el usuario debe volver a presionar `Ctrl` en caso de querer ejecutar otro comando. Este modo de reconocimiento es ideal para ser ejecutado en entornos ruidosos donde el usuario desea controlar qué sonidos pueden ser un comando de voz y cuales deben ser excluidos del proceso. Si bien este modo es muy práctico, atenta directamente contra el objetivo de no requerir interacción

manual. No obstante puede llegar a ser útil para aquellas personas que consideren que una mínima interacción manual no es un inconveniente.

Una vez iniciado el reconocimiento de voz por cualquier de los modos descritos anteriormente, se envía el flujo de sonidos al servicio remoto de reconocimiento de voz.

### **Recepción de texto resultante**

Mientras el flujo de datos correspondientes al sonido capturado es enviado activamente, se reciben transcripciones parciales y no finales correspondientes al sonido procesado. Son estas transcripciones las que muestra la barra de estado para demostrar al usuario que el comando de voz está en proceso de reconocimiento.

La herramienta puede recibir como resultado del proceso de reconocimiento del sonido enviado hasta diez transcripciones posibles, cada una de ellas con un grado de probabilidad de acierto y con una marca de si son resultados finales o parciales.

### **Resolución de comando de voz**

Solo las transcripciones finales con una probabilidad de acierto superiores al 40% son consideradas como comandos de voz. Son tenidas en cuenta sólo las transcripciones finales ya que las parciales, por lo general, como su nombre lo indica, son resultados tempranos del proceso sujeto a ser mejorado luego de sucesivos refinamientos. Desechar transcripciones cuyo grado de acierto sea inferior al 40% fue una medida que se alcanzó mediante ensayo y error durante la fase de prueba.

Regularmente se obtienen múltiples transcripciones que pueden ser comandos de voz y no coinciden con exactitud con los comandos de voz disponibles para ser ejecutados. Por ejemplo, al indicar verbalmente el comando de voz “abrir” se puede obtener la transcripción “habría”. Al buscar por el comando “habría” se encuentra que el mismo no existe, por lo que es necesario realizar aproximaciones con el fin de determinar si existe un comando similar que satisfaga la intención del usuario.

Con el fin de realizar aproximaciones precisas, inicialmente la herramienta transforma todos los comandos disponibles y las transcripciones recibidas en fonemas utilizando el algoritmo Metaphone [21]. Posteriormente utilizando el Coeficiente de Sorensen-Dice [22], se compara los fonemas correspondientes a las transcripciones y los comandos disponibles con el fin de obtener el par (transcripción, comando) que presenten mayor grado de similitud.

Luego de sucesivas pruebas realizadas a través de ensayo y error, se concluyó que el par (transcripción, comando) debe poseer al menos un 70% de nivel de similitud para que el comando sea elegido para ser ejecutado.

La etapa de resolución de comando de voz puede finalizar con dos resultados; la existencia de un comando a ejecutar o la ausencia del mismo. En caso de no existir un comando a ejecutar se informa por pantalla que no existe el comando de voz indicado. De lo contrario se procede a la etapa siguiente.

## Ejecución de comando de voz

Una vez elegido el comando de voz, la reproducción del mismo consiste en ejecutar el código definido para el mismo.

Cuando la función del comando es ejecutada se esperan dos tipos de efectos; externos o internos a la aplicación. Los primeros corresponden, por ejemplo, a cambios realizados en el sitio web actual o la ejecución de acciones en el navegador. Por el contrario los efectos internos a la aplicación son, por ejemplo, la solicitud de cambio de contexto o la actualización del estado del contexto.

## Panel de administración

Una vez instalada la herramienta, la misma posee diversas opciones que pueden resultar de interés ser configuradas por parte del usuario.

Para ingresar al panel de administración basta con hacer click sobre el ícono de la extensión, el cual se encuentra a la derecha de la barra de navegación.



Figura 5.12: Barra de dirección de Google Chrome

## Encendido y apagado

La herramienta es capaz de reconocer los comandos de voz si y sólo si el panel de control de la extensión está abierto en una de las pestañas. Basta con cerrar el mismo para desactivar la aplicación en todo en el navegador.

Se consideró un requisito importante de seguridad que el usuario pueda identificar fácilmente en qué pestaña del navegador se realiza el reconocimiento de voz y que este proceso pueda ser simplemente desactivado en caso de que el usuario desee evitar que se continúe capturando el sonido.

La pestaña correspondiente al panel de administración puede ser cerrada a través de un comando de voz, por lo que oralmente sólo se podría apagar la aplicación, la activación es exclusivamente manual.

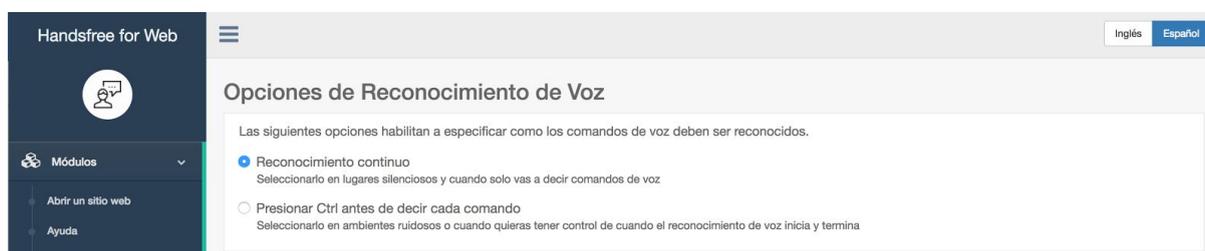


Figura 5.13: Sección de configuración de herramienta

## Configuración de idioma

Actualmente se brinda la opción de elegir entre inglés y español, pero como se describió en el apartado de “Internacionalización” la aplicación puede soportar una gran cantidad de lenguajes.

En el encabezado del panel de administración el usuario puede optar por cambiar el idioma seleccionado. Una vez realizado el cambio de lenguaje, la aplicación a lo largo de todas las pestañas, actualizará automáticamente el idioma en el que debe ser ejecutada.

## Configuración de modo de ejecución

Anteriormente en el apartado de “Recepción y envío de señal sonora” se describió cómo la aplicación soporta dos modos de reconocimiento:

- Reconocimiento de voz continuo
- Reconocimiento de voz luego de presionar una tecla

El usuario puede cambiar de modo en la parte superior del contenido del panel de control. Una vez seleccionado el cambio de modo, la aplicación a lo largo de todas las pestañas actualizará su modo automáticamente.

## Visualización de módulos instalados

A lo largo de la barra lateral se puede observar la lista de módulos instalados. Cada uno de ellos brinda información detallada acerca de los comandos y contextos que define junto con los textos de ayuda y una sección que permite al usuario practicar los comandos.

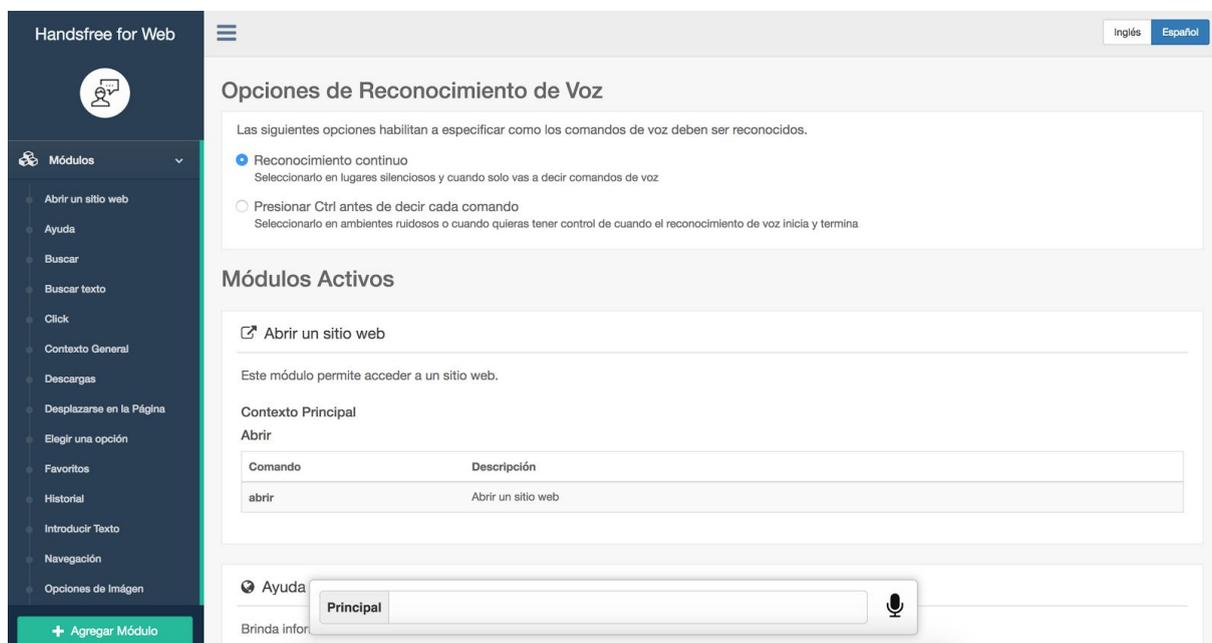


Figura 5.14: Lista de módulos activos

## Administración de módulos

Como se presentará más detalladamente en la sección siguiente, la herramienta brinda la posibilidad de instalar módulos desarrollados por el usuario con el fin de extender la

aplicación. Estos módulos pueden ser instalados al hacer click en el botón inferior izquierdo del panel de administración, el cual presenta la leyenda “Agregar Módulo”. El mismo se muestra automáticamente en la lista de módulos instalados junto con todos sus datos de la misma forma que los módulos incluidos por defecto dentro de la herramienta.

Un módulo agregado por el usuario puede ser removido al hacer click en el botón “Eliminar” que se encuentra en la esquina superior derecha de la sección que describe los comandos del módulo.

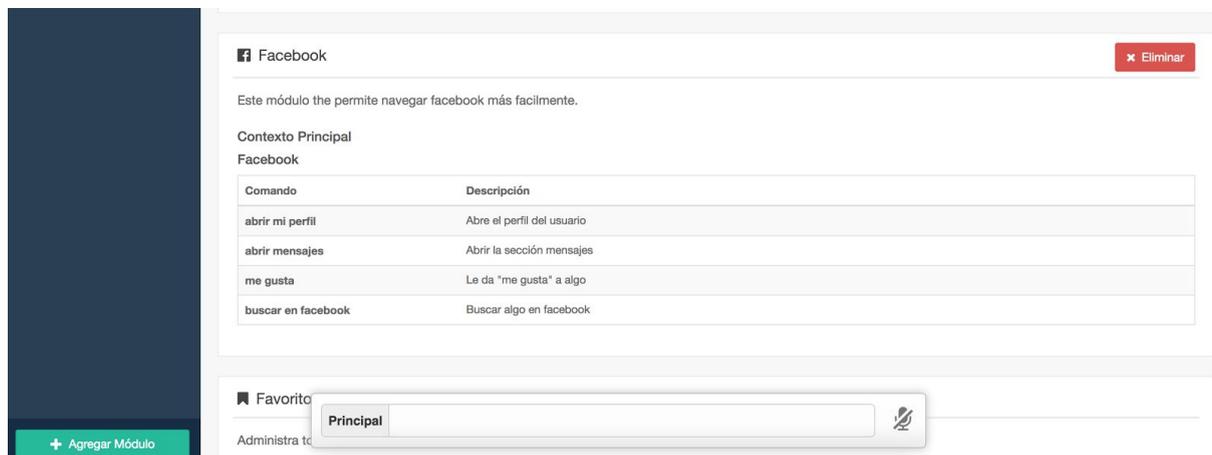


Figura 5.15: Opción remover un módulo previamente agregado por el usuario

## Funcionalidad Soportada

A lo largo del capítulo cuatro, se describió un modelo de interacción web a través del uso exclusivo de comandos de voz sin la necesidad de resignar funcionalidad soportada por los sitios web actuales. En dicho capítulo se realizó un profundo análisis y clasificación de la funcionalidad que es necesario satisfacer. La herramienta desarrollada soporta los tres tipos de funcionalidades mencionadas en el modelo propuesto.

## Funcionalidades exclusivas del navegador

Como se describió en el capítulo anterior definimos funcionalidades exclusivas del navegador como todas aquellas acciones que son efectuadas directamente por el navegador sin intervención de un sitio web.

En esta sección se mencionan las funcionalidades del navegador soportadas por la herramienta a través del uso de comandos de voz.

### Abrir un sitio web

La herramienta permite abrir una dirección web invocando el siguiente comando de voz.

- Abrir  
Solicita la dirección a abrir por pantalla
- [dirección]  
Al especificar la dirección a abrir carga el sitio web en la pestaña actual

## Búsqueda de contenidos en la web

Es posible buscar palabras claves en un buscador de internet.

- **Buscar**  
Solicita indicar las palabras claves a buscar
- **[palabras claves]**  
Al indicar las palabras claves a buscar se abre en la pestaña actual los resultados de la búsqueda

## Gestión de favoritos

A continuación los comandos de voz soportados para la gestión de favoritos.

- **Favoritos**  
Detalla los favoritos guardados
- **[referencia]**  
Dependiendo de la referencia expresada puede ejecutar la apertura de un favorito o la eliminación del mismo.

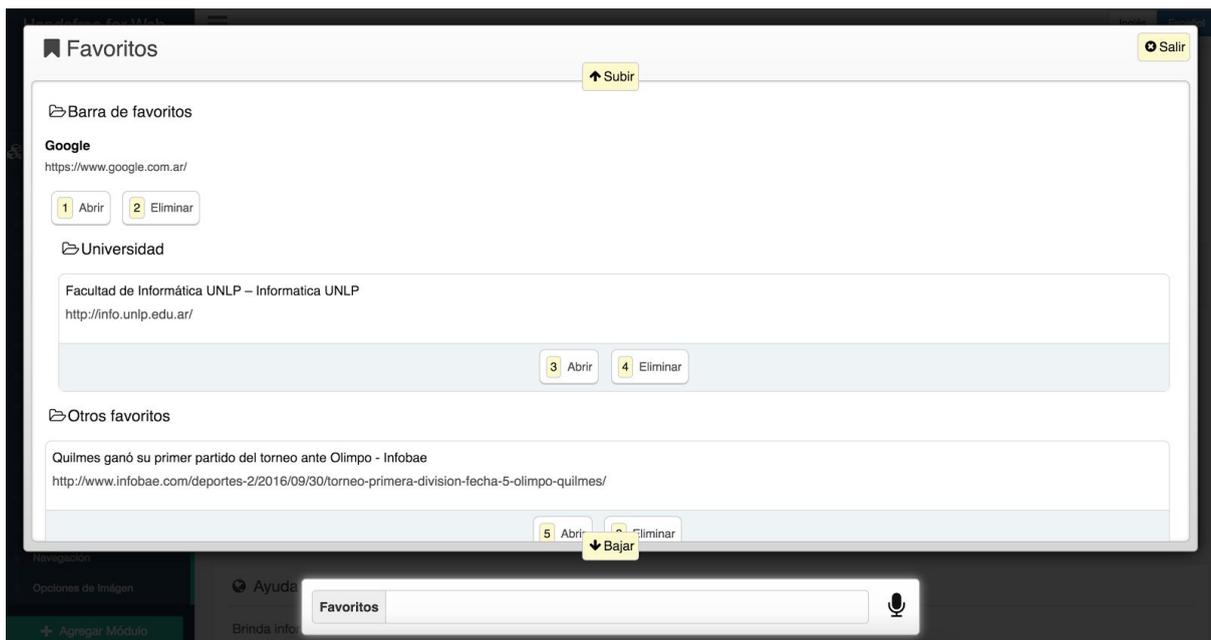


Figura 5.16: Administración de favoritos

## Sitios más visitados

A continuación se presenta los comandos de voz soportados para la visualización de los sitios más frecuentes.

- **Sitios más visitados**  
Detalla los sitios más visualizados por el usuario
- **[referencia]**  
Abre el sitio indicado

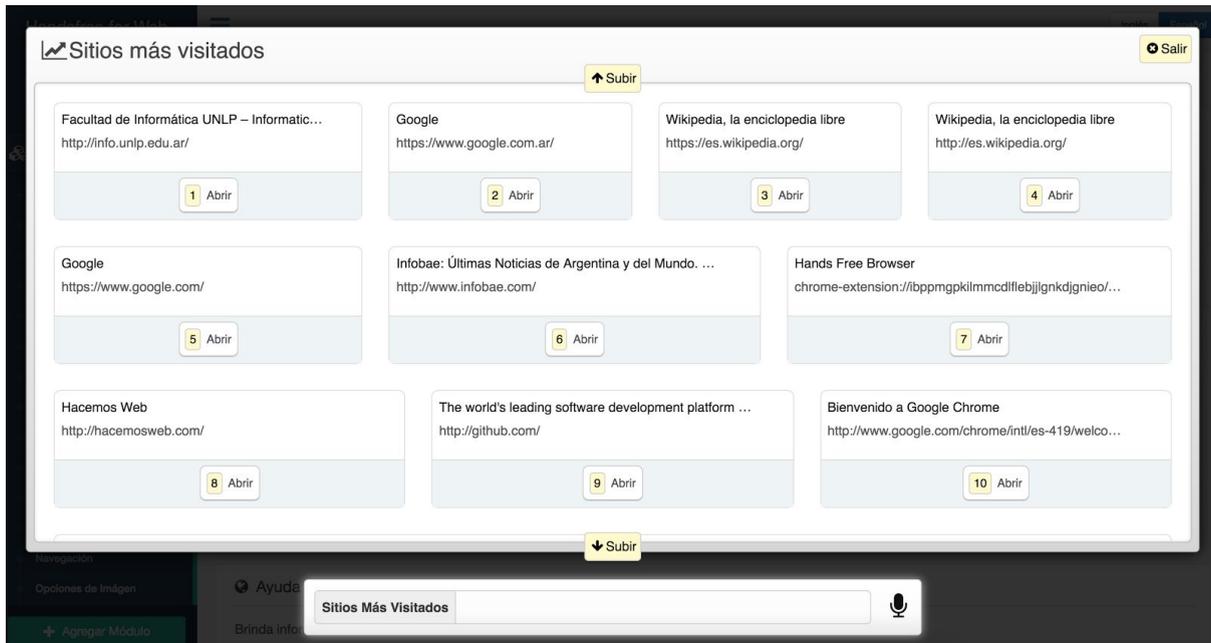


Figura 5.17: Sección sitios más visitados

## Historial

A continuación se enumeran los comandos de voz soportados para la gestión del historial.

- *Historial*  
Detalla el acceso en orden cronológico a sitios web
- *[referencia]*  
Abre el sitio indicado

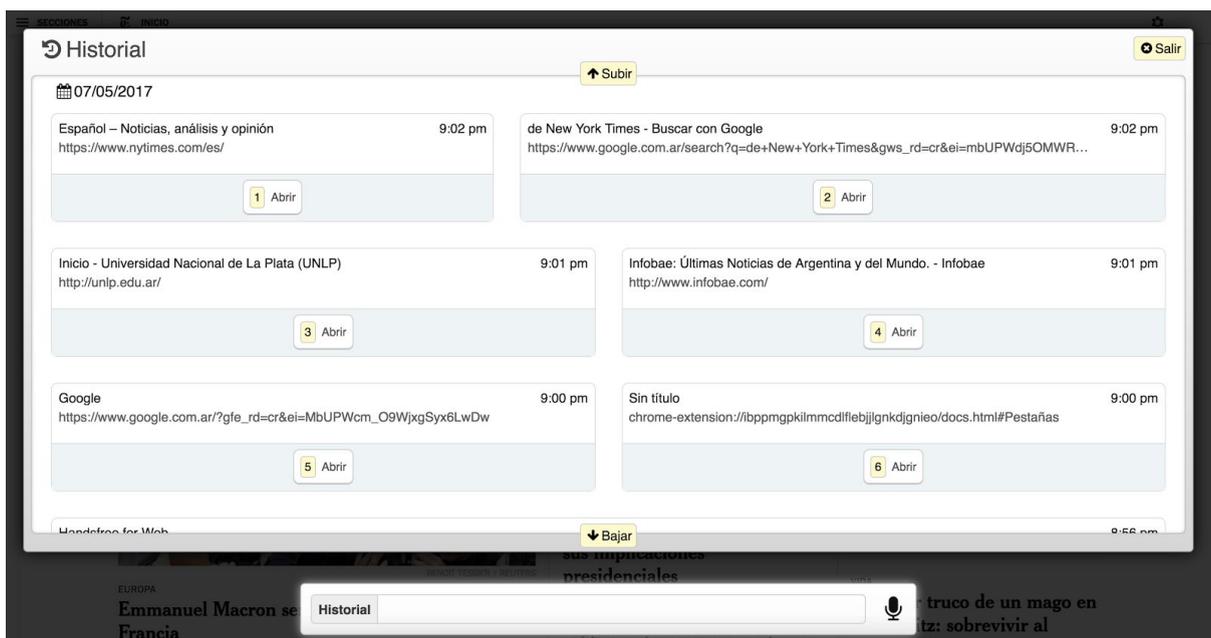


Figura 5.18: Sección Historial

## Descargas

Se presentan a continuación los comandos de voz soportados para la gestión de descargas.

- *Descargas*  
Detalla las descargas realizadas en orden cronológico

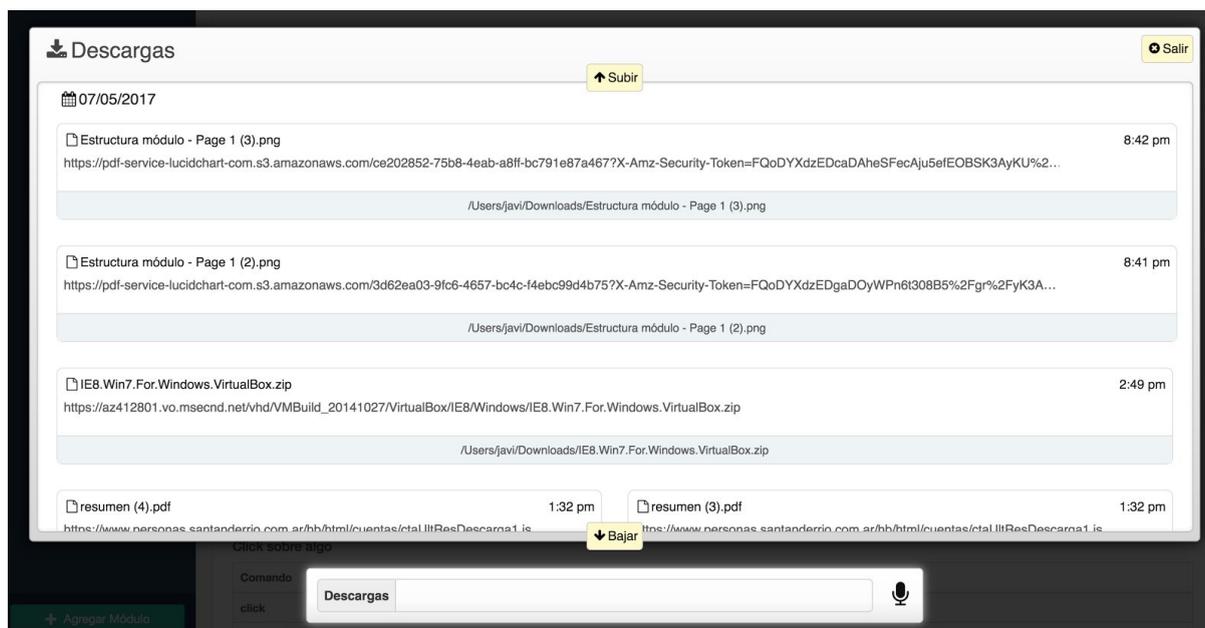


Figura 5.19: Sección descargas

Por cuestiones de seguridad, el navegador no brinda la posibilidad de ejecutar archivos previamente descargados desde el contexto de una extensión de Google Chrome, a no ser que la misma haya gestionado la descarga. Se deja el comando encargado de abrir una descarga como un trabajo a futuro.

## Búsqueda de texto

La búsqueda de texto se realiza en dos etapas, inicialmente se solicita el inicio de la sesión de búsqueda de texto y luego se indica el texto a buscar. Posteriormente se puede navegar los resultados obtenidos.

A continuación, se muestran los comandos de voz soportados para realizar búsquedas de texto.

- *Buscar texto*  
Inicia la sesión de búsqueda de texto

Una vez iniciada la sesión de búsqueda de texto se pueden indicar los siguientes comandos

- *[Texto a buscar] > enviar*

- Busca la primer ocurrencia del texto indicado
- *Siguiente*  
Busca la siguiente ocurrencia del texto indicado
- *Anterior*  
Busca la anterior ocurrencia del texto indicado



Figura 5.20: Buscar un texto

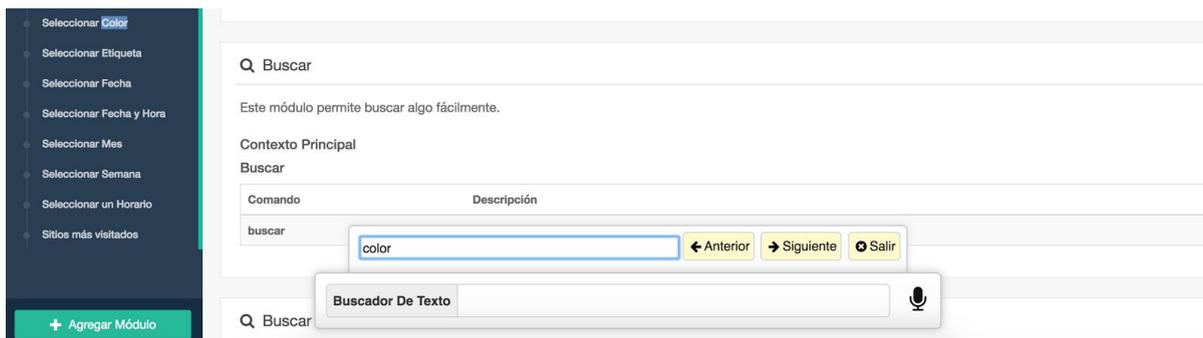


Figura 5.21: Navegar ocurrencias de texto buscado

## Desplazamientos

A continuación se enumeran los comandos de voz soportados para desplazarse verticalmente dentro del sitio web.

- *Bajar*  
Desplaza el contenido hacia abajo
- *Subir*  
Desplaza el contenido hacia arriba
- *Arriba*  
Desplaza el contenido hasta el inicio
- *Abajo*  
Desplaza el contenido hasta el final

## Navegación

Los comandos de voz soportados para la gestión de la navegación de los sitios web son los siguientes:

- *Atrás*  
Carga al sitio anterior
- *Adelante*  
Carga al sitio siguiente
- *Actualizar*  
Recarga el sitio actual
- *Parar*  
Detiene la carga del sitio

## Gestión de pestañas

A continuación se presentan los comandos de voz soportados para la gestión de pestañas.

- *Nueva pestaña*  
Abre una nueva pestaña
- *Cerrar pestaña*  
Cierra una pestaña
- *Siguiente pestaña*  
Cambia a la siguiente pestaña
- *Anterior pestaña*  
Cambia a la anterior pestaña

## Funcionalidades genéricas de los sitios web

En el capítulo anterior se definieron funcionalidades genéricas de los sitios como a aquellas acciones que impactan directamente sobre los componentes web estándares presentes dentro del sitio web visualizado.

En esta sección se describe el soporte a las funcionalidades provistas por los componentes genéricos incluidos en los sitios web.

## Click y Seleccionar

Al realizar un análisis acerca de cómo las personas interactúan con los diferentes componentes web se llegó a la conclusión de que existen dos acciones básicas que inician el proceso de interacción con un elemento presente en un portal. Estas acciones las llamaremos “click” y “seleccionar”.

Manualmente se puede realizar un “click” sobre un elemento presionando el botón izquierdo del mouse cuando el cursor está posicionado sobre el objetivo, o alternativamente en ciertas circunstancias se puede presionar múltiples veces la tecla `TAB` hasta lograr hacer foco en el

elemento deseado, posteriormente para simular el click es necesario presionar la tecla ENTER.

De forma similar, manualmente se puede realizar la acción “seleccionar” sobre un elemento presionando el botón derecho del mouse cuando el cursor está posicionado sobre el objetivo, o alternativamente, en ciertas circunstancias, se puede presionar múltiples veces la tecla TAB hasta lograr hacer foco en el elemento deseado.

Las dos acciones descritas “click” y “seleccionar” vienen incluidas como comando de voz por la herramienta. Las mismas son utilizadas como primer paso en la interacción con los componentes web presentes en todos los sitios.

Por ejemplo al indicar el comando “click”, se muestran en la pantalla un marcador por cada elemento que puede llegar a ser sujeto de interacción luego de recibir un click. Por ejemplo, un botón, un vínculo, un campo de texto o un video.

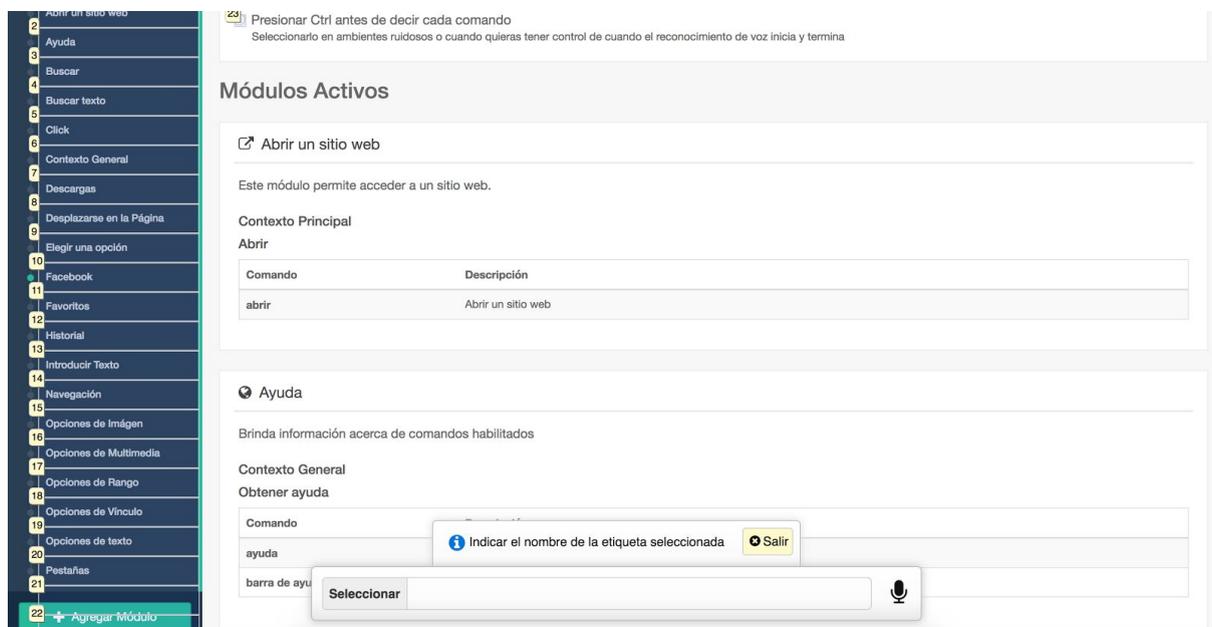


Figura 5.22: Elección de marcadores luego de indicar el comando “click”

Alternativamente al comando “click”, con el fin de facilitar la tarea de búsqueda del marcador por parte de usuario, se brindan comandos de voz que permiten hacer “click” solo en algunos tipos de elementos. Esto es muy útil para al usuario cuando existen una gran cantidad de elementos que puedan a ser sujetos de click en la misma pantalla.

- click audio
- click botón
- click opción
- click color
- click fecha
- click archivo
- click imagen

- click vínculo
- click multimedia
- click rango
- click opción única
- click opciones
- click texto
- click video

En oposición a un contexto donde existen muchos elementos que pueden ser sujetos de click, es relativamente común que solo exista un elemento de cierto tipo. Por ejemplo un video. En este caso si el usuario invoca a el comando “click video”, el video inmediatamente va a recibir la acción click, no es necesario que el usuario indique la referencia de un elemento que es único en la pantalla.

Luego de ejecutar una acción de “click” sobre algunos de los elementos mencionados en la lista, puede ser necesario ejecutar otro u otros comandos para lograr el resultado deseado. Por ejemplo, indicar el texto correspondiente a un campo de formulario conlleva al menos dos comandos de voz. Inicialmente es necesario hacer “click” sobre el campo y luego proceder a dictar el texto deseado. Como caso opuesto, no es necesario invocar sucesivos comandos luego de hacer “click” sobre un botón, salvo que se desee presionarlo múltiples veces por algún razón especial.

De forma similar al comando “click”, encontramos que el comando “seleccionar” muestra por pantalla un marcador por cada elemento que pueda ser sujeto de interacción si el mismo es seleccionado. Por ejemplo, un texto y una imagen.



Figura 5.23: Selección de elementos luego de indicar el comando “seleccionar”

Junto con el comando “seleccionar” y con el fin de facilitar la tarea de búsqueda del marcador por parte de usuario, en los casos en que exista una gran cantidad de elementos que puedan a ser sujetos de selección, se proveen los siguientes comandos.

- seleccionar audio
- seleccionar imagen
- seleccionar vínculo
- seleccionar multimedia
- seleccionar texto
- seleccionar video

De forma opuesta al comando “click”, luego de ejecutar “seleccionar” sobre un elemento, siempre es necesario indicar un comando siguiente o un segundo comando. Por ejemplo, si se seleccionó un texto, el siguiente comando a ejecutar puede ser copiar. Para ello, como se menciona al final de esta sección, se provee el comando “copiar”.

### **Interacción con Vínculos**

Los vínculos o hipervínculos son uno de los elementos más simples que componen un sitio web. Normalmente se utilizan los vínculos para brindar soporte de navegación entre distintos sitios web, o secciones internas del mismo portal.

La herramienta a través del uso de los comandos “click” y “seleccionar” propone las siguientes opciones de interacción:

- click > [referencia]  
Ejecuta el click sobre el vínculo indicado, la acción a realizar depende de la intención programada por el sitio web. Puede ser navegar hacia una nueva dirección, navegar hacia un posición dentro de la misma página, o cualquier otra acción especificada.
- seleccionar > [referencia]  
Habilita al usuario a indicar una acción basada en el vínculo, ver a continuación las opciones disponibles.

Luego de seleccionar un vínculo el usuario puede indicar los siguientes comandos:

- Copiar  
Copia en el portapapeles el texto correspondiente al vínculo
- Copiar dirección  
Copia en el portapapeles la dirección a la que apunta el vínculo
- Abrir en nueva pestaña  
Abre en una nueva pestaña la dirección especificada en el vínculo

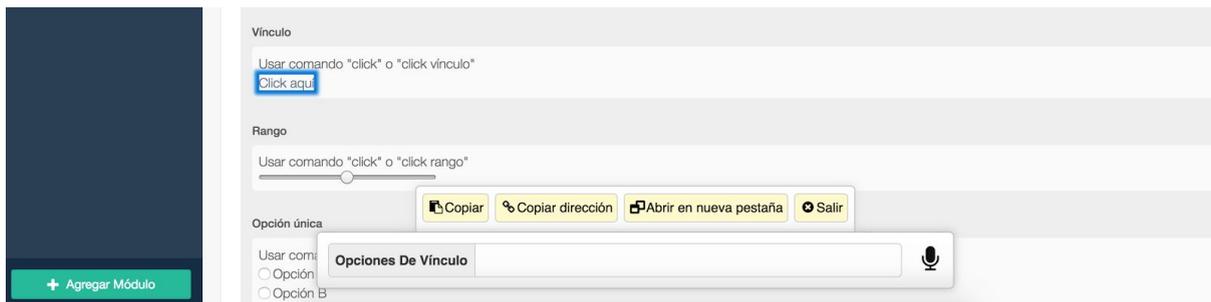


Figura 5.24: Opciones de vínculo

### Interacción con Botones

Un componente web común en los sitios web son los botones. Al contrario de los vínculos ellos pueden ser solo presionados.

Un botón puede ser presionado utilizando el comando de voz

- click > [referencia]  
Presiona el botón indicado

### Interacción con lista de selección de opciones

Al completar un formulario es habitual seleccionar una o varias opciones de un lista. A veces esta diversidad de opciones se presenta como un campo desplegable o como una lista de valores los cuales pueden ser seleccionados usando sucesivamente el mouse o el teclado.

Para seleccionar uno o más valores de una lista de opciones, inicialmente es necesario hacer click sobre la lista invocando el comando especificado a continuación.

- click > [referencia]  
Inicia el modo de selección de opciones

Una vez realizado el click sobre la lista, se muestra por pantalla todas las opciones de selección disponibles. El usuario puede indicar sus preferencias mencionando los marcadores exhibidos.

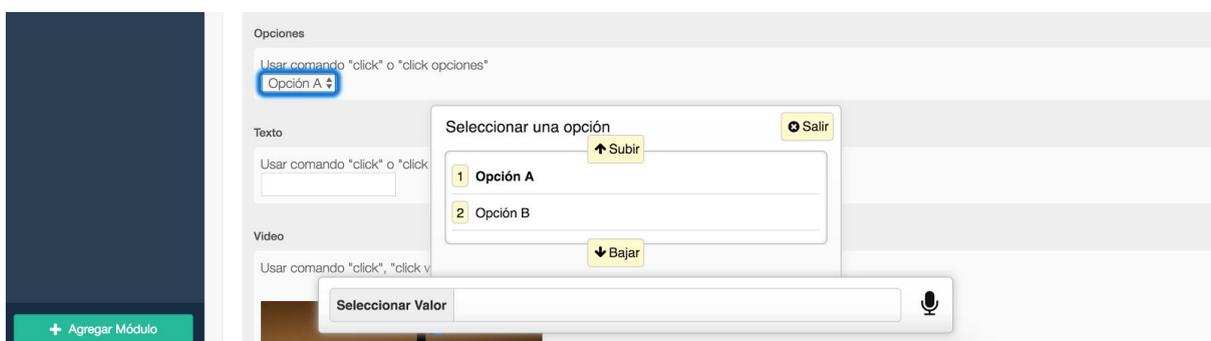


Figura 5.25: Selección de opciones de una lista

En caso de haber más opciones de las que se pueden mostrar por pantalla, el usuario puede mencionar los comandos de voz “Subir” y “Bajar” para ver más opciones.

### **Interacción con campos de selección de opciones**

Probablemente luego del campo de texto, los elementos web presentes en formularios más difundidos son las casillas de selección, las cuales vienen en dos formatos; de selección única y de selección múltiple. Las primeras por lo general son redondas y las segundas cuadradas. Al igual que los botones, una casilla de selección puede ser solamente presionada. Por esta razón la herramienta permite interactuar con estos elementos solo a través del uso del comando:

- click > [referencia]  
Marca o desmarca la opción indicada

### **Interacción con reproductores multimedia**

El estándar HTML 5 definió dos componentes web útiles para reproducir contenido de audio y video. En la actualidad todos los navegadores populares soportan estos componentes y es por ello que la mayoría de los sitios web modernos reproducen el contenido multimedia usando los reproductores genéricos provistos.

Los reproductores soportados por HTML presentan una gran cantidad de funcionalidades, todas ellas son cubiertas con los siguientes comandos de voz soportados por la herramienta:

- Click > [referencia]  
Inicia el modo de selección de opciones de reproducción
- Seleccionar > [referencia]  
Inicia el modo de selección de opciones de reproducción

Una vez seleccionado el reproductor se puede invocar los siguientes comandos de voz:

#### **Control de reproducción**

- reproducir
- pausar
- parar
- rebobinar
- adelantar
- rebobinar 30 segundos
- rebobinar 1 minuto
- rebobinar 3 minutos
- rebobinar 5 minutos
- rebobinar 10 minutos
- rebobinar 15 minutos

- rebobinar 20 minutos
- rebobinar 30 minutos
- rebobinar 45 minutos
- rebobinar 60 minutos
- rebobinar 1 hora
- adelantar 30 segundos
- adelantar 1 minute
- adelantar 3 minutos
- adelantar 5 minutos
- adelantar 10 minutos
- adelantar 15 minutos
- adelantar 20 minutos
- adelantar 30 minutos
- adelantar 45 minutos
- adelantar 60 minutos
- adelantar 1 hora

#### Control de volumen

- silencio
- subir volumen
- bajar volumen



Figura 5.26: Opciones de reproductor multimedia

#### Interacción con campos de selección de color

Si bien los campos de selección de color no se han difundido en la actualidad, los mismos son partes del estándar HTML 5 y están soportados por la mayoría de los navegadores modernos. Estos campos permiten al usuario seleccionar un color incluido en una paleta de colores.

La herramienta posibilita la interacción con el componente de selección de color a través del uso de los siguientes comandos:

- Click > [referencia]  
Inicia session de selección de color

Una vez hecho “click” sobre el campo de color, se permite seleccionar el color elegido a través de la invocación de dos comandos correspondientes, los cuales corresponden a los ejes horizontales y verticales del color deseado:

- [número del 1 al 10]  
Selecciona la intensidad del color
- [número del 11 al 30]  
Selecciona el color

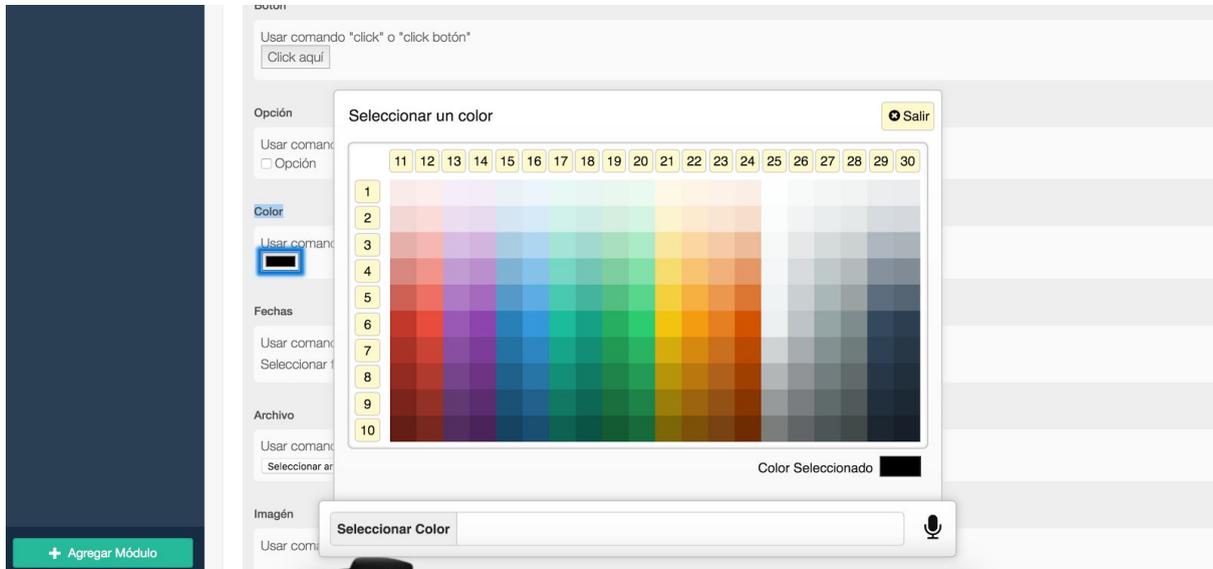


Figura 5.27: Selección de un color

### Interacción con campos de selección de un valor incluido en un rango

Un elemento definido por HTML 5 muy útil, aunque todavía no es muy popular, es el campo de selección de un valor de un rango determinado.

La herramienta permite seleccionar el valor del rango al indicar los siguientes comandos:

- Click > [referencia]  
Al hacer click sobre el elemento de rango se inicia una sesión donde se puede seleccionar un valor contemplado dentro del rango indicado

Luego de hacer click sobre el campo de rango, se pueden ejecutar los siguientes comandos de voz:

- [valor deseado]  
Selecciona el valor deseado en caso de ser un valor válido



Figura 5.28: Selección de un valor perteneciente a un rango de valores

## Interacción con campos de texto

Definir e implementar la interacción a través de comandos de voz con un campo de texto fue una tarea compleja. Por lo general la interacción con el campo de texto se realiza utilizando casi todas las teclas del teclado, incluso existen combinaciones de estas que son útiles para ejecutar acciones que no se pueden realizar con solo presionar una tecla. Por ejemplo escribir una letra mayúscula o el símbolo arroba que está presente en todas las direcciones de email.

Inicialmente se tuvo en cuenta que los campos de texto vienen en dos formatos; de una sola línea y multilínea. Esta multiplicidad de presentación introduce la necesidad de brindar la opción de saltos de línea dependiendo el tipo de elemento.

Dado que la herramienta se sirve de comandos de voz para interactuar con el componente, es deseable que el usuario pueda introducir texto de dos modos distintos:

- Dictado de texto de forma libre
- Dictado de texto letra por letra

Un aspecto importante a tener en cuenta fue que al dictar un texto puede ocurrir que la transcripción del mismo no sea perfectamente precisa por lo que es necesario brindar funcionalidades que permitan deshacer el texto recientemente escrito para que el usuario pueda re-dictar el texto nuevamente.

De forma similar a como se interactúa con los otros campos de formulario, inicialmente es necesario ejecutar un “click” sobre el elemento de texto.

- Click > [referencia]  
Inicia el modo de inserción de texto

Luego de iniciado el modo de inserción de texto sobre un campo determinado, es posible ejecutar los siguientes comandos:

- borrar  
Borra la última palabra
- borrar línea  
Borra la última línea de texto
- borrar todo

- Borra todo el texto
- pegar  
Pega texto previamente copiado
- deshacer  
Deshace la última acción
- rehacer  
Restaura la acción revertida
- enviar  
Envía el formulario al cual el campo de texto corresponde
- [número del 1 al 54]  
Ejecuta la acción indicada en la tecla del teclado exhibido cuando se está en modo de inserción de texto.
- [texto libre]  
Transcribe el texto mencionado



Figura 5.29: Introducción de texto en un campo de texto de una sola línea

### Interacción con campos de calendario

El estándar HTML 5 provee una diversa variedad de componentes web que tienen como objetivo asistir al usuario en la introducción de información referente a datos temporales.

En este apartado, con el fin de no ser extenso y repetitivo, se va a describir la interacción con el campo de fecha y hora. La elección de este campo es debido a que es el que mayores opciones de interacción presenta. No obstante la herramienta soporta la interacción con todos los campos relacionados con la inserción de datos temporales, los cuales son:

- Selección de fecha (día-mes-año)
- Selección de fecha y hora (día-mes-año hora:minuto am/pm)
- Selección de mes (mes-año)
- Selección de semana (número de semana-año)
- Selección de horario (hora:minuto am/pm)

Al igual que la mayoría de los elementos web, para interactuar con ellos es necesario hacer click para iniciar la sesión de interacción:

- Click > [referencia]  
Inicia el modo de interacción con el campo de fecha y hora

Una vez realizado el click sobre el elemento, se despliega un cuadro de diálogo, el cual permite visualizar los distintos comandos de voz que pueden ser ejecutados con el fin de ajustar la fecha y hora deseada. A continuación se presentan los comandos disponibles:

- Mes > [1 .. 12]  
Permite seleccionar el mes
- Día > [1 ... 31]  
Permite seleccionar el día
- Año > [número]  
Permite seleccionar el año
- Hora > [0 .. 24]  
Permite seleccionar la hora
- Minutos > [0 ... 60]  
Permite seleccionar los minutos



Figura 5.29: Selección de campos de formulario

## Funcionalidades específicas de los sitios web

Como se describió en el capítulo anterior, las funcionalidades específicas de los sitios web son todas aquellas acciones derivadas de un componente web no estándar o aquellas que derivan de la ejecución de funcionalidad de tipo específico.

Siguiendo la propuesta de soporte de funcionalidad indicada previamente, en esta sección se describe la gestión de módulos personalizados de comandos de voz provistos por el usuario y el soporte de una API de comandos de voz nativos del sitio web.

### Módulos personalizados de comandos de voz

Como se especificó en la sección “Diseño de la herramienta”, todo contexto y comando de voz es provisto por un módulo. La herramienta por defecto provee una gran cantidad de

módulos, los cuales aseguran contar con el soporte de las funcionalidades del navegador y las genéricas de los sitios web.

### Módulos

La única diferencia entre un módulo personalizado provisto por el usuario y uno que provee la herramienta es que el código correspondiente al primero puede ser cargado y descargado cuando el usuario lo requiera. Por el contrario los módulos por defecto incluidos en la aplicación no pueden ser deshabilitados o borrados, sin embargo, pueden ser extendidos o sobrescritos por módulos personalizados.

A continuación, se brindan detalles técnicos acerca de las componentes de un módulo personalizado.

Inicialmente un módulo es un archivo de texto plano que debe definir una estructura específica a través de código escrito en Javascript. Por esta razón la extensión del archivo debe ser “.js”

La estructura general de un módulo es un objeto javascript con ciertas propiedades específicas que él mismo debe definir. El archivo correspondiente al módulo personalizado puede indicar directamente el objeto u opcionalmente una función, que al ser ejecutada, retorna el objeto correspondiente al módulo.

### Módulo simple

```
{  
  //module properties  
}
```

Sección de Código 5.1: Módulo de comandos de voz simple

### Módulo dinámico

```
function () {  
  return {  
    //module properties  
  }  
})
```

Sección de Código 5.2: Módulo de comandos de voz dinámico

Todo módulo debe definir un conjunto de propiedades con ciertos valores específicos. A continuación el detalle de los mismos.

Propiedad	Tipo propiedad	Tipo de dato	Descripción
name	requerida	string	Indica el nombre del módulo
description	opcional	string	Indica un texto informativo que describe la utilidad del módulo

icon	opcional	string	Referencia a un icono de la librería font-awesome [23]
isActive	opcional	function	Al retornar true o false permite controlar cuando el módulo debe estar activo
contexts	opcional	array	Define los contextos junto con sus comandos de voz que el módulo provee.
i18n	opcional	object	Define las traducciones para los textos de este módulo. No aplica a los textos del contexto o comandos.

A continuación como ejemplo se muestra el primer nivel de un módulo encargado de dar de alta un favorito. Este módulo está activo solamente en los sitios cuyo dominio es “google.com”.

```

{
  name: 'i18n-name',
  description: 'i18n-description',
  isActive: function () {
    return window.location.hostname === 'google.com';
  },
  icon: 'fa fa-bookmark',
  contexts: [
    //contexts
  ],
  i18n: {
    en: {
      'name': 'Add to Bookmarks',
      'description': 'Manage when a website is part of the favorite
list'
    },
    es: {
      'name': 'Agregar a Favoritos',
      'description': 'Administra cuando un sitio web está en la lista de
favoritos'
    }
  }
}

```

Sección de Código 5.3: Ejemplo de definición de un módulo de comandos de voz

### Contextos

Cada módulo es capaz de definir una lista de contextos de comandos de voz que soporta o extiende. La definición de un contexto brinda principalmente información acerca del mismo y una lista de comandos que soporta.

A continuación, se brindan las propiedades que componen un contexto.

Propiedad	Tipo propiedad	Tipo de dato	Descripción
context	requerida	string	Indica el nombre clave del contexto.
name	opcional	string	Indica el nombre del contexto que se muestra al usuario.
commands	opcional	array	Lista de comandos de voz
i18n	opcional	object	Define las traducciones para los textos del contexto.
setup	opcional	function	Se ejecuta previamente a que el contexto sea cargado.
teardown	opcional	function	Se ejecuta antes de que el contexto deja de estar cargado.
switchOnSelectElement	opcional	function	Cada vez que el usuario selecciona un elemento de un sitio web se ejecuta esta función. En caso de retornar true el contexto será cargado.

A continuación, se detalla los contextos que definen el módulo que brinda comandos de voz que tienen como fin agregar a favoritos sitios web que se encuentran dentro del dominio "google.com".

```

{
  name: 'i18n-name',
  description: 'i18n-description',
  isActive: function () {
    return window.location.hostname === 'google.com';
  },
  icon: 'fa fa-bookmark',
  contexts: [{
    context: 'root',
    name: 'i18n-name',
    setup: function (state) {
      //do something on setup
    },
    teardown: function (state) {
      //do something on teardown
    },
    commands: [
      //command list

```

```

    ],
    i18n: {
      en: {
        'name': 'main'
      },
      es: {
        'name': 'principal'
      }
    }
  }],
  i18n: {
    en: {
      'name': 'Add to Bookmarks',
      'description': 'Manage when a website is part of the favorite
list'
    },
    es: {
      'name': 'Agregar a Favoritos',
      'description': 'Administra cuando un sitio web está en la lista de
favoritos'
    }
  }
}
}
}

```

Sección de Código 5.4: Ejemplo de definición de módulo que posee un contexto

### Comandos

La función principal de un contexto consiste en definir comandos. Para ello, es necesario brindar información acerca del comando junto con un código que será ejecutado cuando el nombre asociado sea invocado por el usuario.

A continuación, se presentan las propiedades que componen un comando:

Propiedad	Tipo propiedad	Tipo de dato	Descripción
name	requerida	string	Indica el nombre del comando
action	requerida	function	Función que será ejecutada cuando el usuario invoca el comando
isActive	opcional	function	Al retornar true o false permite controlar cuando el comando se encuentra activo.
group	opcional	string	Nombre de grupo al que pertenece el comando. Útil para mostrarle al usuario los comandos divididos por grupos según alguna clasificación.

help	opcional	string	Descripción del comando
switchToContext	opcional	string	Una vez ejecutado el comando se cambia al contexto indicado.

Para finalizar se presenta un módulo completo, el cual incluye un comando de ejemplo.

```

{
  name: 'i18n-name',
  description: 'i18n-description',
  isActive: function () {
    return window.location.hostname === 'google.com';
  },
  icon: 'fa fa-bookmark',
  contexts: [{
    context: 'root',
    name: 'i18n-name',
    setup: function (state) {
      //do something on setup
    },
    teardown: function (state) {
      //do something on teardown
    },
    commands: [{
      name: 'i18n-command.add-bookmark',
      group: 'i18n-group.add-bookmark',
      help: 'i18n-help.add-bookmark',
      action: function (state) {
        //save bookmark
      }
    }
  ]],
  i18n: {
    en: {
      'name': 'main',
      'command.add-bookmark': 'add bookmark',
      'help.add-bookmark': 'Saves the current page as a bookmark',
      'group.add-bookmark': 'Bookmarks'
    },
    es: {
      'name': 'principal',
      'command.add-bookmark': 'agregar a favoritos',
      'help.add-bookmark': 'Guarda el sitio actual en favoritos',
      'group.add-bookmark': 'Favoritos'
    }
  }
}

```

```

    }
  }],
  i18n: {
    en: {
      'name': 'Add to Bookmarks',
      'description': 'Manage when a website is part of the favorite
list'
    },
    es: {
      'name': 'Agregar a Favoritos',
      'description': 'Administra cuando un sitio web está en la lista de
favoritos'
    }
  }
};

```

Sección de Código 5.5: Ejemplo de definición de un módulo completo de comandos de voz

### *Contextos y comandos especiales*

Existen en la aplicación dos contextos llamados “root” y “global” los cuales poseen características especiales.

El contexto con nombre “root” es considerado el principal. La aplicación siempre se encuentra inicialmente en ese contexto y, por lo general, es el contexto donde mayor tiempo se encuentra el usuario.

Por otra parte, el contexto “global” en realidad no existe como tal, este nunca va a poder ser cargado. Si un módulo define comandos o traducciones para este contexto, lo que realmente está haciendo es definirlos para todos los contextos en simultáneo, es decir que todos los contextos son extendidos. Este mecanismo es muy útil para definir comandos comunes a toda la aplicación sin importar el contexto actual. Por ejemplo, el comando “ayuda”, “salir”, “cerrar”, son acciones que siempre deberían estar disponibles. Dentro de la ejecución del comando se tiene acceso al estado del contexto actual, según el cual se puede determinar que acción realizar basándose en el contexto cargado. Por ejemplo, si el usuario invoca el comando global “ayuda”, en el código de ejecución del mismo se puede observar cual es el contexto actual y brindar ayudas apropiadas al mismo.

Existe un tipo de comando que es especial, el mismo lleva el nombre de “\*” (asterisco), este comando es ejecutado siempre que no exista un comando que coincida con la frase indicada por el usuario. Por ejemplo, en el módulo útil para agregar un sitio como favorito descrito anteriormente, existe un comando soportado que es “agregar a favoritos”. Puede llegar a ser deseable la capacidad de ejecutar un comando que pueda deshacer esta acción una vez realizada. Para ello solamente habría que agregar el comando “eliminar de favoritos”. Si bien este comando cumple el objetivo, puede que por requerimientos del sistema, sea deseable ejecutar esta acción si el usuario indica una frase que contiene la palabra clave “eliminar” o “deshacer”. Sería imposible definir por extensión las frases que el usuario puede mencionar con alguna de estas palabras. Para esto se suelen utilizar

expresiones regulares aunque, para este caso en particular, una vez obtenido la transcripción del comando de voz indicado, solo basta con una simple búsqueda de texto dentro de la frase y así determinar si el comando “eliminar de favoritos” debe ser ejecutado.

De esta forma, el módulo contará con dos comandos; “agregar a favoritos” y “\*”. La función correspondiente al comando “\*” tendrá acceso a toda frase invocada por el usuario que no coincida con alguno de los comandos definidos, en este caso solamente “agregar a favoritos”. Una vez que la función reciba el texto invocado por el usuario, podrá evaluar si alguna de las palabras claves se encuentran en ella y a continuación proceder con la eliminación del favorito en caso afirmativo.

### **API de comandos de voz nativos del sitio web**

Complementariamente a los módulos personalizados de comandos de voz, la herramienta cuenta con un mecanismo por el cual el sitio web activo puede definir nuevos comandos de voz aptos para ser ejecutados al ser invocados por el usuario. De esta forma se brinda la posibilidad de que un sitio web exponga el soporte a la funcionalidad específica a través de acciones definidas de forma nativa dentro del código del portal.

Cuando un usuario invoca un comando de voz, la búsqueda del mismo se hace sobre los comandos de voz pertenecientes al contexto cargado y la lista de comandos definidos por el sitio web. De la misma manera que los comandos de voz definidos por los módulos personalizados tienen mayor prioridad que los definidos por la aplicación, los comandos brindados por los sitios web tienen prioridad sobre los demás.

La herramienta realiza la búsqueda de comandos disponibles brindados por el sitio web ante la ejecución de cada comando. Dada esta situación, son los sitios webs los responsables de mantener la lista de comandos de voz actualizada y en concordancia con el contexto actual del usuario. Por ejemplo, si el sitio decide brindar diversos comandos de voz correspondientes a acciones de la sección que el usuario está viendo, es de esperarse que al acceder a otra sección, la lista de comandos de voz se actualice para reflejar los nuevos comandos disponibles.

Para proveer comandos de voz nativos, un sitio web necesita definir una lista de comandos invocando la función global llamada “`handsfreeCommands`” disponible una vez la aplicación esté cargada.

Los comandos nativos definidos poseen las siguientes restricciones en comparación con los comandos implementados por un módulo personalizado:

- No poseen acceso al estado del contexto en el cual se encuentra el usuario
- No pueden solicitar cambios de contextos
- La herramienta no brinda servicios de internacionalización, por lo que es deseable que el comando esté definido en el idioma del usuario que por lo general coincide con el del sitio web.

Debido a estas restricciones la estructura de un comando nativo es más reducida que la presentada por un comando correspondiente a un módulo. En la tabla siguiente se describen las propiedades de los comandos nativos.

Propiedad	Tipo propiedad	Tipo de dato	Descripción
name	requerida	string	Indica el nombre del comando
action	requerida	function	Función que será ejecutada cuando el usuario invoca el comando
group	opcional	string	Nombre de grupo al que pertenece el comando. Útil para mostrarle al usuario los comandos divididos por grupos según alguna clasificación.
help	opcional	string	Descripción del comando

A continuación se indica un ejemplo de un sitio web que define un comando de voz que permite al usuario acceder a la sección de contacto del portal.

```

<!DOCTYPE HTML>
<html>
<head>
  <meta http-equiv="Content-Type" content="text/html; charset=UTF-8" />
  <title>Website</title>
</head>
<body>
  <!--
    Page content
  !-->
  <script type="text/javascript">
    window.onload = function() {
      handsfreeCommands([
        {
          name: 'ir a contacto',
          help: 'Visualiza la sección Contacto',
          group: 'Secciones del sitio',
          action: function() {
            window.location.href = '/contacto.html';
          }
        }
      ]);
    };
  </script>
</body>
</html>

```

Sección de Código 5.6: Ejemplo de definición de un comando de voz nativo

## Asistencia al usuario

La aplicación desarrollada plantea una interfaz visual y verbal particular que requiere de ciertas acciones específicas por parte del usuario para que esta cumpla su propósito.

Es este apartado se describen los múltiples recursos implementados que tienen como objetivo inducir al usuario a llevar a cabo las operaciones deseadas de una forma natural.

## Soporte de múltiples idiomas

El primer impedimento significativo que puede surgir entre el usuario y la aplicación es la disparidad de idiomas. Con el objetivo de mitigar este problema se tomaron las siguientes decisiones:

- Idioma de la aplicación por defecto seleccionado automáticamente según el idioma del navegador
- Posibilidad de cambiar el idioma de la aplicación.

En la gran mayoría de los casos el idioma nativo del usuario coincide con el del navegador instalado. Dado que la herramienta se considera una extensión del navegador sería natural compartir la configuración del idioma con la aplicación. Sin embargo, la herramienta propone una interacción verbal que para ser precisa requiere de usuarios nativos de la lengua configurada. Es por esta razón que, independientemente del lenguaje del navegador, es posible seleccionar el idioma de la herramienta.

Al cambiar de idioma todas las interfaces de la aplicación son actualizadas automáticamente según la nueva configuración.

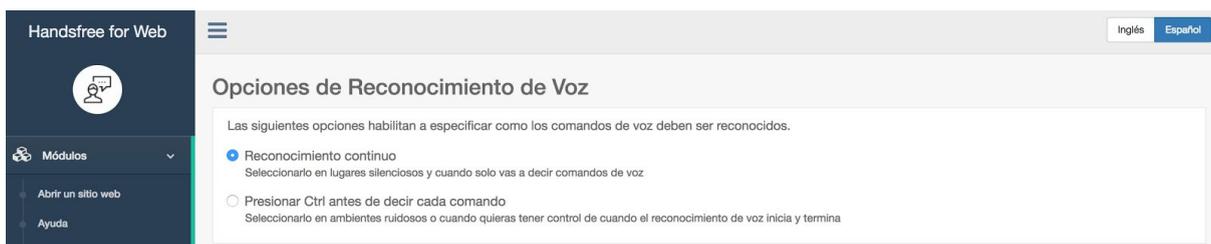


Figura 5.31: Opción de cambio de idioma en panel de administración

## Información de contexto

Al interactuar con la herramienta se brinda información contextual acerca de la actividad del usuario con el objetivo de reafirmar y confirmar que las acciones invocadas están siendo ejecutadas.

A continuación se presenta la información contextual brindada por la aplicación:

- Estado del micrófono: puede estar encendido o apagado.

- Estado del proceso de reconocimiento de voz: se muestra por pantalla los resultados parciales del mismo a medida que el usuario invoca una frase.
- Confirmación de ejecución de comando de voz
- Nombre del contexto actual
- Elemento seleccionado en la pantalla si lo hubiera
- Cuadros de diálogo y marcadores exhibidos luego de la ejecución de los comandos que requieran sucesivos comandos

Para ver más información acerca de los distintos componentes visuales ver la subsección “Interfaz Visual” correspondiente a la sección “Diseño de la herramienta” en este mismo capítulo.

## Comandos disponibles

La resolución de comandos disponibles se basa en las acciones definidas para el contexto actual. Estos comandos provienen de tres diferentes fuentes:

1. Comandos nativos definidos por el sitio web
2. Comandos provistos por módulos personalizados
3. Comandos definidos en la aplicación

En ciertos contextos de ejecución puede haber una gran cantidad de comandos disponibles para ser ejecutados. La herramienta provee tres distintos tipos de componentes útiles para demostrar al usuario los comandos que puede ejecutar.

### Cuadro de diálogo ayuda

Al invocar el comando “ayuda”, independientemente del contexto en el que se encuentre el usuario, se visualiza por pantalla un cuadro de diálogo que contiene la lista de comandos disponibles a ser ejecutados en el contexto actual y la descripción de los mismos.

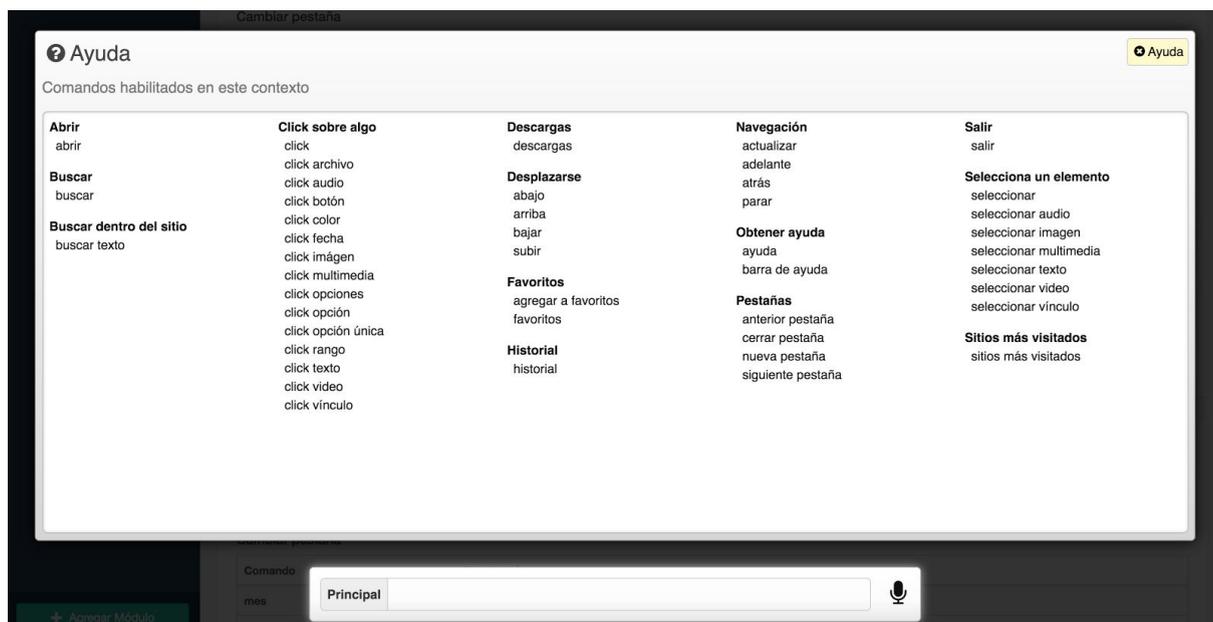


Figura 5.32: Sección de ayuda

## Barra de ayuda

Los usuarios que tienen su primera experiencia con la aplicación y que empiezan a interiorizarse con la gama de comandos de voz disponibles, pueden encontrar conveniente la posibilidad de visualizar de forma constante una lista de comandos habilitados para su ejecución.

Al invocar el comando “barra de ayuda”, independientemente del contexto actual, se visualiza por pantalla la lista de comandos disponibles en una barra lateral. Este componente persiste luego de la ejecución de sucesivos comandos de voz en el sitio actual hasta que el usuario invoque nuevamente el comando “barra de ayuda”.

En esta situación el comando “salir” puede llegar a estar reservado por una sesión iniciada por otro comando mientras la barra de ayuda es visible. Debido a esta circunstancia, utilizar el comando “salir” para ocultar la barra puede llegar a ser contraproducente para las intenciones del usuario.

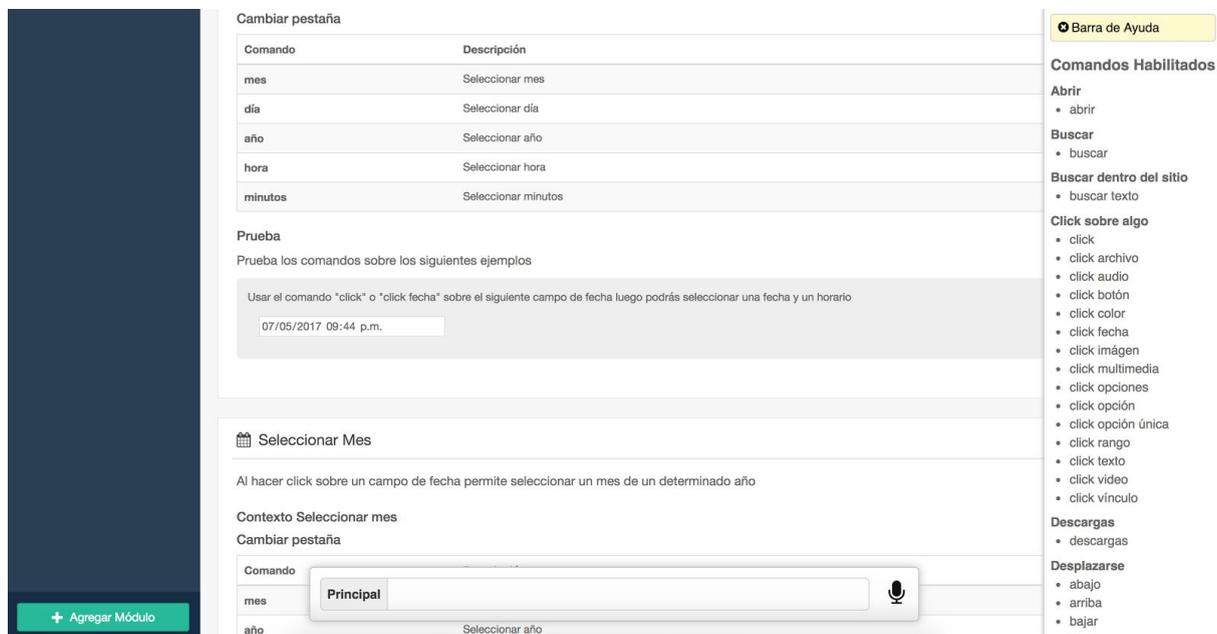


Figura 5.33: Barra de ayuda

## Marcadores

Usualmente los marcadores son presentados en la pantalla para indicar los posibles comandos de voz que el usuario puede llegar a desear invocar para cumplir una función.

Todo texto contenido dentro de un marcador corresponde a un comando de voz disponible. Es necesario tener en cuenta que por lo general puede haber más comandos disponibles que no son exhibidos en forma de un marcador. Esta limitación por lo general proviene del espacio disponible para posicionar los marcadores en la pantalla.

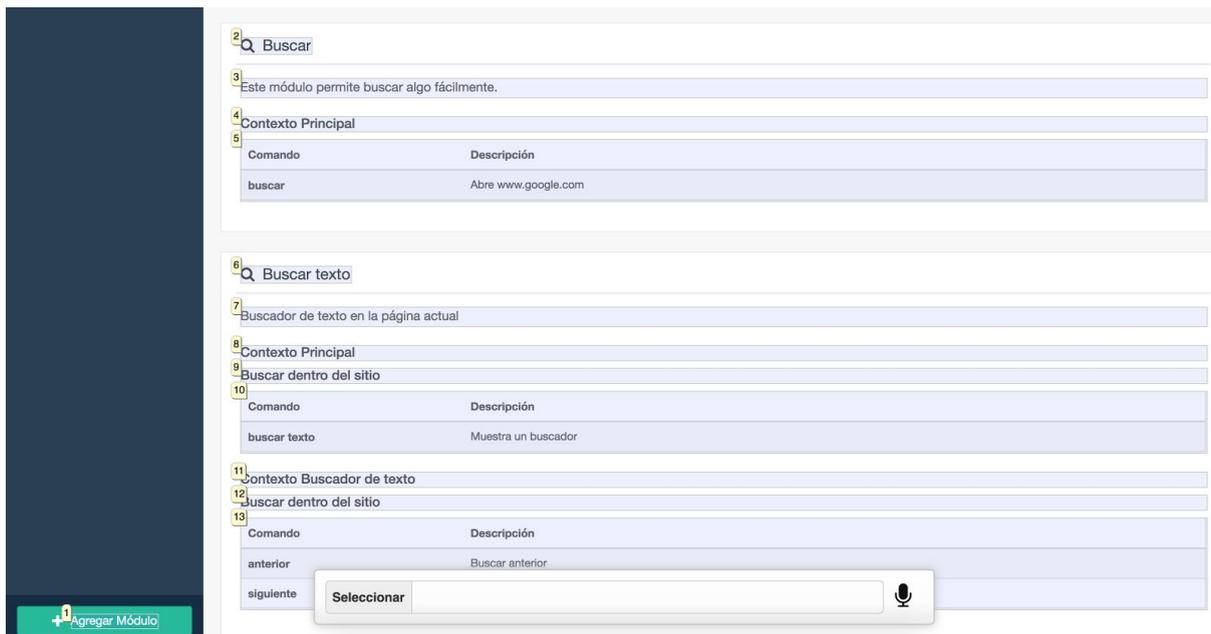


Figura 5.33: Marcadores de elementos seleccionables

## Guía de módulos y comandos instalados

Con el objetivo de facilitar el aprendizaje de los comandos disponibles para interactuar con los diferentes escenarios que presenta la navegación web, se incluye dentro del panel de administración de la aplicación, una descripción detallada de todos los módulos instalados en la herramienta.

Cada módulo descrito presenta información básica acerca del mismo como título y descripción, junto con la especificación de los contextos y comandos de voz que soporta. Opcionalmente los módulos pueden indicar una sección que presenta diversos componentes donde el usuario puede probar los comandos de voz instalados.

Para ver más información acerca del panel de administración consulte la subsección “Panel de administración” correspondiente a la sección “Diseño de la herramienta” en este mismo capítulo.

## Documentación técnica

La implementación de módulos personalizados de comandos de voz, como así también la especificación de comandos nativos por parte de los sitios web, requiere de documentación técnica que debe estar al alcance de la comunidad de forma pública.

Una descripción detallada acerca de cómo implementar módulos y comandos nativos se incluye en la sección Documentación del sitio web [www.handsfreeforweb.com](http://www.handsfreeforweb.com)

## Cuestiones de seguridad

El uso no cuidadoso de la herramienta desarrollada puede producir una brecha de seguridad significativamente importante. La misma es capaz de exponer información

sensible a los atacantes tanto correspondiente a los sitios web como también al entorno físico donde el usuario utiliza la aplicación.

A continuación, se describen los aspectos de seguridad relevantes para la aplicación según los distintos tipos de componentes que pueden llegar a ser expuestos a vulnerabilidades.

## Reconocimiento de Voz

Las interfaces de voz plantean nuevas cuestiones de seguridad a tener en cuenta. En particular a la herramienta, debido a que esta no brinda respuestas verbales, es necesario solo analizar el proceso de reconocimiento de voz.

La especificación de la Api de Reconocimiento de Voz de HTML 5 [18] establece directivas acerca de:

- Cómo el usuario debe consentir el inicio del reconocimiento voz
- Señales visuales que el navegador debe brindar para demostrar que se está capturando sonido utilizando el micrófono

El navegador Google Chrome al cargar la extensión por primera vez solicita por única vez el permiso de captura de sonido, el cual el usuario debe aceptar para poder utilizar la aplicación.

Siguiendo las directivas de la especificación, el navegador muestra un indicador en la pestaña que está en proceso activo de capturar sonidos.



Figura 5.33: Indicador de captura de sonido en pestaña del navegador

Como complemento a las consideraciones tomadas por el navegador, la herramienta realiza el proceso de reconocimiento de voz en la pestaña correspondiente al Panel de Administración por lo que en primera instancia el mismo debe ser abierto para que la captura de sonido sea efectuada.

Dependiendo del modo de reconocimiento de voz seleccionado por el usuario, la captura de sonido puede ser constante o discontinua. El primero sucede cuando el modo seleccionado es el llamado "Reconocimiento continuo". En caso contrario cuando se opta por la opción "Capturar el comando de voz luego de presionar una tecla", la captura de sonido se realiza en forma interrumpida cada vez que el usuario lo desea.

Una vez capturado el sonido, el mismo es enviado a los servidores remotos de Google a través de un canal seguro encriptado para luego recibir la transcripción del comando por otro canal seguro. Dado que no se tiene visualización acerca de qué sucede con los datos

que forman parte del proceso de reconocimiento de voz en el servidor remoto, esto plantea un gran interrogante acerca del destino de ellos y por ello un problema de seguridad.

La Api de Reconocimiento de Voz permite indicar otro servicio de reconocimiento que no sea el asignado por defecto, sin embargo hasta el momento la herramienta no permite este cambio de configuración por parte del usuario. Ver capítulo ocho, Trabajos futuros, sección Mejoras al reconocimiento de comandos de voz.

### **Permisos solicitados por la herramienta**

Con el fin de dar soporte a funcionalidad propia del navegador web y de los sitios web, la herramienta requiere el acceso a los siguientes datos correspondientes a la sesión del usuario:

- Contenido de todos los sitios web a los cuales acceda el usuario
- Gestión de Favoritos
- Gestión del portapapeles
- Lista de descargas realizadas
- Gestión del Historial
- Gestión de Pestañas
- Lista de sitios de acceso más frecuente

Si bien la lista es de reducida extensión, prácticamente la misma brinda el acceso a toda actividad realizada por el usuario. Por esta razón, en caso de haber una vulnerabilidad, el acceso no autorizado a esta información supondría un grave problema de seguridad.

### **Módulos personalizados**

Los comandos de voz definidos por un módulo personalizado poseen el mismo nivel de acceso a información que aquellos que vienen soportados por defecto dentro de la aplicación. Por esta razón es requisito fundamental que antes de importar un módulo se audite el código del mismo o se verifique que éste provenga de una fuente confiable.

### **Comandos nativos de un sitio web**

En oposición a los comandos de voz incluidos o importados dentro de la aplicación los comandos nativos no tienen acceso a información de la aplicación o del contexto de ejecución, incluso tampoco tienen acceso a la interfaz visual de la herramienta. Un comando de voz nativo se ejecuta en el contexto de ejecución del sitio web al igual que todas las funciones definidas en el portal, estas no tienen acceso o conocimiento del estado de la herramienta.

# Capítulo 6

## Pruebas de usabilidad

Dada la naturaleza de uso general que propone la aplicación desarrollada, se consideró de gran importancia evaluar la usabilidad de la misma cuando es operada por los distintos tipos de usuarios.

### Objetivos de la prueba

Las diversas características de la herramienta desarrollada enunciadas a lo largo de este trabajo, motivaron la elección de los siguientes objetivos que se desean validar:

- Evaluar el grado de complejidad que presenta la instalación de la herramienta
- Analizar la comprensión del usuario acerca de la utilidad de la herramienta
- Confirmar la elección satisfactoria de la plataforma en la que se ejecuta la aplicación
- Confirmar el correcto funcionamiento de la herramienta en los lenguajes soportados
- Obtener el grado de acierto correspondiente al reconocimiento de comandos de voz según los distintos oradores
- Determinar el nivel de satisfacción que presentan los usuarios con respecto a la velocidad de interacción que propone la aplicación
- Confirmar si los usuarios presentan interés acerca de los mecanismos por los cuales se puede extender la aplicación
- Confirmar si los comandos de voz provistos por la aplicación satisfacen las necesidades de navegación requeridas por los usuarios
- Evaluar el nivel de aceptación por parte del usuario de la información contextual brindada
- Validar si la herramienta puede ser una solución a los problemas de interacción manual
- Determinar el grado de satisfacción del uso de la aplicación por parte de los usuarios

### Evaluación

Con el fin de satisfacer los objetivos planteados para la prueba, se confeccionó una evaluación compuesta por las siguientes cuatro etapas:

1. Descripción acerca del propósito y uso de la aplicación
2. Obtención del perfil del usuario
3. Instalación y prueba de la herramienta
4. Preguntas acerca de la experiencia de usuario

Con el propósito de que esta sección no sea demasiado extensa, no se presenta la evaluación completa, en cambio se brinda una descripción de las etapas de evaluación. Para más información consultar en el Anexo 3 la evaluación completa realizada.

## **Descripción acerca del propósito y funcionamiento de la aplicación**

Con el objetivo de introducir a la persona en la temática de la herramienta y reducir las posibles dudas acerca de la evaluación, inicialmente se le describe la problemática de interacción manual que presenta la navegación actual y la alternativa propuesta por herramienta.

## **Obtención de perfil del usuario**

Una vez que la persona que va a ser evaluada se muestra interiorizada acerca de los objetivos que persigue la herramienta y como está propone llevarlos a cabo, se procede a realizar preguntas personales con el fin de obtener el perfil de usuario que la misma presenta.

En esta etapa, preguntas como las mencionadas a continuación son enunciadas con el fin de conocer más acerca de la relación que tiene el entrevistado con internet.

- ¿Con que fin accede regularmente a internet?
- ¿En algún momento percibió limitaciones o restricciones físicas a la hora de navegar la web?

## **Instalación y prueba de la herramienta**

La tercera etapa consiste en solicitarle a la persona que utilice activamente la herramienta, para ello es necesario que el entrevistado realice los pasos habituales que todo futuro usuario de la aplicación deberá realizar:

1. Instalar la aplicación
2. Modificar algún aspecto de la configuración por defecto en caso de ser necesario
3. Navegar entre diferentes sitios web utilizando comandos de voz
4. Consultar datos de la sesión del usuario en el navegador tales como favoritos, historial, descargas
5. Interactuar con la gran mayoría de los componentes definidos por el estándar HTML y que dan forma a la web actual

## **Evaluación de la experiencia de usuario**

Finalmente en esta última etapa se consulta al usuario acerca de las múltiples características de la herramienta, y la experiencia de uso percibida.

A continuación se muestran algunas de las preguntas correspondientes a esta etapa:

- ¿Cómo calificaría la ayuda brindada por la herramienta a la hora de navegar la web?
- ¿Cómo consideraría la experiencia de uso de la aplicación?

## **Selección de los participantes**

Con el fin de validar el funcionamiento de la herramienta se seleccionaron personas de múltiples rango etario, género e idioma.

Si bien existen ciertos trastornos físicos que pueden provocarle a una persona dificultades en el habla, como se describió en el capítulo dos, a la hora de invocar comandos de voz por lo general una persona con discapacidad motriz y una sin discapacidad están en igualdad de condiciones. Ambas pueden expresarse de forma oral con total facilidad. Por esta razón se consideró no estrictamente necesario evaluar el desempeño de la aplicación mientras es operada por una persona con discapacidad motriz.

A continuación, se presentan los entrevistados:

<b>Nombre</b>	<b>Edad</b>	<b>Género</b>	<b>Idioma nativo</b>	<b>Idioma secundario</b>
Maitena Perez	8	Femenino	Español	-
Alicia Vicente	60	Femenino	Español	-
Carlos Perez	64	Masculino	Español	-
Belén De Grandis	27	Femenino	Español	Italiano, Inglés
Ivan Parra	23	Masculino	Español	Inglés

Nota: Tanto la voz del entrevistado como la del entrevistador junto con la captura de pantalla durante todo el proceso de evaluación fueron capturados en forma de video. Estas grabaciones se entregan como material adjunto a este documento.

## **Análisis de los resultados**

A lo largo de las pruebas realizadas con los diferentes usuarios se interactuó con los siguientes tópicos que hacen a la navegación web:

- Gestión de pestañas
- Gestión de favoritos
- Gestión del historial
- Acceso a la lista de descargas
- Acceso a la lista de sitios más visitados
- Búsqueda de contenidos en internet
- Apertura de sitios web
- Click sobre elementos que componen los sitios web
- Interacción con textos e imágenes presentes en los diferentes sitios web
- Control de contenido multimedia
- Desplazamientos verticales dentro de los portales
- Interacción con diversos campos de formularios
- Dictado y edición de textos

- Gestión de la navegación

Atendiendo a los objetivos de la prueba propuestos al inicio de este capítulo, a continuación se describen los resultados obtenidos y las conclusiones a las que se llegaron luego de analizar las actividades realizadas por los usuarios.

### **Grado de complejidad que presenta la instalación de la herramienta**

El 100% de los entrevistados, luego de realizada la prueba de la herramienta, manifestaron que el proceso de instalación era muy simple de realizar.

Uno de los criterios de selección, por los cuales se optó por las extensiones de Google Chrome como modelo de aplicación, fue que estas eran de muy simple instalación. Ver capítulo cinco, sección Tecnologías utilizadas.

### **Analizar la comprensión del usuario acerca de la utilidad de la herramienta**

El 100% de los entrevistados demostró ser capaz de utilizar la herramienta. El grado de versatilidad en el uso de la aplicación dependió de su familiarización con el uso de programas informáticos.

Para que un persona use la herramienta de una forma versátil, es necesario que la misma se interiorice en la gama de comandos de voz que pueden ser invocados y dedique tiempo de uso para acostumbrarse al nuevo paradigma de interacción.

### **Confirmar la elección satisfactoria de la plataforma en la que se ejecuta la aplicación**

El 100% de los entrevistados utilizan diariamente Google Chrome. Lo cual sugiere que la elección del navegador de Google como entorno elegido fue adecuado.

### **Confirmar el correcto funcionamiento de la herramienta en los lenguajes soportados**

Durante las pruebas la aplicación se operó en los idiomas inglés y español, en ambos funcionó correctamente.

### **Obtener el grado de acierto correspondiente al reconocimiento de comandos de voz según los distintos oradores**

A continuación, se presenta un análisis de los comandos de voz invocados por los diferentes oradores durante el transcurso de las evaluaciones.

	<b>Maitena</b>	<b>Alicia</b>	<b>Carlos</b>	<b>Belén</b>	<b>Ivan</b>
Idioma	Español	Español	Español	Español	Inglés
Comandos Invocados	56	92	30	38	94
Comandos interpretados correctamente	51	83	27	33	80

Comandos interpretados incorrectamente	5	9	3	5	15
Tasa de acierto	91%	90%	90%	86%	85%

Se puede observar que en promedio la tasa de acierto de comandos de voz invocados en español fue cercana al 90% y 85% en el caso de inglés.

Cuando se realizaron las evaluaciones en español, se notó que ciertas personas tenían problemas a la hora de invocar el comando de voz “click”, probablemente porque esta es una palabra de origen inglés y requiere de cierta pronunciación particular para que sea reconocida correctamente. Para resolver esta situación se definió un alias o sinónimo dentro de la aplicación llamado “presionar”. Las personas que presenten inconvenientes a la hora de invocar la palabra “click” pueden indicar alternativamente “presionar”.

Ante este cambio, se volvió a realizar una prueba de usabilidad con Alicia, quien había sido la más afectada por los inconvenientes a la hora de reconocer el comando “click”. Los resultados fueron totalmente satisfactorios, el 100% de las veces que se mencionó “presionar” en vez de “click” fue reconocido correctamente.

Si omitimos los comandos “click” interpretados incorrectamente de los resultados obtenidos en las evaluaciones y proyectamos los resultados obtenidos con el nuevo alias de comando “presionar”. Obtenemos la siguiente tabla:

	<b>Maitena</b>	<b>Alicia</b>	<b>Carlos</b>	<b>Belén</b>	<b>Ivan</b>
Idioma	Español	Español	Español	Español	Inglés
Comandos Invocados	56	92	30	38	94
Comandos interpretados correctamente	54	92	30	38	80
Comandos interpretados incorrectamente	2	0	0	0	15
Tasa de acierto	96%	100%	100%	100%	85%

Estos últimos resultados indican que la tasa de reconocimiento de comandos de voz en español es cercana al 100% cuando se utiliza “presionar” en vez de “click”.

La tasa de reconocimiento de comandos de voz en inglés puede ser significativamente mejorados si:

- Los comandos de voz son pronunciados por una persona que habla inglés de forma nativa

- Se realizarán ciertas mejoras en la aplicación relacionadas con el reconocimiento de números. Por ejemplo, más de la mitad de los comandos interpretados incorrectamente fueron debido a que la aplicación fue incapaz de relacionar que en inglés “number two” es lo mismo que “2”. O que “two” es igual a “2”.

Una vez resuelto estos problemas la tasa de reconocimiento de comandos de voz expresados en inglés, al igual que en español, estaría cercano a 100%.

### **Determinar el nivel de satisfacción que presentan los usuarios con respecto a la velocidad de interacción que propone la aplicación**

El 100% de los entrevistados manifestó percibir que la aplicación funciona ágilmente y que los resultados son visualizados inmediatamente luego de expresar un comando de voz.

Efectivamente, si se cuenta con una buena conexión a internet, los resultados parciales y finales producto del proceso de reconocimiento de voz, se obtienen en el segundo posterior a ser el comando de voz invocado.

### **Confirmar si los usuarios presentan interés acerca de los mecanismos por los cuales se puede extender la aplicación**

Las personas entrevistadas manifestaron interés en ser notificados ante la posibilidad de extender la aplicación con nuevos comandos de voz.

### **Confirmar si los comandos de voz provistos por la aplicación satisfacen las necesidades de navegación requeridas por los usuarios**

La mayoría de los entrevistados pudieron realizar la mayor parte de las actividades deseadas utilizando la herramienta. Algunos, en ciertos casos, se vieron impedidos de desempeñar una acción debido a que era necesario interactuar con algún componente personalizado de un sitio web. Por ejemplo, al querer comprar un pasaje de avión era necesario seleccionar la tarjeta de crédito deseada a través de una lista de selección que no era propia del estándar HTML 5 y por ende la aplicación era incapaz de interactuar con ella.

### **Evaluar el nivel de aceptación por parte del usuario de la información contextual brindada**

Todas las personas evaluadas destacaron la simplicidad de la aplicación y se mostraron conformes con el nivel de ayuda e información contextual brindada por la herramienta. Es importante destacar que ninguno de los entrevistados poseía experiencia previa utilizando la herramienta, por lo que durante la entrevista fueron asistidos por el entrevistador.

### **Validar si la herramienta puede ser una solución a los problemas de interacción manual**

Dado que todos las personas evaluadas pudieron navegar la web y realizar la mayoría de las actividades deseadas utilizando exclusivamente comandos de voz, se concluye que la herramienta puede ser una alternativa válida de interacción manual.

### **Determinar el grado de satisfacción en el uso de la aplicación por parte de los usuarios**

Las personas que tuvieron la oportunidad de navegar la web utilizando la herramienta se mostraron gratamente sorprendidos y al mismo tiempo satisfechos por el uso de la aplicación. Es importante notar que debido a que ninguno de ellos poseía experiencia previa utilizando la herramienta, inicialmente fue necesario asistir en el uso de la extensión, pero luego todos los entrevistados fueron capaces de operar la sin ayuda.

# Capítulo 7

## Difusión de la Herramienta

Dada la naturaleza social de la aplicación desarrollada, se consideró de vital importancia difundir la misma a través de los canales usuales de publicación para este tipo de herramientas.

A continuación, se describen los canales formales de difusión donde se publicó la extensión desarrollada, información sobre el uso de la misma y una descripción de las características más relevantes.

### Chrome Web Store

Google en el año 2010 presentó un repositorio online de aplicaciones web aptas para ser instaladas en el navegador Google Chrome. En la actualidad, este repositorio de acceso público, cuenta con miles de extensiones publicadas por diferentes tipos de organizaciones o desarrolladores con el fin de que estén a disposición del público general.

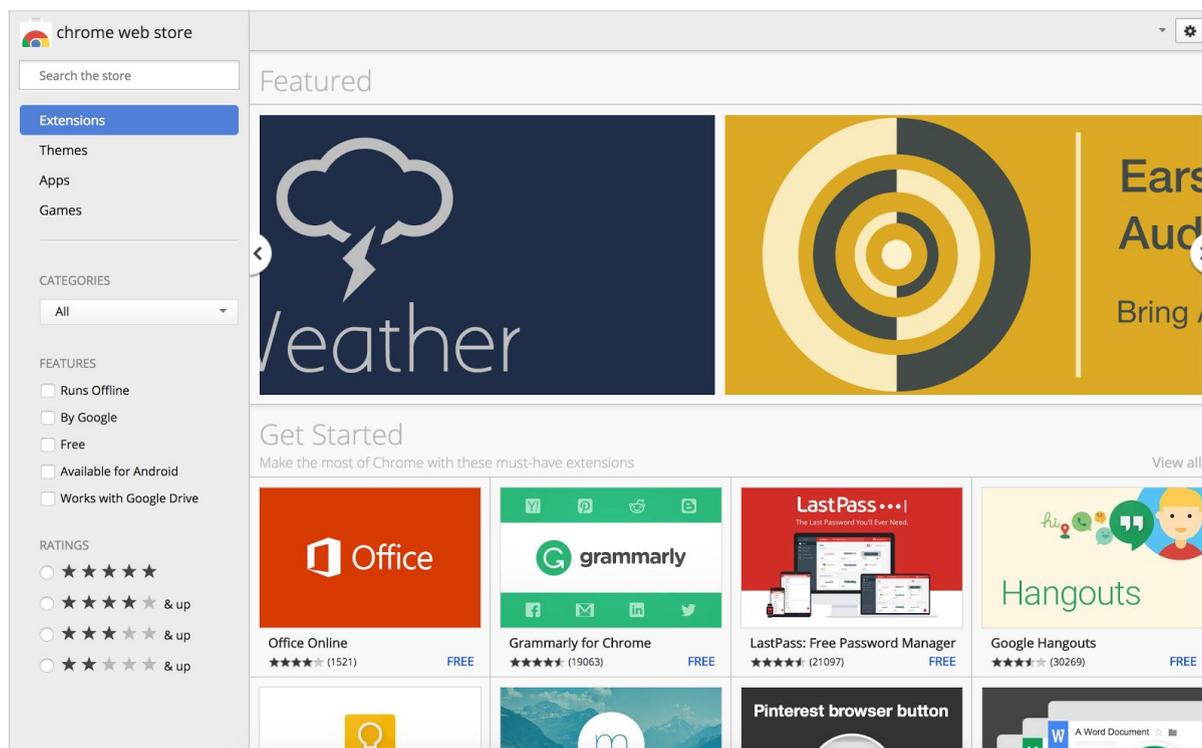


Figura 7.1: Página principal de Chrome Web Store

Dependiendo de las intenciones comerciales del autor de la aplicación, las extensiones publicadas pueden ser de instalación gratuita o paga. Además de brindar servicios de difusión y categorización de las aplicaciones, la web store provee un mecanismo de

instalación extremadamente simple y un sistema de evaluación por parte de los usuarios que permite rápidamente conocer la opinión de los mismos acerca del funcionamiento de la aplicación.

La extensión Handsfree for Web fue publicada en los idiomas inglés y español bajo la categoría accesibilidad. Junto con la posibilidad de instalar la aplicación, se brinda al usuario; un video demostrativo, una descripción acerca de la utilidad de la herramienta, la dirección del sitio web de la aplicación y datos de contacto en caso de que el usuario desee hacer un comentario. Complementariamente, una vez instalada la herramienta, el repositorio provee al usuario la opción de indicar una valoración acerca del funcionamiento de la aplicación.

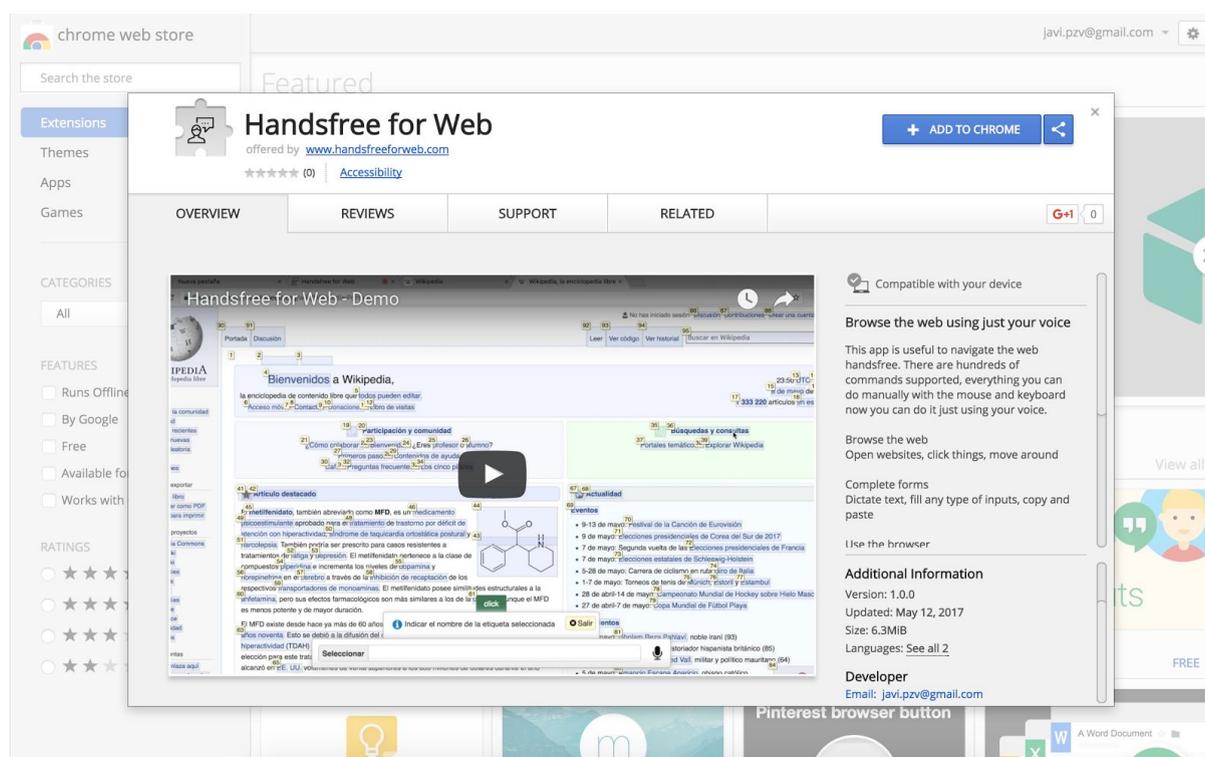


Figura 7.2: Extensión Handsfree for Web en Chrome Web Store

## Sitio web oficial de la aplicación

Con el fin de brindar información oficial acerca de la herramienta, se desarrolló y publicó el sitio web [www.handsfreeforweb.com](http://www.handsfreeforweb.com), donde se deja a disposición del usuario los recursos que se detallan a continuación.

### Demo de la aplicación

Al inicio del sitio web, se presenta un video demostrativo de algunas de las funcionalidades más importantes brindadas por la aplicación. En este video, como ejemplo de navegación, se visitan algunos sitios web, se controla la reproducción de un video, se realizan búsquedas y finalmente se agrega un sitio web a favoritos.

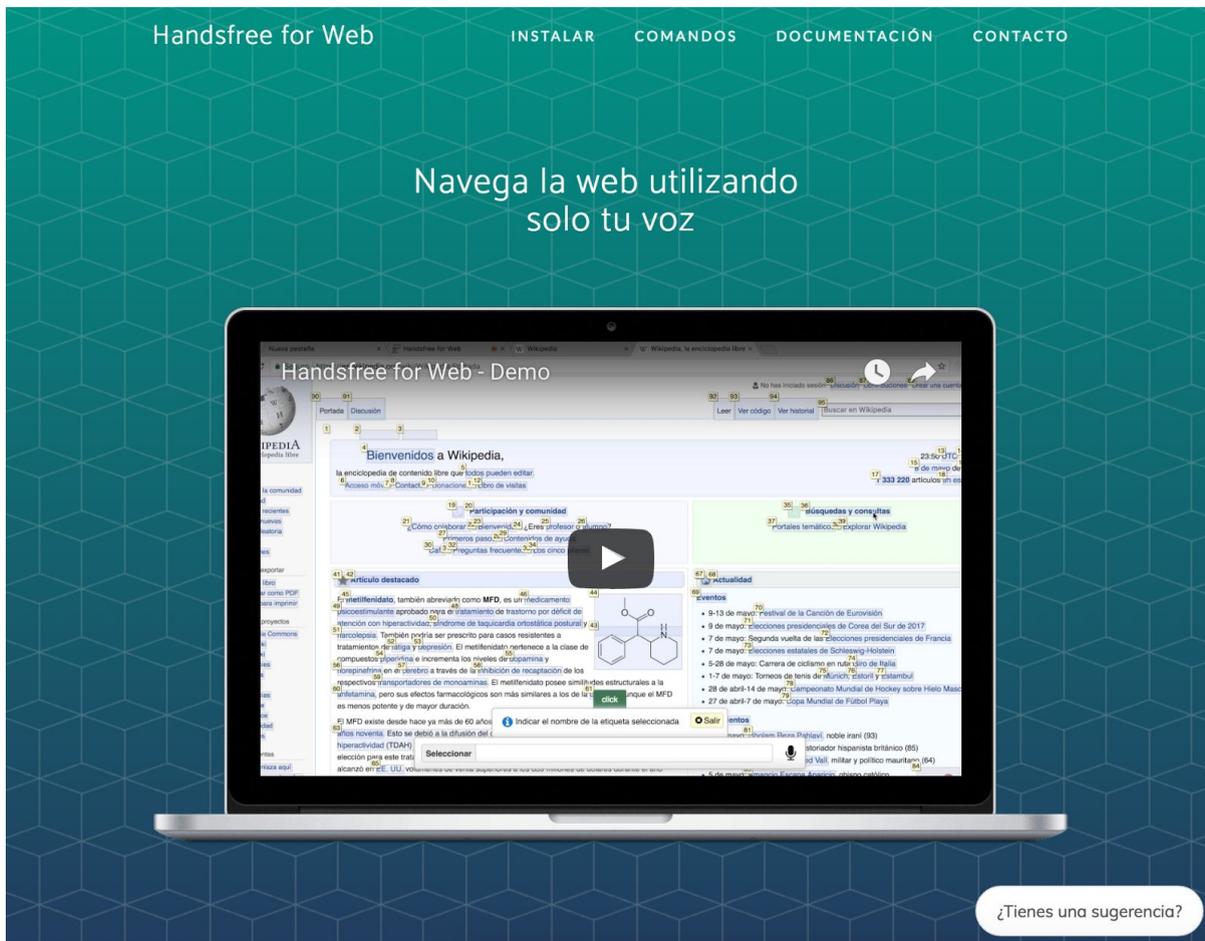


Figura 7.3: Página principal [handsfreeforweb.com](http://handsfreeforweb.com)

### Botón de instalación

Para comodidad del usuario, la instalación de la herramienta puede ser efectuada desde el mismo sitio web, para ello es suficiente con hacer click en el botón “Instalar” y aceptar el cuadro de diálogo donde se indican los permisos del navegador requeridos por la extensión.

Luego de que el usuario acepta la instalación, el navegador se encarga de descargar la aplicación, instalarla y activar la ejecución de la misma sin necesidad de intervención del usuario.



Figura 7.4: Botón de instalación de la herramienta

### Guía de comandos de voz

Junto con la descripción de las funcionalidades principales que soporta la aplicación, se brinda una guía detallada de los comandos de voz que el usuario puede ejecutar con el fin de realizar las distintas acciones que la navegación web requiere.

**Guía de Comandos de Voz**  
Comandos de voz soportados por defecto

- Asistencia al usuario
- Abrir un sitio web
- Buscar contenido en la web
- Hacer click sobre algo
- Seleccionar algo
- Desplazarse dentro de la página
- Control de navegación
- Introducción de texto
- Controlar contenido multimedia
- Gestión de Pestañas
- Gestión de Descargas
- Gestión del Historial
- Gestión de Favoritos
- Acceso a sitios más visitados
- Opciones de Imágen

A continuación se brinda un lista resumida de los principales comandos de voz clasificados según el propósito de los mismos.

**Asistencia al usuario**

En todo momento de la navegación se puede mencionar los siguientes comandos

Comando	Descripción
ayuda	Muestra la lista de comandos de voz disponibles para ser ejecutados
barra de ayuda	Muestra la lista de comandos habilitados en una barra lateral
salir	Sale de la sesión en curso

**Abrir un sitio web**

Para acceder a un sitio web basta con mencioar el comando "abrir" y luego indicar la dirección del ditio web. Por ejemplo: "wikipedia.com"

Comando	Descripción
---------	-------------

¿Tienes una sugerencia?

Figura 7.5: Sección guía de comandos de Voz

### Guía de implementación de comandos de voz

Con el fin de brindar a los usuarios técnicos la posibilidad de crear sus propios comandos de voz, se brinda una sección de documentación donde se desarrollan los siguientes conceptos:

- Comandos de Voz
- Contextos
- Módulos personalizados de comandos de voz

- API de comandos de voz nativos de un sitio web

### Información acerca de esta tesis

En el pie de página del sitio web se puede encontrar un vínculo a este documento e información acerca del autor y el contexto académico en el cual se desarrolló la tesis.

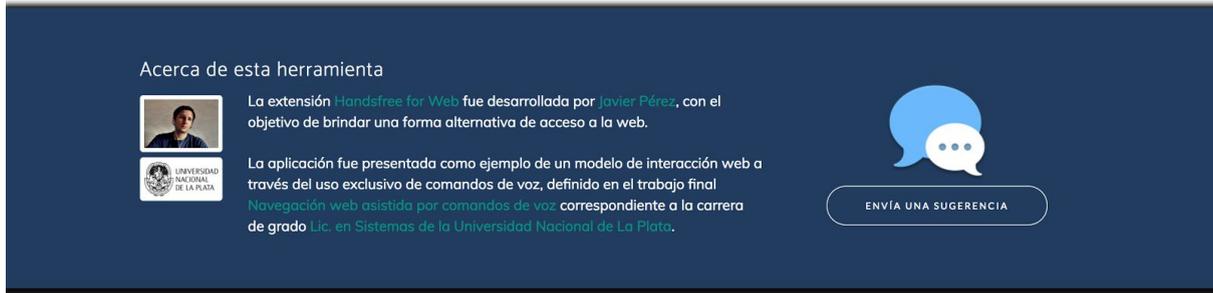


Figura 7.6: Información acerca de esta tesis

### Buzón de sugerencias

Finalmente se provee un sistema que permite al usuario enviar una sugerencia o entrar en contacto con el administrador del sitio web.

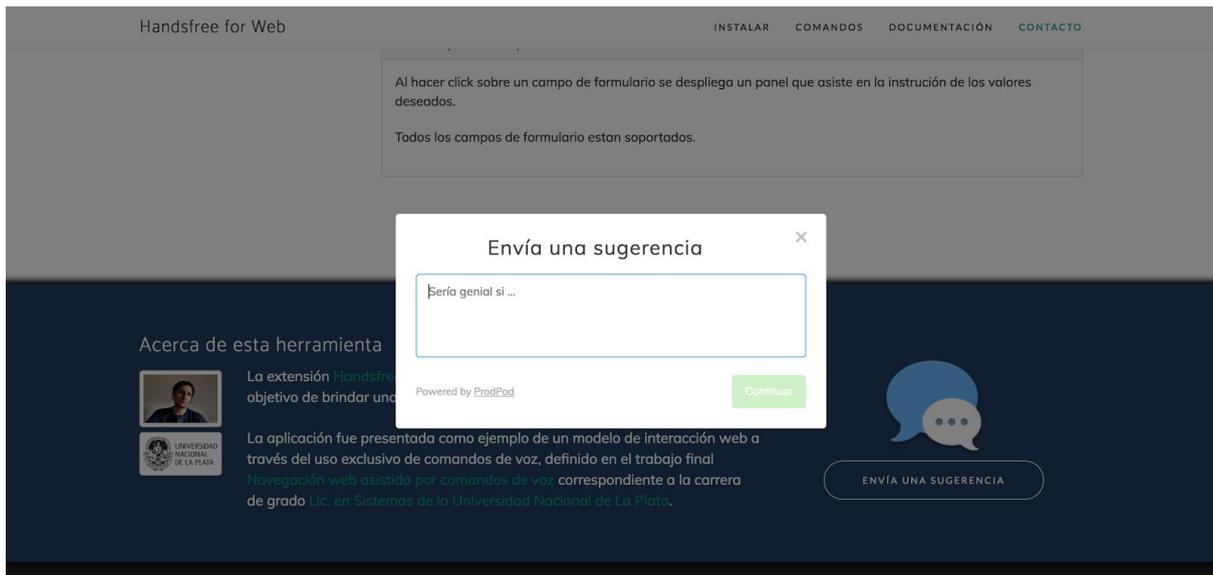


Figura 7.7: Buzón de sugerencias

# Capítulo 8

## Trabajos futuros

En este trabajo de tesis se han abordado diversos temas relacionados con la navegación web dirigida por comandos de voz en contraposición a una interacción convencional. Se han provisto modelos de navegación a través del uso de la voz y una herramienta que permite al usuario interactuar con los diferentes tipos de sitios web invocando comandos. La diversidad de tópicos involucrados brinda la posibilidad de enunciar una gran variedad de líneas de trabajos futuras a desarrollar. A continuación, se enumeran las que se consideraron más relevantes o de mayor interés.

### Portaciones de la herramienta a otras plataformas

La aplicación Handsfree for Web, como se especificó al inicio del capítulo cinco, está disponible exclusivamente para la edición de escritorio de Google Chrome. Si bien el navegador elegido es el más popular y se puede instalar de forma gratuita en todos los sistemas operativos de uso masivo, la plataforma elegida no es la preferida por los usuarios a la hora de navegar. Desde el año 2016 la navegación web en dispositivos móviles superó a la realizada desde computadoras de escritorio [24] y todo indica que con el tiempo esta tendencia se seguirá profundizando.

Debido al uso masivo de internet desde dispositivos móviles, resulta de sumo interés portar la herramienta a este tipo de plataformas. Es importante destacar que las personas con discapacidades motoras carecen de ayudas técnicas versátiles para este tipo de plataforma, lo cual hace aún más necesaria la adaptación.

Técnicamente hablando, portar la herramienta a dispositivos móviles es todo un desafío, debido a que las principales tecnologías de las cuales se sirve la aplicación desarrollada no están disponibles en estos dispositivos. Por ejemplo, la API de reconocimiento de voz del estándar HTML 5 es soportada solo por Google Chrome Mobile sin embargo, este navegador, no soporta extensiones. Alternativamente Firefox Mobile provee la posibilidad de desarrollar extensiones para este navegador, pero no soporta la posibilidad de hacer reconocimiento de voz [25].

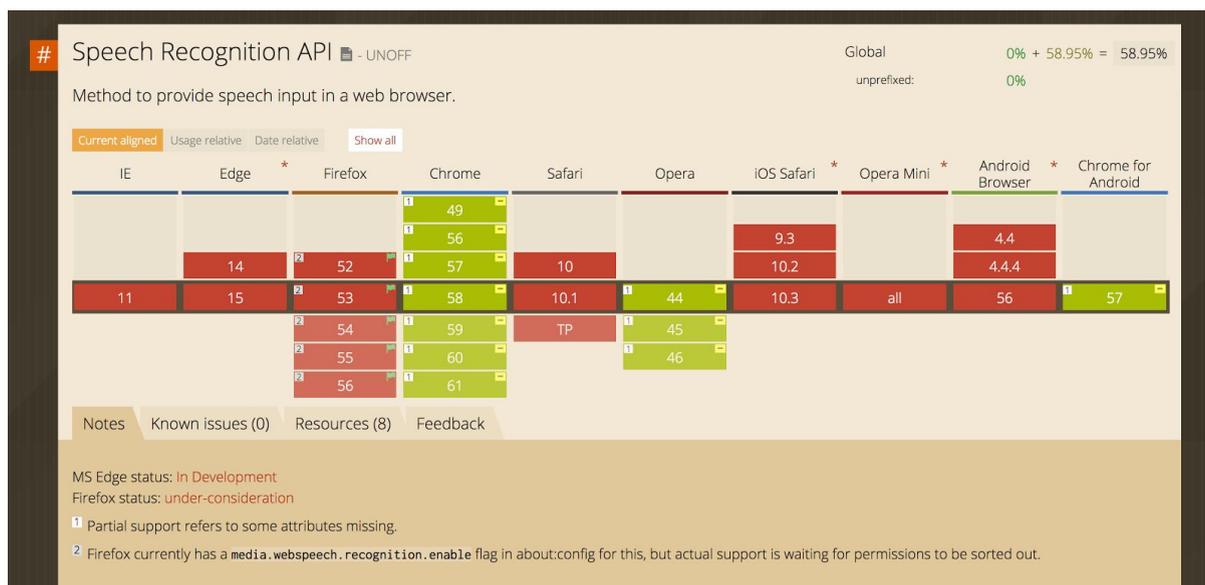


Figura 8.1: Soporte de reconocimiento de voz [25]

## Mejoras al reconocimiento de comandos de voz

En el capítulo cinco, sección “Proceso de ejecución de un comando de voz”, se describió la forma actual en la que se infiere el comando de voz a partir del texto resultante de la expresión de usuario. Se considera que este proceso puede ser mejorado desde los siguientes aspectos:

### Mejorar la captura del sonido

Desde el punto de vista del hardware el medio convencional por el cual se captura la voz del usuario es el micrófono, el cual puede ser interno o externo. Estos micrófonos, al no ser de uso profesional, poseen deficiencias a la hora de capturar el sonido.

Atendiendo al software, el controlador instalado en el sistema operativo suele aplicar filtros con la intención de disminuir el ruido ambiental. Si bien el contexto en el cual el usuario utiliza la herramienta no es un ambiente controlado y estaría fuera del alcance de este trabajo, se podría proveer al mismo un dispositivo especializado de captura de audio con filtros de sonidos apropiados tales como; balance de graves y agudos, reducción de ruido ambiental, y otras normalizaciones. De esta forma la captura específica de la voz de la persona sería más eficiente y como consecuencia de ello se obtendrían mejores resultados a la hora de realizar el reconocimiento de voz.

Para más información ver capítulo tres, sección “Dificultades en el proceso de reconocimiento de voz”

### Mejorar el reconocimiento de voz

La API de reconocimiento de voz definida por el estándar HTML 5 permite la definición de un servicio de reconocimiento diferente al provisto por defecto. En concreto la herramienta

desarrollada como parte de este trabajo, no re-define el servicio, por lo que utiliza el servicio de reconocimiento de Google.

Con el fin de mejorar los resultados del proceso de reconocimiento, se podrían utilizar otros sistemas alternativos que existen en el mercado, ver capítulo tres sección “Sistema de reconocimiento de voz”. En la actualidad existen diversos sistemas de reconocimiento de voz, cada uno de ellos presentan diversas características, las cuales son relevantes a la hora de obtener mejores tiempos de respuesta y tasas de reconocimiento de voz.

### **Mejorar la inferencia del comando de voz**

Una vez obtenida la transcripción resultante del proceso de reconocimiento de voz que dio lugar a la expresión verbal del comando por parte del usuario, es necesario determinar en concreto cual es el comando de voz a ejecutar. Por lo general, la identificación del comando no concuerda en un 100% con la transcripción recibida por el servicio de reconocimiento de voz. Por ejemplo, cuando el usuario dice “click”, se suele obtener la transcripción “clip”. Al intentar determinar el comando correspondiente a “clip” se encuentra que el mismo no existe.

Al no contar con sistemas de reconocimiento de voz con tasas de acierto del 100%, es importante analizar los resultados obtenidos, con el fin de inferir cuál fue la intención del usuario, y a partir de ello ejecutar el comando de voz que tenga una mayor probabilidad de acierto.

La aplicación desarrollada realiza comparaciones fonéticas entre los posibles comandos que pueden ser ejecutados y las transcripciones resultantes de la expresión del usuario, con el fin de determinar los posibles comandos de voz que él mismo quiso ejecutar. El comando de voz de mayor probabilidad de acierto es el elegido, ver capítulo cinco, sub sección “Resolución de comando de voz” correspondiente a la sección “Proceso de ejecución de un comando de voz”.

Si bien la herramienta obtiene buenos resultados a la hora de inferir los comandos, este proceso puede ser mejorado. A continuación, se describen algunas posibles mejoras.

Luego de la ejecución de un comando, el usuario podría indicar si el comando ejecutado fue el deseado. De esta forma, se podrían almacenar los casos fallidos de resolución de comandos con el fin de no volver a reproducir el fallo.

Cada vez que una persona menciona un comando de voz desconocido para la aplicación, se le podría mostrar una lista de comandos de voz disponibles y consultarle qué comando de voz deseó invocar. De esta forma, la próxima vez que el usuario invoque las mismas palabras, se contará con la información necesaria para inferir el comando apropiadamente.

La correcta selección de los comandos de voz es un aspecto importante a tener en cuenta. A la hora de inferir los comandos, se obtienen mejores resultados si la pronunciación de los mismos es totalmente diferente. Por ejemplo, si analizamos el comando “dos”, correspondiente al número 2, es en cierta forma similar al comando “favoritos”. En caso de

visualizar errores en la resolución de comandos, se podrían cambiar alguno de ellos a “número dos” o “lista de favoritos” respectivamente.

## **Pruebas automatizadas de una interfaz de voz**

Tanto la aplicación como los módulos de comandos de voz implementados pueden ser probados automáticamente a través de los llamados tests, los cuales se clasifican según su propósito en; test de unidad, test funcionales y test de integración.

Si bien no se encuentra dentro del alcance de esta sección los diversos aspectos relacionados con la evaluación de las aplicaciones, se propone un nuevo tipo de test de integración.

Haciendo uso de síntesis de voz [19], se pueden invocar los comandos de voz soportados y, posteriormente, evaluar el correcto funcionamiento de la aplicación. De este modo, se puede evaluar en su totalidad automáticamente la interfaz oral implementada por la herramienta.

## **Repositorio de módulos personalizados de comandos de voz**

En el modelo de navegación web a través del uso de comandos de voz desarrollado en el capítulo cuatro, se enunció la necesidad de implementar módulos de comandos de voz que soporten funcionalidades específicas de los sitios web.

Posteriormente, en el capítulo cinco se describen los módulos de comandos de voz que pueden ser instalados en la herramienta, los cuales van más allá de lo propuesto en el modelo y no solamente permiten que los desarrolladores particulares implementen comandos de voz específicos para funcionalidades web, sino que también pueden sobrescribir los comandos provistos por la aplicación.

Resulta de interés proveer un mecanismo por el cual los desarrolladores de módulos puedan dejar a disposición del público en general los comandos implementados y así, de esta forma, luego de la instalación de estos módulos, los usuarios pueden contar con una aplicación personalizada que se adapte mejor a sus necesidades.

## **Mejorar soporte de elementos HTML**

A la hora de implementar la herramienta de navegación web, se hizo un gran esfuerzo por soportar todos los elementos HTML definidos por el estándar. Sin embargo, debido a la naturaleza compleja de uso e integración de los llamados atributos Aria, el soporte para estos por parte de la aplicación no es completo. Una situación similar, y en parte agravada por reglas de seguridad implementadas por el navegador, es la incapacidad de realizar operaciones dentro de un iframe.

## **Handsfree for Web para personas no videntes**

Las personas no videntes navegan la web utilizando una ayuda técnica denominada lector de pantalla [26]. Este software, a través de síntesis de voz, lee los contenidos del sitio web y explica las acciones realizadas al interactuar con el teclado y el mouse. De esta forma, una persona no vidente puede acceder, con ciertas dificultades, a los contenidos presentes en los sitios web.

Distinta es la situación de una persona no vidente que sufre discapacidad motriz. Debido a la incapacidad de utilizar el mouse con precisión o de presionar los atajos de teclado requeridos, estas personas suelen tener impedimentos para utilizar un lector de pantalla.

Handsfree for Web propone una alternativa de interacción web vía comandos de voz en contraposición a la manual. La herramienta hace uso de indicadores visuales para guiar al usuario durante la navegación web. Por ejemplo, luego de invocar el comando “click” es necesario indicar el número correspondiente a la referencia visual del elemento que se desea presionar.

Con el fin de brindar la posibilidad de navegar a personas con discapacidad visual severa, los componentes gráficos presentados por la aplicación pueden ser reemplazados o acompañados con expresiones verbales producidas con síntesis de voz.

De forma complementaria, es necesario proveer funcionalidades propias de un lector de pantalla o asignar comandos de voz que emulen los comportamientos esperados por un lector de pantalla previamente instalado por el usuario.

# Capítulo 9

## Conclusiones

A partir del análisis de interfaces de voz realizado en este trabajo, se concluye que a pesar de existir diversos aplicativos capaces de proveer al usuario la capacidad de controlar la computadora a través del uso de la voz, no brindan la posibilidad de interactuar plenamente con el navegador y los contenidos web como lo puede hacer una persona manualmente.

Con el objetivo de desarrollar una herramienta que permita la navegación web dirigida por comandos de voz, fue necesario inicialmente analizar las operaciones que hacen a esta actividad. Se describió por extensión todas las funcionalidades de usuario provistas por el navegador Google Chrome y los componentes de interacción estándares presentes en los sitios web.

Luego de enunciar y clasificar las acciones manuales que realiza un usuario cuando interactúa con el navegador, se determinó que era necesario desarrollar un modelo navegacional, dirigido por comandos de voz, capaz de brindar un comando o una secuencia de ellos para dar soporte a las funcionalidades implementadas por el navegador y los sitios web.

Al analizar las funcionalidades presentes en los sitios web, se puso en evidencia que, además de la interacción con componentes estándares, el usuario interactúa con componentes específicos propios de cada sitio web en particular. Esta interacción, compuesta de acciones de tipo personalizado, no puede ser descrita en su totalidad por lo que dificulta la posibilidad de definir un conjunto de comandos de voz que le den soporte. Para ello se definieron dos mecanismos por los cuales se puede brindar soporte a las funcionalidades específicas de los sitios web; módulos personalizados de comandos de voz y API de comandos voz nativos definidos por el sitio web.

En concordancia con el objetivo general de este trabajo, y siguiendo el modelo navegacional dirigido por comandos de voz mencionado, se desarrolló una herramienta que permite al usuario navegar la web exclusivamente invocando comandos de voz.

La aplicación desarrollada, además de dar soporte de comandos de voz a las funcionalidades del navegador y estándares de los sitios web, permite la incorporación de módulos personalizados de comandos de voz por parte del usuario y comandos nativos provistos por el sitio web mientras el usuario interactúa con el mismo. En un principio la herramienta se presenta en dos idiomas, inglés y español, pero fue desarrollada para dar soporte a otros idiomas y dialectos sencillamente.

Si bien la extensión desarrollada como parte de este trabajo final permite al usuario un medio alternativo de acceso a los contenidos web, también propone y hace posible un cambio de paradigma de interacción, donde las interfaces web dejan de ser diseñadas para ser operadas manualmente y empiezan a ser parte de una agente inteligente capaz de comunicarse verbalmente.

# Anexo 1

## Soporte de comandos de voz a funcionalidad exclusiva del navegador

A continuación, se define por extensión el conjunto general de funcionalidades soportadas por el navegador Google Chrome, junto con una descripción de las operaciones manuales que debe realizar el usuario y una alternativa de interacción utilizando comandos de voz.

### Gestión de Favoritos

Habitualmente un usuario selecciona una dirección web para guardarla como favorito o marcador y, de esta forma, poder visitarla posteriormente.

#### Guardar un Favorito

La forma usual para denotar el sitio web actual como favorito consiste en hacer click en la estrella que se encuentra a la derecha de la dirección web y luego de seleccionar opcionalmente una carpeta, aceptar la incorporación del nuevo favorito.

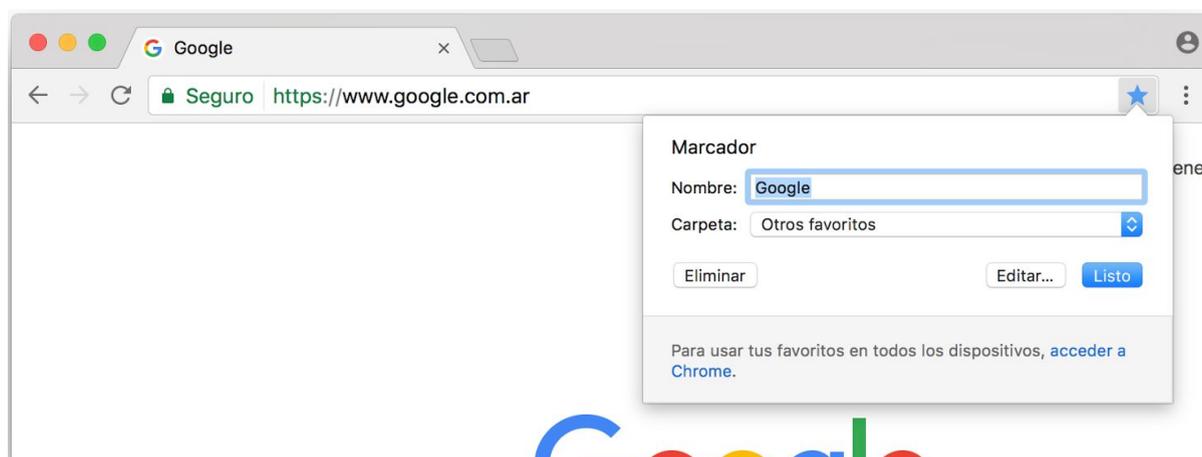


Figura 10.01: Alta de un favorito

Comandos de voz equivalentes a la operación manual:

1. “agregar a favoritos”  
Agrega el sitio web a la lista principal de favoritos
2. “{mencionar el nombre de carpeta existente o nueva carpeta}”  
Cambia la carpeta destino donde será almacenado el favorito

## Ver favoritos guardados

La sección favoritos tiene como objetivo listar todos los sitios guardados por el usuario organizados según las distintas carpetas creadas. Opcionalmente el usuario puede visualizar los favoritos guardados en la barra de favoritos.



Figura 10.02: Barra de favoritos

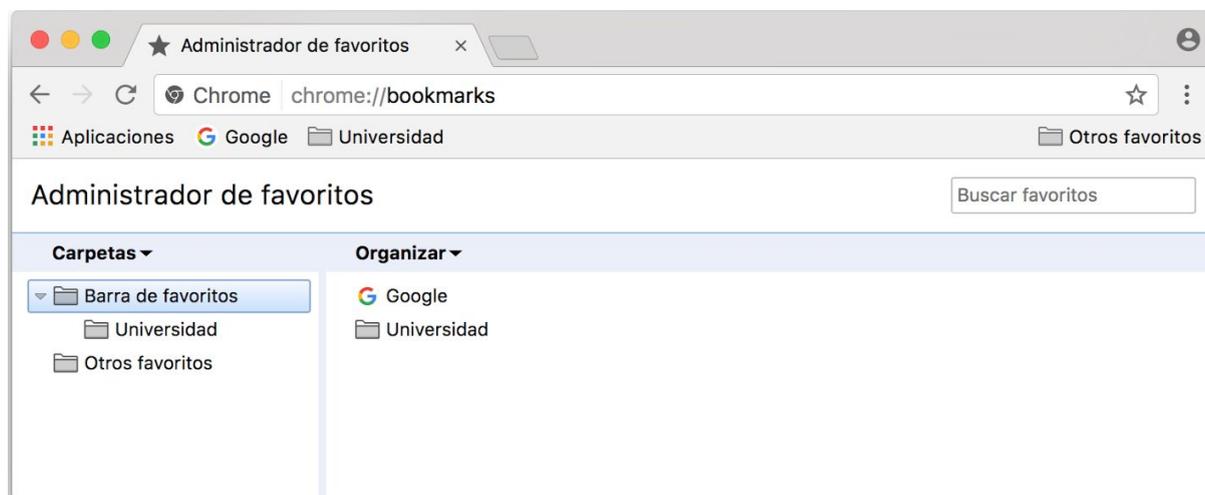


Figura 10.03: Administrador de favoritos

Comandos de voz equivalentes a la operación manual:

- “Favoritos”  
Lista todos los favoritos guardados

## Eliminar un favorito

Una vez localizado un favorito, ya sea en la barra de favoritos o en la sección de favoritos del navegador, el usuario está habilitado a eliminar el mismo si ya no desea conservarlo.

Para ello es necesario hacer doble click sobre el mismo y luego elegir la opción “Eliminar” del menú contextual.

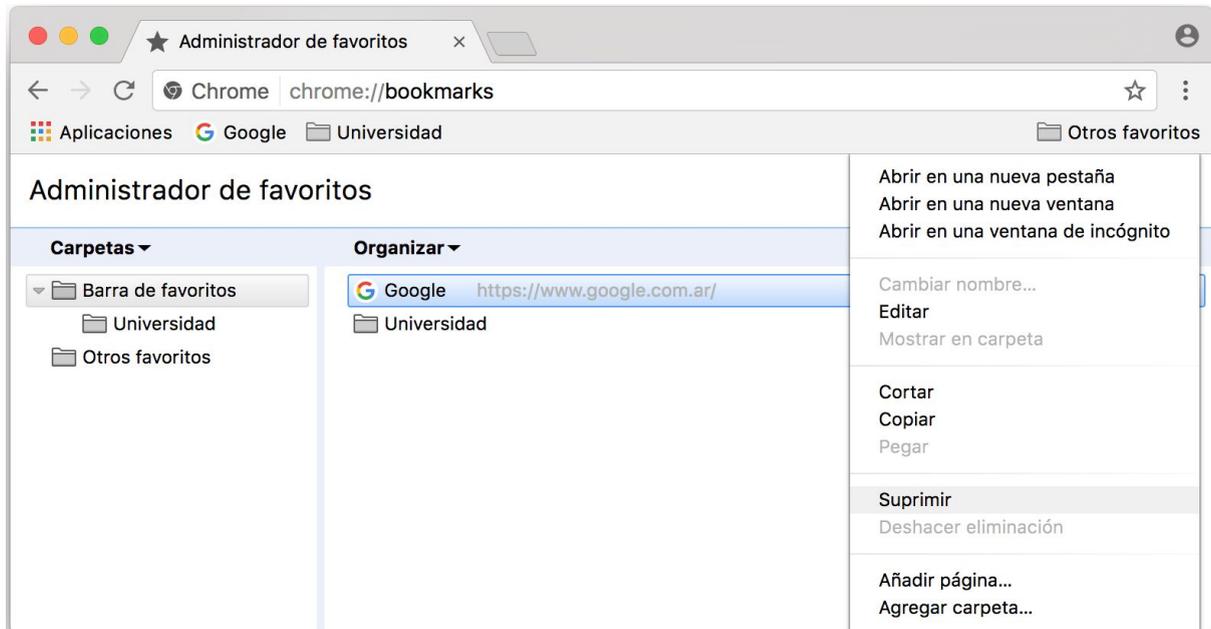


Figura 10.04: Opción eliminar de un favorito

Comandos de voz equivalentes a la operación manual:

1. “Favoritos”  
Lista todos los favoritos guardados
2. “{mencionar referencia correspondiente al botón eliminar del favorito deseado}”  
Indicar el marcador correspondiente al botón eliminar del favorito que se desea eliminar

### Eliminar una carpeta

Una vez localizada una carpeta, ya sea en la barra de favoritos o en la sección de favoritos del navegador, el usuario está habilitado a eliminar la misma si ya no desea conservarla. Para ello es necesario hacer doble click sobre el nombre y luego elegir la opción “Eliminar” del menú contextual.

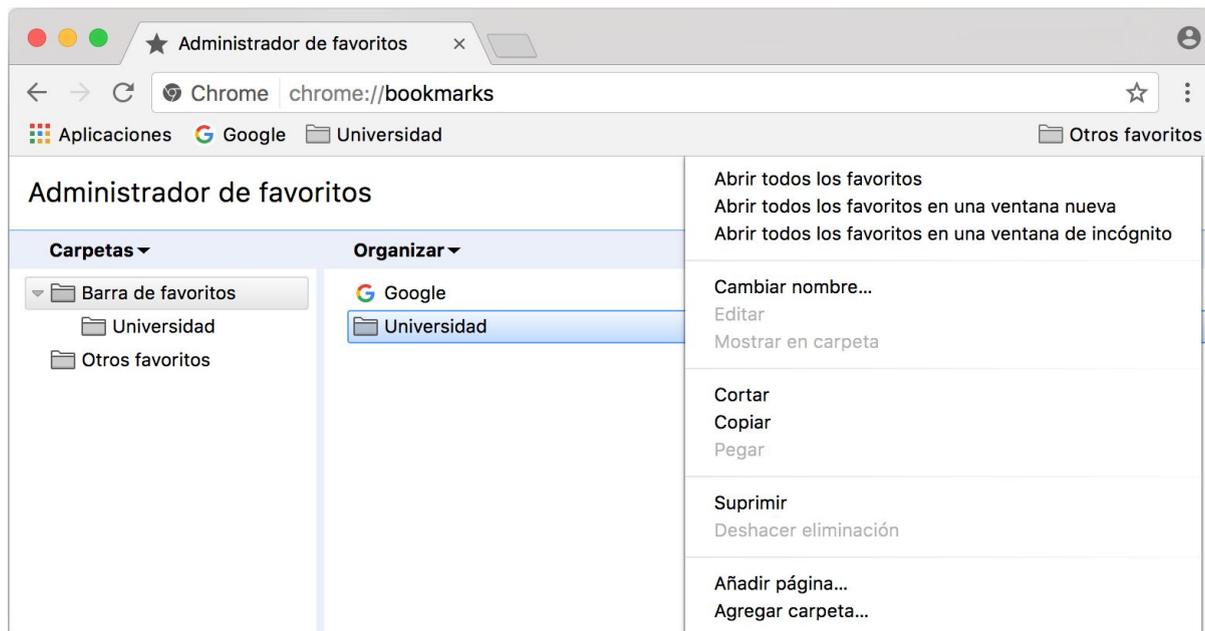


Figura 10.05: Opción eliminar de una carpeta

Comandos de voz equivalentes a la operación manual:

1. “Favoritos”  
Lista todos los favoritos guardados
2. “{mencionar referencia correspondiente a la opción eliminar de la carpeta}”  
Indicar el marcador correspondiente a la opción eliminar de la carpeta que se desea remover

### Mover un favorito de una carpeta a otra

Una vez localizado un favorito el usuario está habilitado a cambiar la ubicación del mismo a otra carpeta, para ello es necesario arrastrarlo usando el mouse hasta la carpeta destino.

Comandos de voz equivalente a la operación manual:

1. “Favoritos”  
Lista todos los favoritos guardados
2. “{mencionar referencia correspondiente a la opción mover del favorito deseado}”  
Indicar el marcador de la opción mover correspondiente al favorito que se desea mover
3. “{mencionar nombre de carpeta existente o de nueva carpeta}”  
Indica la nueva carpeta destino

### Abrir un favorito

Una vez localizado un favorito, el usuario por lo general deberá abrir el mismo, para esto basta con hacer click sobre el mismo.

Comandos de voz equivalentes a la operación manual:

1. “Favoritos”  
Lista todos los favoritos guardados
2. “{mencionar referencia correspondiente al favorito}”  
Indicar el marcador correspondiente al favorito que se desea abrir

## Buscar favoritos

En ciertas ocasiones el usuario puede requerir realizar una búsqueda de favoritos por nombre dentro del panel de administración de la sección de marcadores.

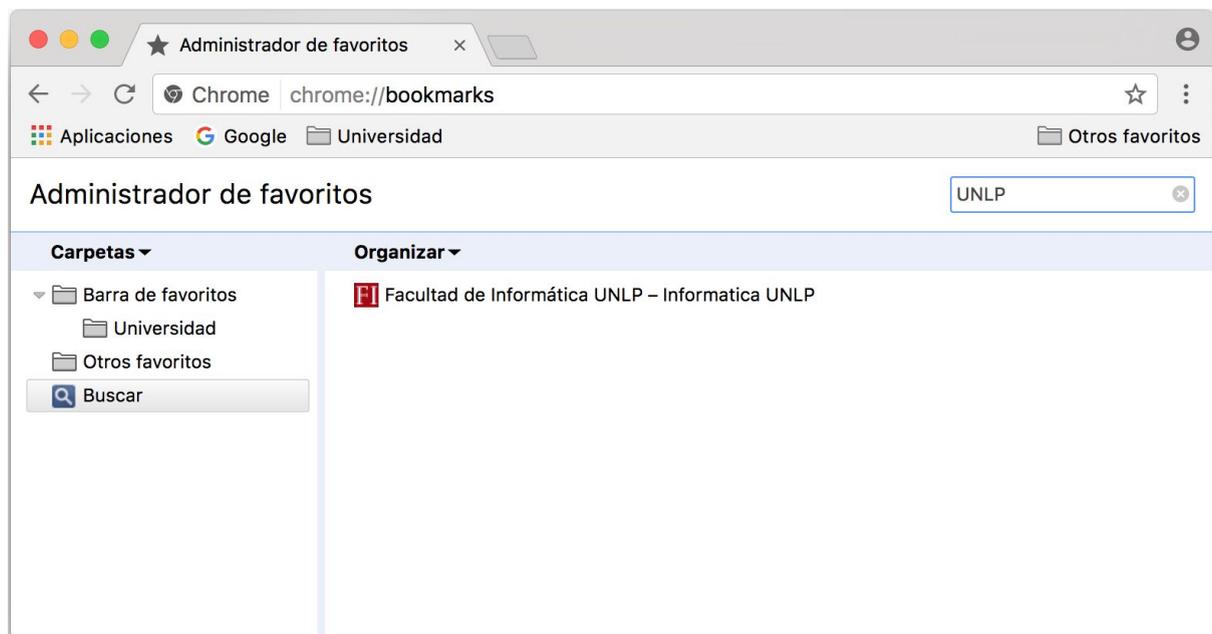


Figura 10.06: Campo de búsqueda en Administrador de Favoritos

Comandos de voz equivalentes a la operación manual:

1. “Favoritos”  
Lista todos los favoritos guardados
2. “Buscar”  
Permite dictar un parámetro de búsqueda
3. “{mencionar nombre de favorito}”  
Al mencionar el nombre del favorito deseado lista los resultados de la búsqueda

## Gestión de Descargas

Es habitual que los usuarios descarguen contenidos mientras navegan por internet. Actualmente Google Chrome provee una sección de administración de descargas realizadas.

## Ver archivos descargados

La sección descargas tiene como objetivo listar todos los archivos descargados por el usuario en orden cronológico.

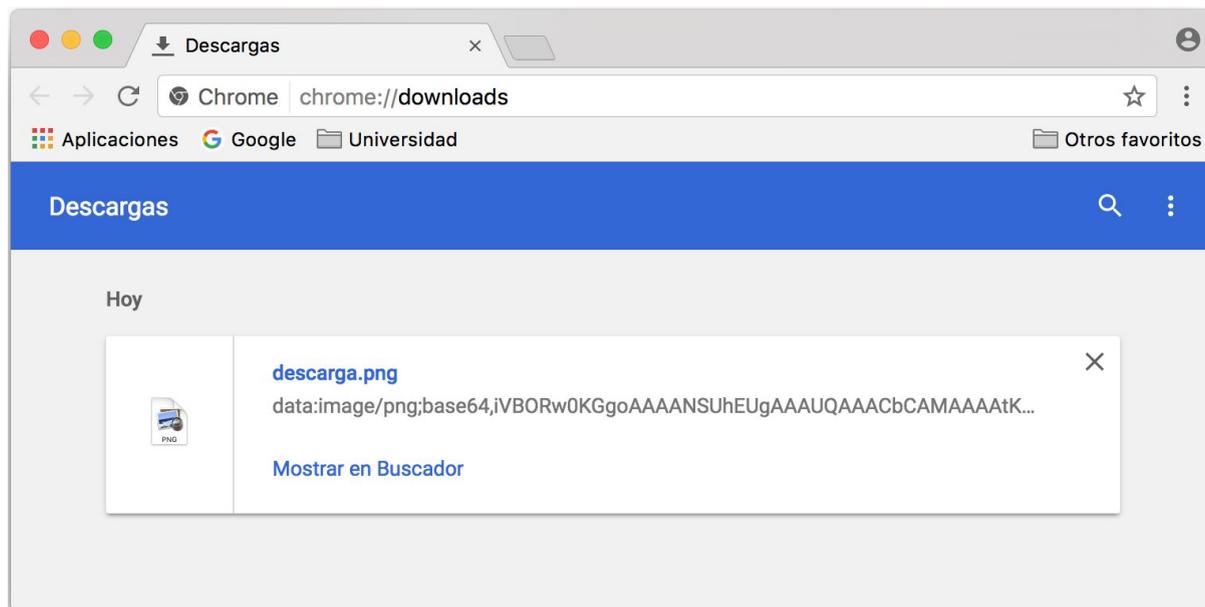


Figura 10.07: Sección de Descargas

Comandos de voz equivalentes a la operación manual:

- “Descargas”  
Lista todos los archivos descargados

## Abrir un archivo descargado

Una vez localizado un archivo descargado el usuario puede desear ejecutar el mismo desde el navegador, para ello es necesario hacer click sobre el nombre del mismo mismo.

Comandos de voz equivalentes a la operación manual:

1. “Descargas”  
Lista todos los archivos descargados
2. “{mencionar referencia correspondiente al archivo}”  
Indicar el marcador correspondiente al archivo que se desea abrir

## Eliminar un archivo descargado

Una vez localizado el archivo que se desea remover de la lista de descargas, es necesario hacer click sobre la opción eliminar correspondiente al mismo.

Comandos de voz equivalentes a la operación manual:

1. “Descargas”  
Lista todos los archivos descargados
2. “{mencionar referencia correspondiente a la opción eliminar del archivo}”  
Indicar el marcador correspondiente a la opción eliminar del archivo que se desea remover de la lista

## **Buscar archivos descargados**

En ciertas ocasiones el usuario puede requerir realizar una búsqueda de archivos por nombre dentro del panel de descargas.

Comandos de voz equivalentes a la operación manual:

1. “Archivos”  
Lista todos los archivos descargados
2. “Buscar”  
Permite dictar un parámetro de búsqueda
3. “{mencionar nombre del archivo}”  
Al mencionar el nombre del archivo deseado lista los resultados de la búsqueda

## **Gestión de Historial de Navegación**

Habitualmente los usuarios revisan el historial de navegación con el fin de buscar un sitio web visitado con anterioridad.

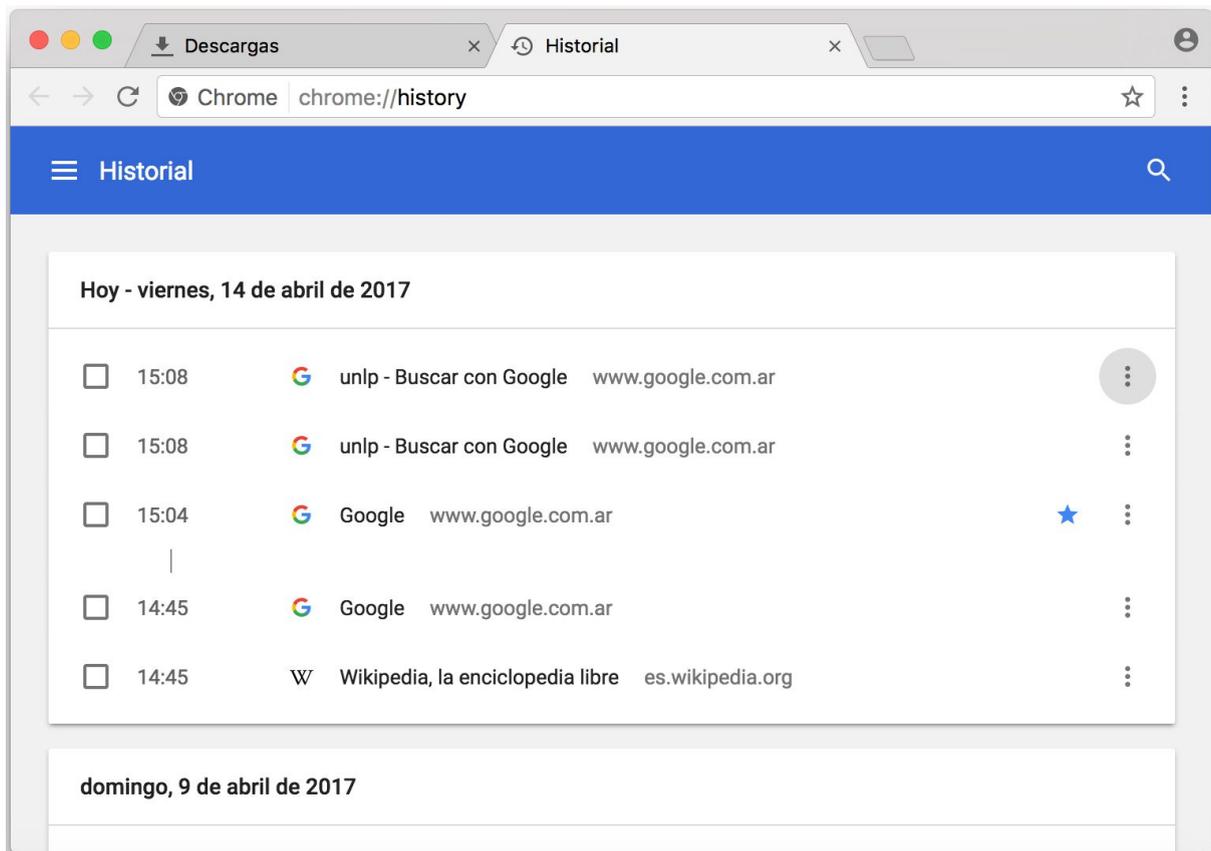


Figura 10.08: Sección Historial

## Ver sitios web visitados

La sección Historial tiene como objetivo listar todos los sitios visitados por el usuario en orden cronológico.

Comandos de voz equivalentes a la operación manual:

- “Historial”  
Lista todos los sitios visitados

## Abrir un sitio visitado

Una vez localizada la dirección correspondiente al sitio visitado, el usuario puede desear abrir nuevamente el mismo, para ello es necesario hacer click sobre el nombre de la entrada de navegación.

Comandos de voz equivalentes a la operación manual:

1. “Historial”  
Lista todos los sitios visitados
2. “{mencionar referencia correspondiente al sitio visitado}”  
Indicar el marcador correspondiente al sitio que se desea abrir

## Eliminar un sitio visitado

Una vez localizada la entrada de navegación que se desea remover de la lista de sitios visitados, es necesario hacer click sobre la opción eliminar correspondiente al mismo.

Comandos de voz equivalentes a la operación manual:

1. “Historia”  
Lista todos los sitios visitados
2. “{mencionar referencia correspondiente a la opción eliminar de la entrada de navegación}”  
Indicar el marcador correspondiente a la opción eliminar de la entrada de navegación que se desea remover de la lista

## Buscar sitios visitados

En ciertas ocasiones el usuario puede requerir realizar una búsqueda en el historial de navegación.

Comandos de voz equivalentes a la operación manual:

1. “Historial”  
Lista todos los sitios visitados
2. “Buscar”  
Permite dictar un parámetro de búsqueda
3. “{mencionar nombre del sitio}”  
Al mencionar el nombre del sitio deseado lista los resultados de la búsqueda

## Navegación

Con el fin de poder visitar diferentes sitios web, el navegador brinda al usuario la posibilidad de ejecutar diversas acciones que tienen como fin administrar la navegabilidad de los sitios web. Estas funcionalidades son expuestas a través de la llamada barra de navegación.

### Abrir un sitio web indicando la dirección

Por lo general la acción inicial de toda sesión de navegación es indicar la dirección web a cargar. Para ello se indica la dirección en la barra de navegación y luego se presiona la tecla ENTER.



Figura 10.09: Barra de navegación

Comandos de voz equivalentes a la operación manual:

1. “Abrir”  
Permite ingresar la dirección del sitio web a visitar
2. “{dictar dirección del sitio web}”  
Indicar la dirección del sitio web a cargar

## Buscar un texto en internet

Por lo general los usuarios indican una dirección web dentro de la barra de navegación, sin embargo, si el texto indicado no cumple con el formato de una dirección, el navegador procede a buscar el texto en internet a través del buscador preestablecido.

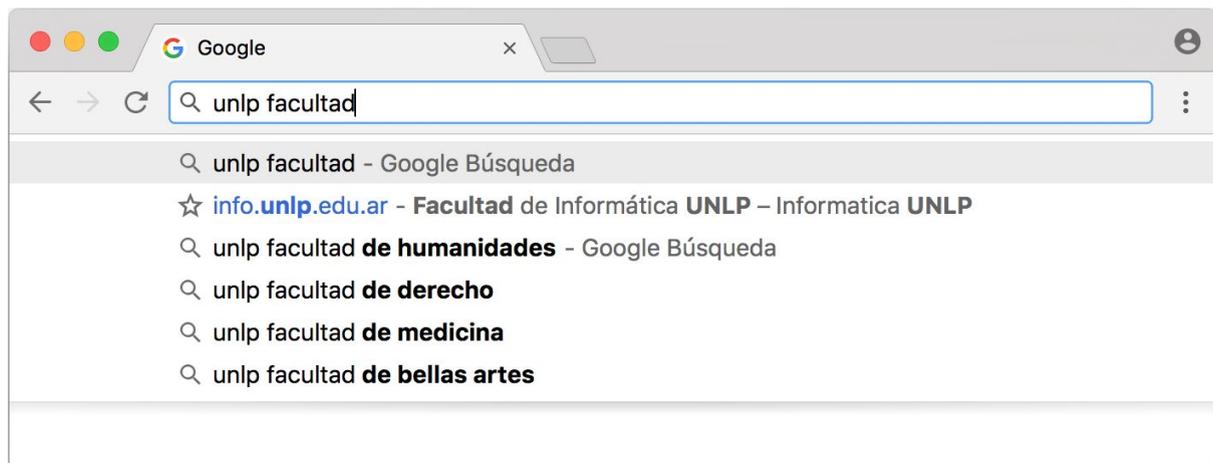


Figura 10.10: Barra de búsqueda

Comandos de voz equivalentes a la operación manual:

1. “Buscar”  
Permite ingresar los parámetros de búsqueda
2. “{dictar texto a buscar}”  
Indicar el texto a buscar

## Ir atrás

Consiste en navegar hacia la dirección web visitada con anterioridad.

Se puede ejecutar esta acción haciendo click en el botón que exhibe una flecha apuntando hacia la izquierda en la barra de navegación, o presionando la tecla BACKSPACE.

Comandos de voz equivalentes a la operación manual:

- “Atrás”  
Abre la dirección web cargada anteriormente.

## Ir adelante

Luego de ejecutar la acción Ir atrás, se puede volver a la dirección web visitada más recientemente haciendo click en el botón que exhibe una flecha apuntando hacia la derecha en la barra de navegación.

Comandos de voz equivalentes a la operación manual:

- “Adelante”  
Abre la dirección que estaba cargada antes de hacer click atrás

## Recargar sitio web

Habitualmente se desea refrescar un sitio web, posiblemente porque hubo un error de carga en el sitio o porque simplemente se desea cargar una versión más reciente.

Se puede ejecutar esta acción haciendo click en el botón que exhibe una flecha apuntando hacia el inicio de la misma o presionando las teclas `Ctrl+R`.

Comandos de voz equivalentes a la operación manual:

- “Recargar”  
Vuelve a cargar la dirección actual

## Detener carga de sitio web

En ciertas circunstancias se desea detener la carga de un sitio web.

Se puede ejecutar esta acción haciendo click en el botón que exhibe una cruz o presionando las teclas `Esc`.



Figura 10.11: Botón detener en barra de navegación

Comandos de voz equivalentes a la operación manual:

- “Detener”  
Interrumpe el proceso de carga del sitio web

## Buscar texto en página web

Habitualmente se desea buscar la ocurrencia de un texto dentro del contenido de la página web que se está visualizando.

Se puede ejecutar esta acción haciendo click en el menú contextual del navegador o presionando las teclas `Ctrl+F`. Luego es necesario escribir en el campo de texto las

palabras buscadas. Opcionalmente se puede hacer foco en los resultados encontrados haciendo click en los botones arriba y abajo que están a la derecha del campo de texto.

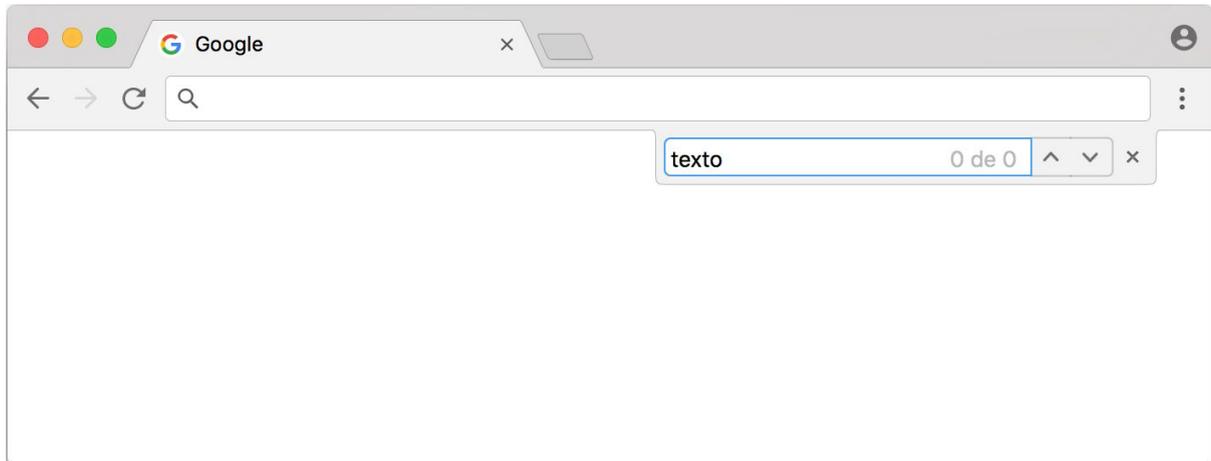


Figura 10.12: Barra de búsqueda de texto

Comandos de voz equivalentes a la operación manual:

1. “Buscar texto”  
Inicia el modo de búsqueda de texto en el sitio web actual
2. “{dictar texto a buscar}”  
Busca la primer ocurrencia del texto indicado
3. “Siguiente”  
Busca la siguiente ocurrencia del texto indicado
4. “Anterior”  
Opcionalmente se puede hacer foco en la anterior ocurrencia encontrada

## Imprimir

Una acción habitual para algunos usuarios es imprimir ciertas páginas webs. Esta acción puede ser ejecutada desde el menú contextual del navegador, o haciendo click en `Ctrl+P`

Comandos de voz equivalentes a la operación manual:

1. “Imprimir”  
Abre el administrador de impresión
2. “{mencionar marcadores correspondientes a las opciones de impresión}”  
Configura la impresión según las opciones disponibles
3. Aceptar  
Inicia la impresión según los parámetros configurados

## Gestión de Pestañas

Habitualmente los usuarios no realizan una sesión de navegación en una única pestaña, sino que, por el contrario, mantienen múltiples pestañas abiertas al mismo tiempo. A pesar

de la multiplicidad de pestañas abiertas en paralelo que el navegador puede tener en un momento dado, el usuario sólo navega una de ellas a la vez.

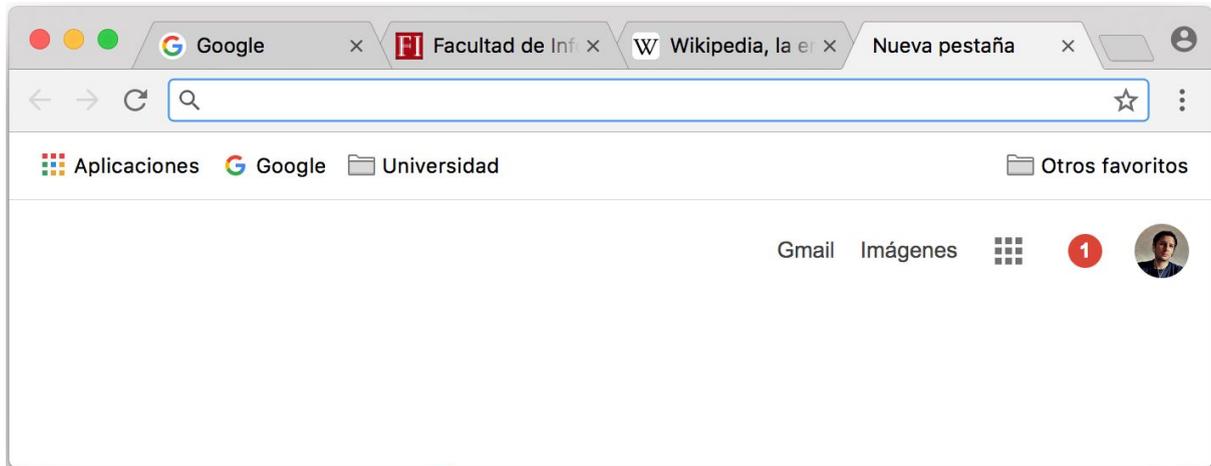


Figura 10.13: Pestañas

A continuación se detallan las operaciones que realiza el usuario con relación a las pestañas.

### **Abrir nueva pestaña**

Se puede abrir una nueva pestaña haciendo click en el botón de nueva pestaña que se encuentra a la derecha de las pestañas abiertas. O presionando las teclas `Ctrl+T`.

Comandos de voz equivalentes a la operación manual:

- “Nueva pestaña”  
Abre una nueva pestaña

### **Cerrar pestaña**

Se puede cerrar una pestaña existente haciendo click en la cruz que se encuentra a la derecha de su identificador, o presionando las teclas `Ctrl+W`.

Comandos de voz equivalentes a la operación manual:

- “Cerrar pestaña”  
Cierra la pestaña actual

### **Cambiar pestaña activa**

Se puede cambiar la pestaña activa haciendo click en el rótulo de la nueva pestaña activa deseada o presionando las teclas `Ctrl + ALT + ->` para seleccionar la pestaña de la derecha, o `Ctrl + ALT + ->` de la izquierda.

Comandos de voz equivalentes a la operación manual:

- “Siguiente pestaña”  
Activa la pestaña que está inmediatamente a la derecha de la pestaña actual o la primera en caso no haber una.
- “Anterior pestaña”  
Activa la pestaña que está inmediatamente a la izquierda de la pestaña actual o la última en caso no haber una.

## Opciones en Menú Contextual

Durante una sesión de navegación habitualmente se ejecutan opciones del menú contextual, a continuación se enumeran las más habituales.

### Copiar

Luego de seleccionar un texto o hacer foco en una imagen, se puede presionar el click derecho del mouse para desplegar el menú contextual y luego seleccionar la opción copiar.

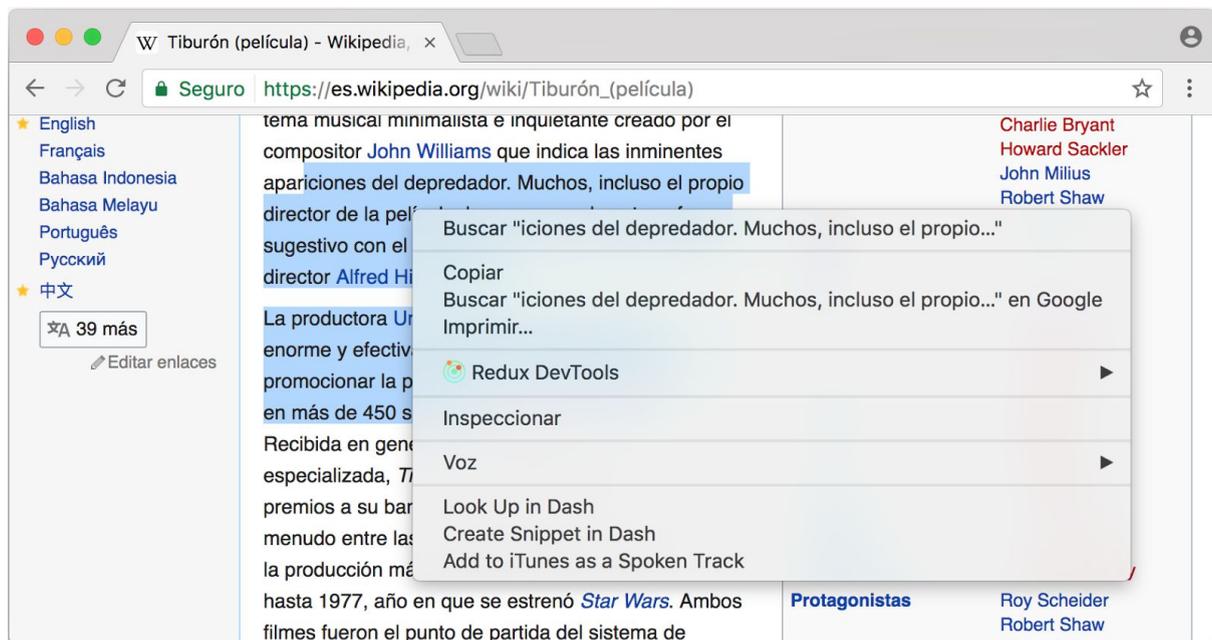


Figura 10.14: Menú contextual luego de seleccionar un texto

Comandos de voz equivalentes a la operación manual:

1. “Seleccionar”  
Muestra por pantalla un marcador por cada elemento que puede ser seleccionable
2. “{Indicar el nombre del marcador correspondiente elemento que se desea copiar}”  
Muestra por pantalla la lista de opciones disponibles luego de seleccionar un elemento
3. “Copiar”  
Guarda en el portapapeles el contenido correspondiente al elemento seleccionado.

## Pegar

Se puede pegar contenido previamente copiado al hacer foco en algún campo de entrada, luego presionar el click derecho del mouse y al desplegarse el menú contextual, seleccionar la opción pegar.

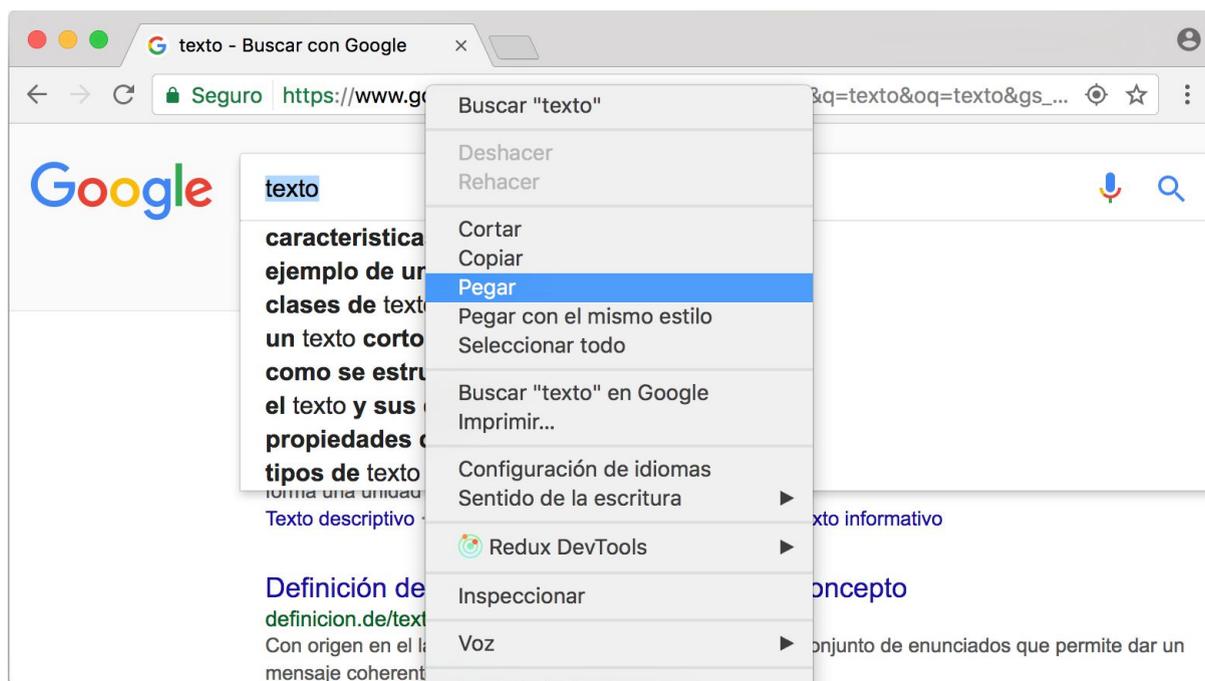


Figura 10.15: Menú contextual de un campo de introducción de texto

Comandos de voz equivalentes a la operación manual:

1. "Seleccionar"  
Muestra por pantalla un marcador por cada elemento que puede ser seleccionable
2. "{Indicar el nombre del marcador correspondiente elemento de entrada deseado}"  
Muestra por pantalla la lista de opciones disponibles luego de seleccionar un elemento
3. "Pegar"  
Introduce en el elemento de entrada el contenido previamente guardado en el portapapeles

## Abrir en nueva pestaña

A la hora de abrir un enlace, es habitual necesitar que la carga del mismo se realice en una nueva pestaña y de esta forma poder continuar la navegación en el sitio web actual y posteriormente en la nueva pestaña.

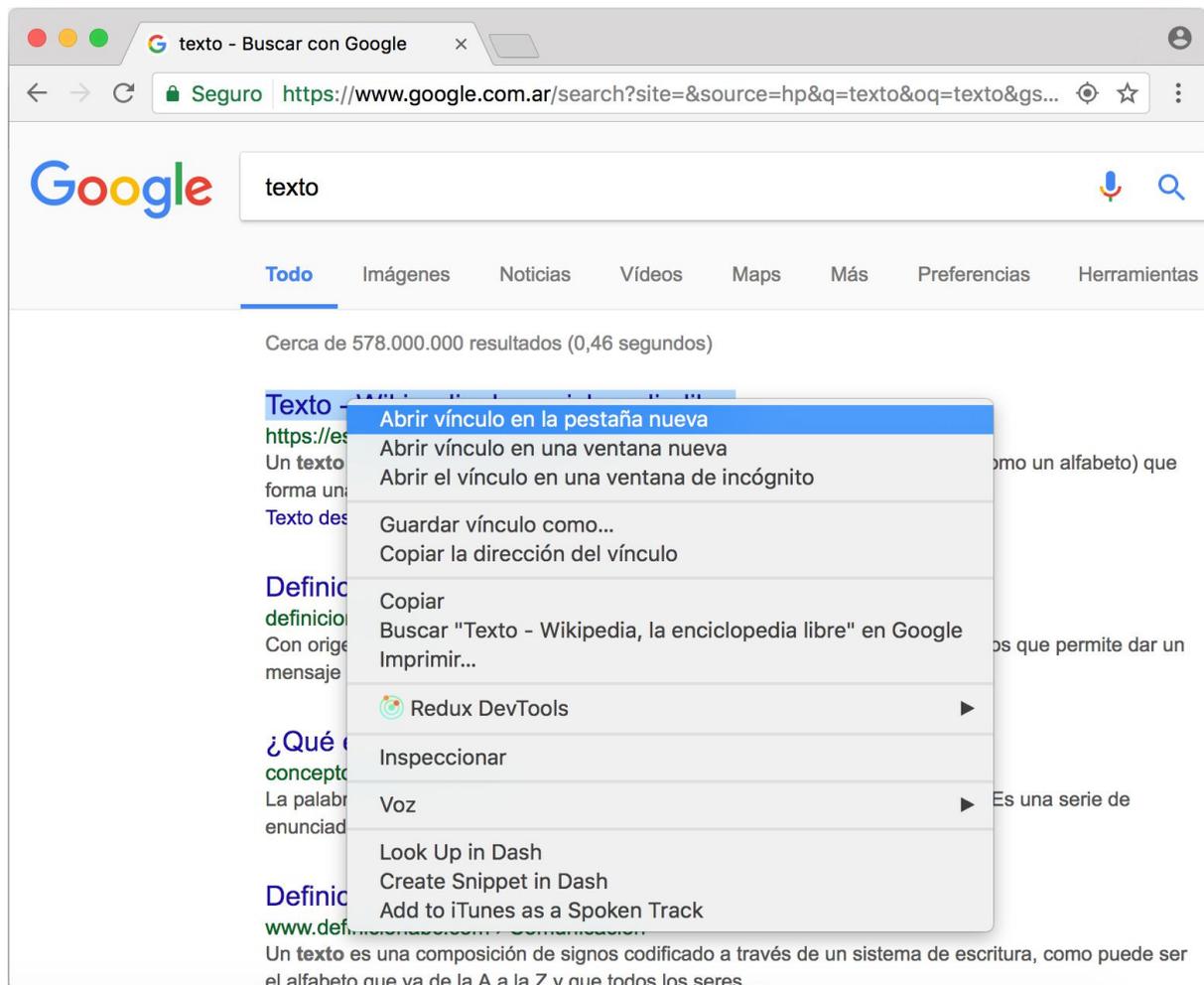


Figura 10.16: Menú contextual de un vínculo

Comandos de voz equivalentes a la operación manual:

1. “Seleccionar”  
Muestra por pantalla un marcador por cada elemento que puede ser seleccionable
2. “{Indicar el nombre del marcador correspondiente al enlace que se desea abrir}”  
Muestra por pantalla la lista de opciones disponibles luego de seleccionar un elemento
3. “Abrir en nueva pestaña”  
Carga la dirección del enlace seleccionado en una nueva pestaña

### Desplazamientos dentro del sitio web

Al ingresar al sitio web, difícilmente el mismo puede exhibir todo el contenido dentro del tamaño de la pantalla, por lo general es necesario desplazar el mismo hacia abajo o arriba para poder ver el contenido en su totalidad.

Existen diversas formas para desplazarse en una página web:

- Haciendo click en los botones de las barras de desplazamiento
- Presionando las teclas *arriba*, *abajo*, *av pag*, *re pag* o *SPACE* del teclado.

- Alternativamente se puede ir al inicio de la página presionando `Ctrl + arriba` o ir al final presionando `Ctrl + abajo`.

Comandos de voz equivalentes a la operación manual:

- “Subir”  
Desplaza el contenido hacia abajo
- “Bajar”  
Desplaza el contenido hacia arriba
- “Arriba”  
Desplaza la pantalla hasta alcanzar la parte superior del sitio web
- “Abajo”  
Desplaza la pantalla hasta alcanzar la parte inferior del sitio web

# Anexo 2

## Soporte de comandos de voz a funcionalidad genérica de los sitios web

A continuación se define por extensión el conjunto de componentes soportados por HTML 5 junto con una descripción de la interacción manual que debe realizar el usuario para operar con ellos y una alternativa de interacción utilizando comandos de voz.

### Elementos <input>

Un componente web definido con el tag “input” denota un campo de entrada o control el cual es muy útil, por ejemplo, para componer un formulario html.

Un elemento input puede presentarse en diferentes formatos. A continuación se presentan todas sus variantes:

#### Elemento Button

El elemento de tipo `button` representa un botón sin estado.



Figura 11.1: Representación visual de un botón.

```
<input type="button" value="Botón" />
```

Sección de Código 11.1: Código html correspondiente a un botón.

Rol Aria relacionado: `button`

Operaciones manuales admitidas:

- Click sobre el botón utilizando el mouse.
- Luego de hacer foco en el botón utilizando la tecla TAB del teclado se puede simular el click presionando la tecla ENTER del teclado.

Comandos de voz equivalentes a la operación manual:

1. “Click”  
Visualiza una referencia por cada elemento que puede ser presionado.
2. “{indicar referencia correspondiente al botón que se desea presionar}”  
Ejecuta la acción de click sobre el elemento referenciado.

## Elemento Checkbox

El elemento de tipo `checkbox` representa un campo de entrada que puede tener dos estados: habilitado o deshabilitado.



Figura 11.2: Representación visual de un checkbox desmarcado

```
<input type="checkbox" />
```

Sección de Código 11.2: Código html correspondiente a un checkbox

Rol Aria relacionado: `checkbox`, `menuitemcheckbox`

Operaciones manuales admitidas:

- Click sobre el checkbox utilizando el mouse.
- Luego de hacer foco en el checkbox utilizando la tecla TAB del teclado se puede simular el click presionando la tecla ENTER del teclado.

Comandos de voz equivalentes a la operación manual:

1. "Click"  
Visualiza una referencia por cada elemento que puede ser presionado.
2. "{indicar referencia correspondiente a la opción que se desea presionar}"  
Ejecuta la acción de click sobre el elemento referenciado.

## Elemento Color

El elemento de tipo `color` representa un campo de entrada cuyo valor es la especificación de un color en formato RGB. Al hacer click sobre el campo, se abre una ventana independiente a la del navegador donde se puede seleccionar el color deseado.

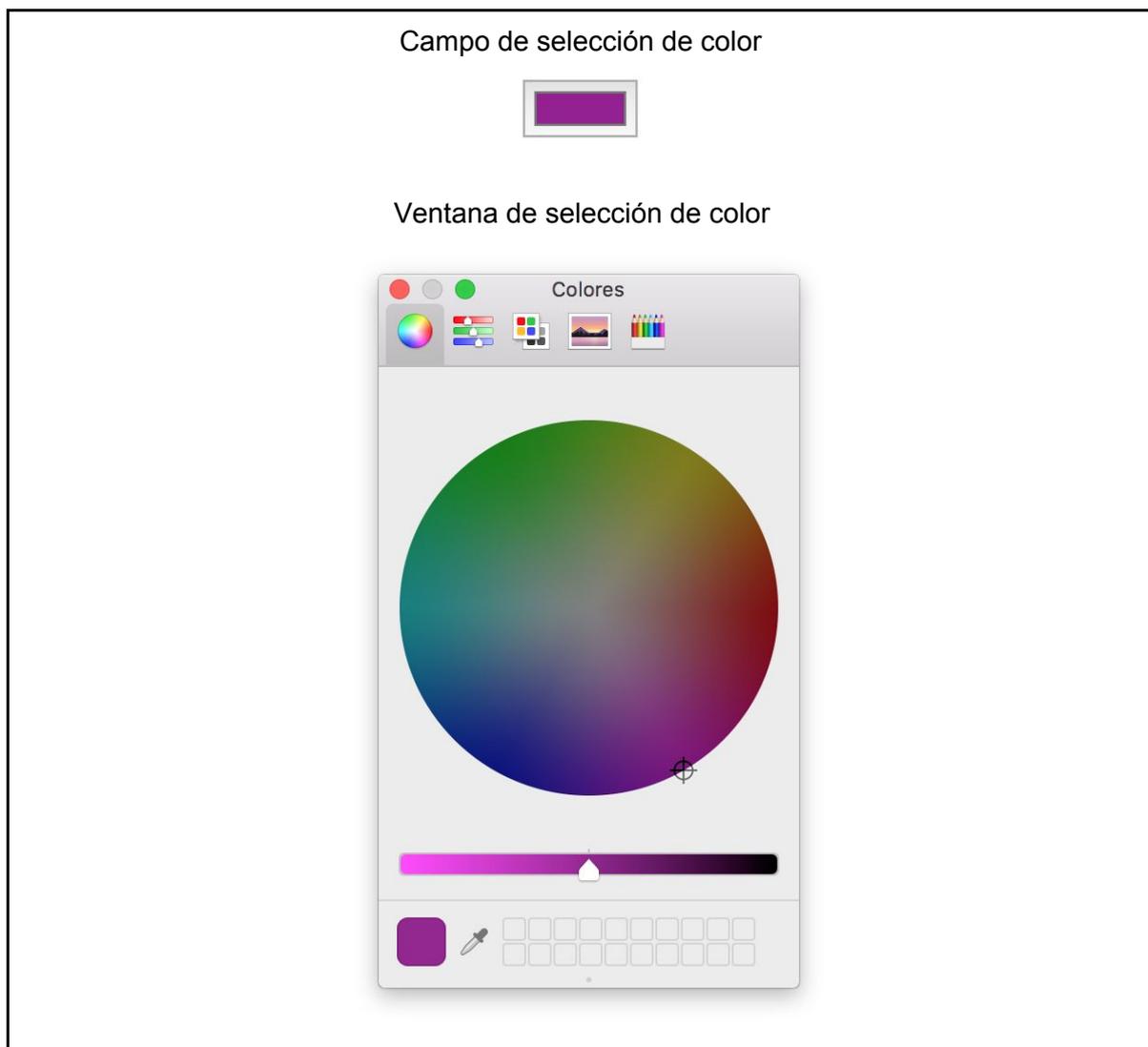


Figura 11.3: Representación visual de campo de selección de color.

```
<input type="color" />
```

Sección de Código 11.3: Código html correspondiente a un campo de selección de color.

Operaciones manuales admitidas:

- Click sobre el campo de color utilizando el mouse.
- Luego de hacer foco en el campo de color, utilizando la tecla TAB del teclado, se puede simular el click presionando la tecla ENTER del teclado.

- Se puede seleccionar un color de la paleta de colores exhibidas en la ventana de selección únicamente utilizando el mouse. No es posible la navegación por teclado dentro de la ventana de selección.

Nota: La ventana de selección de color no forma parte de la estructura del sitio web, es una ventana independiente administrada por el navegador.

Comandos de voz equivalentes a la operación manual:

1. "Click"  
Visualiza una referencia por cada elemento que puede ser presionado.
2. "{indicar referencia correspondiente al campo de introducción de color}"  
Presenta visualmente una lista de colores disponibles de ser seleccionados identificados por una referencia.
3. "{indicar referencia correspondiente al color deseado}"  
Selecciona el color mencionado.

## Elemento Date

El tipo `date` representa un campo de entrada cuyo valor es una fecha en formato `dd/mm/aaaa`. El campo permite ingresar la fecha de múltiples formas; se puede hacer foco sobre el campo y escribir la fecha, interactuar con las flechas hasta alcanzar la fecha deseada o seleccionar una fecha desde la ventana calendario.



Figura 11.4: Representación visual de campo de selección de fecha

```
<input type="date" value="2014-02-09" />
```

Sección de Código 11.4: Código html correspondiente a un campo de selección de fecha

Operaciones manuales admitidas:

- Luego de hacer click sobre el campo de fecha utilizando el mouse o el teclado, se puede escribir la fecha deseada.
- Se puede hacer click sobre las flechas del componente hasta alcanzar la fecha deseada.
- Se puede hacer click en la flecha del extremo derecho, para desplegar la ventana calendario. Luego únicamente utilizando el mouse se puede seleccionar la fecha deseada.

Nota: La ventana calendario donde se puede seleccionar la fecha no forma parte de la estructura del sitio web, es una ventana independiente administrada por el navegador.

Comandos de voz equivalentes a la operación manual:

1. "Click"  
Visualiza una referencia por cada elemento que puede ser presionado.
2. "{indicar referencia correspondiente al campo de introducción de fecha}"  
Solicita por pantalla mencionar la fecha deseada.
3. "{mencionar fecha deseada}"  
Selecciona la fecha indicada.

## Elemento Datetime

El elemento de tipo `datetime` representa un campo de entrada de solo texto. Actualmente este componente no muestra diferencias visuales o de comportamiento respecto al tipo `text`.

```
<input type="datetime" />
```

Sección de Código 11.5: Código html correspondiente a un campo de selección de fecha y hora

Ver elemento de tipo `text` en esta misma sección para más detalles.

## Elemento Datetime-local

El elemento de tipo `datetime-local` representa un campo de entrada cuyo valor es una fecha en formato `dd/mm/aaaa, hh:mm:ss am|pm`. El campo permite ingresar la fecha de múltiples formas; se puede hacer foco sobre el campo y escribir la fecha, interactuar con las flechas hasta alcanzar la fecha deseada o seleccionar una fecha desde la ventana calendario.

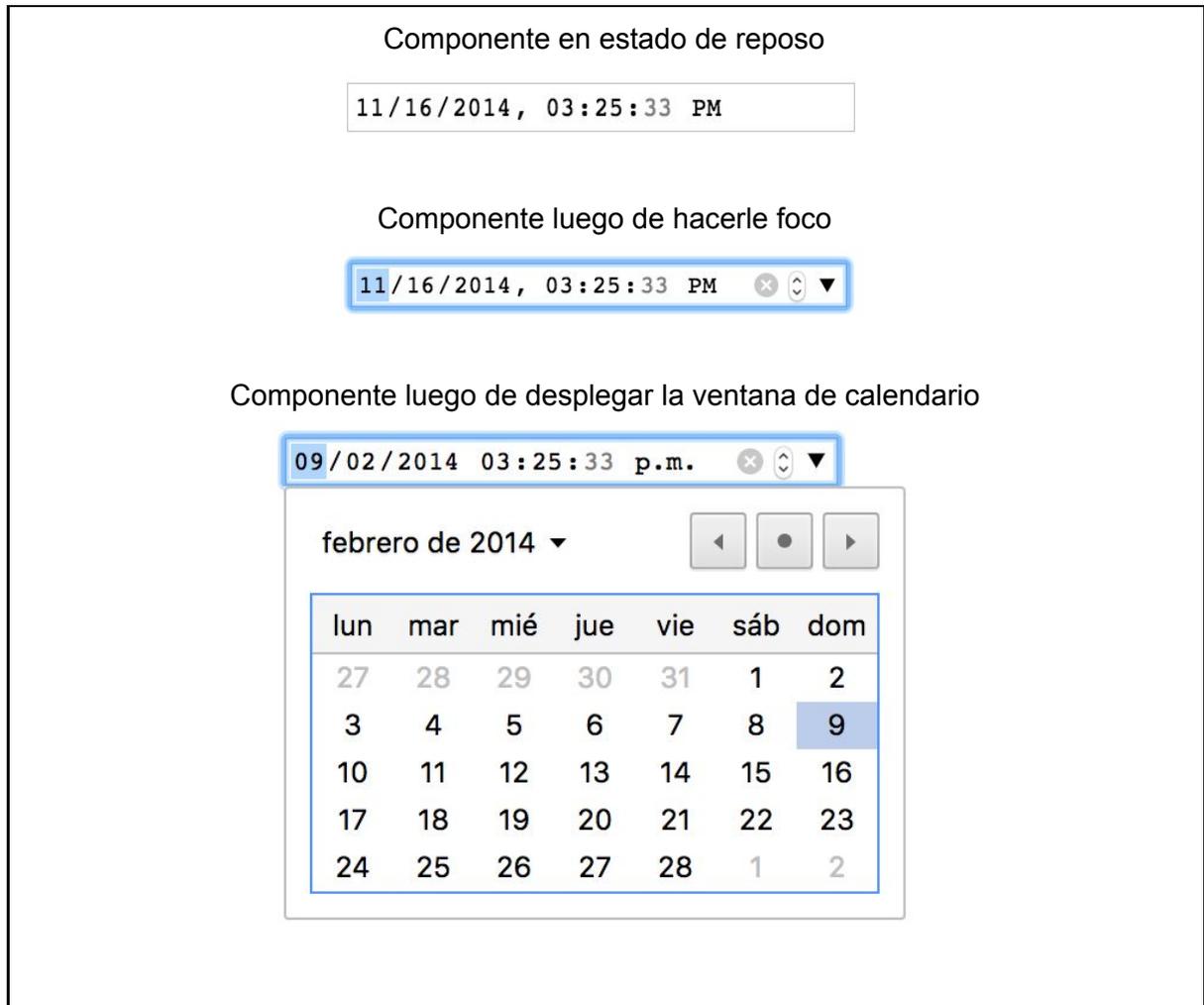


Figura 11.5: Representación visual de campo de selección de fecha y hora.

```
<input type="datetime-local" value="2014-02-09T15:25:33" />
```

Sección de Código 11.6: Código html correspondiente a un campo de selección de fecha y hora.

Operaciones manuales admitidas:

- Luego de hacerle click sobre el campo de fecha utilizando el mouse o el teclado, se puede escribir la fecha deseada.
- Se puede hacer click sobre las flechas del componente hasta alcanzar la fecha deseada.
- Se puede hacer click en la flecha del extremo derecho para desplegar la ventana calendario, luego únicamente utilizando el mouse se puede seleccionar la fecha deseada.

Nota: La ventana calendario donde se puede seleccionar la fecha no forma parte de la estructura del sitio web, es una ventana independiente administrada por el navegador.

Comandos de voz equivalentes a la operación manual:

1. “Click”  
Visualiza una referencia por cada elemento que puede ser presionado.
2. “{indicar referencia correspondiente al campo de introducción de fecha y hora}”  
Solicita por pantalla mencionar la fecha y hora deseada.
3. “{mencionar fecha y hora deseada}”  
Selecciona la fecha y hora indicada.

## Elemento Email

El elemento de tipo `email` representa un campo de entrada de solo texto. Actualmente este componente no muestra diferencias visuales o de comportamiento respecto al tipo `text`.

```
<input type="email" />
```

Sección de Código 11.7: Código html correspondiente a un campo de inserción de email

Ver elemento de tipo `text` en esta misma sección para más detalles.

## Elemento File

El elemento de tipo `file` representa un campo de entrada que permite adjuntar la referencia a un archivo local.



Seleccionar archivo No se eligió archivo

Figura 11.6: Representación visual de campo de selección de archivo.

```
<input type="file" />
```

Sección de Código 11.8: Código html correspondiente a un campo de inserción de email.

Operaciones manuales admitidas:

- Click sobre el botón utilizando el mouse.
- Luego de hacer foco en el botón utilizando la tecla TAB del teclado, se puede simular el click presionando la tecla ENTER del teclado.

Comandos de voz equivalentes a la operación manual:

1. “Click”  
Visualiza una referencia por cada elemento que puede ser presionado.
2. “{indicar referencia correspondiente al campo de selección de archivo}”  
Abre diálogo de selección de archivo provisto por el navegador.

## Elemento Month

El elemento de tipo `month` representa un campo de entrada cuyo valor es un mes acompañado de un año. El campo permite ingresar el mes y el año de múltiples formas; se puede hacer foco sobre el campo y escribir el mes y año, interactuar con las flechas hasta alcanzar la fecha deseada o seleccionar una fecha desde la ventana calendario.

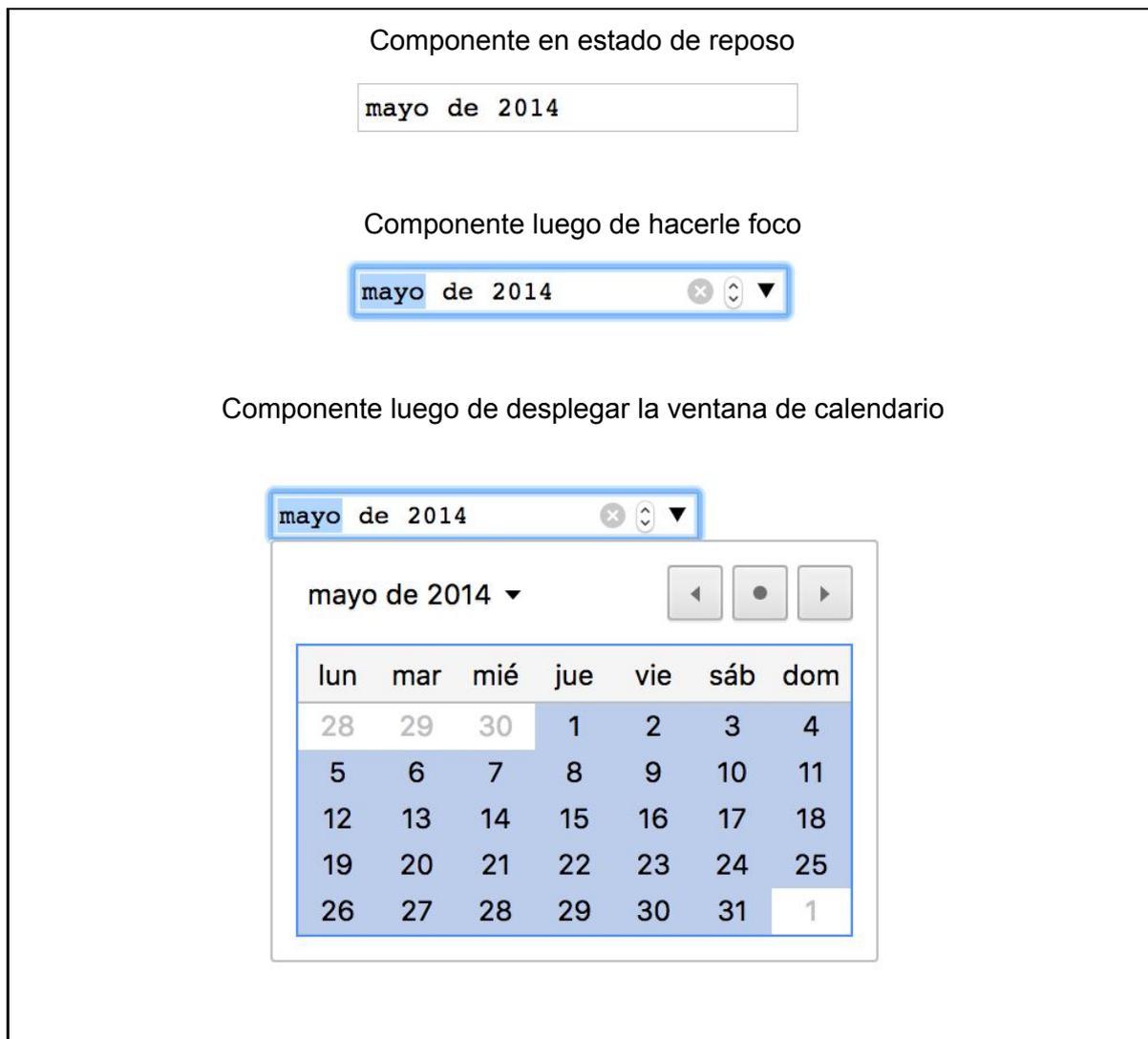


Figura 11.7: Representación visual de campo de selección de mes de un año.

```
<input type="month" type="2014-05" />
```

Sección de Código 11.9: Código html correspondiente a un campo de selección de mes de un año.

Operaciones manuales admitidas:

- Luego de hacer click sobre el campo utilizando el mouse o el teclado, se puede escribir el mes y el año.

- Se puede hacer click sobre las flechas del componente hasta alcanzar la fecha deseada.
- Se puede hacer click en la flecha del extremo derecho para desplegar la ventana calendario, luego únicamente utilizando el mouse se puede seleccionar la fecha deseada.

Nota: La ventana calendario donde se puede seleccionar la fecha no forma parte de la estructura del sitio web, es una ventana independiente administrada por el navegador.

Comandos de voz equivalente a la operación manual:

1. "Click"  
Visualiza una referencia por cada elemento que puede ser presionado.
2. "{indicar referencia correspondiente al campo de introducción mes y año}"  
Solicita por pantalla mencionar mes y año deseado.
3. "{mencionar mes y año deseado}"  
Selecciona el mes y año deseado.

## Elemento Number

El elemento de tipo `number` representa un campo de entrada cuyo valor es un número. El campo permite ingresar exclusivamente números enteros o decimales de múltiples formas; se puede hacer foco sobre el campo y escribir el valor o interactuar con las flechas hasta alcanzar el valor deseado.

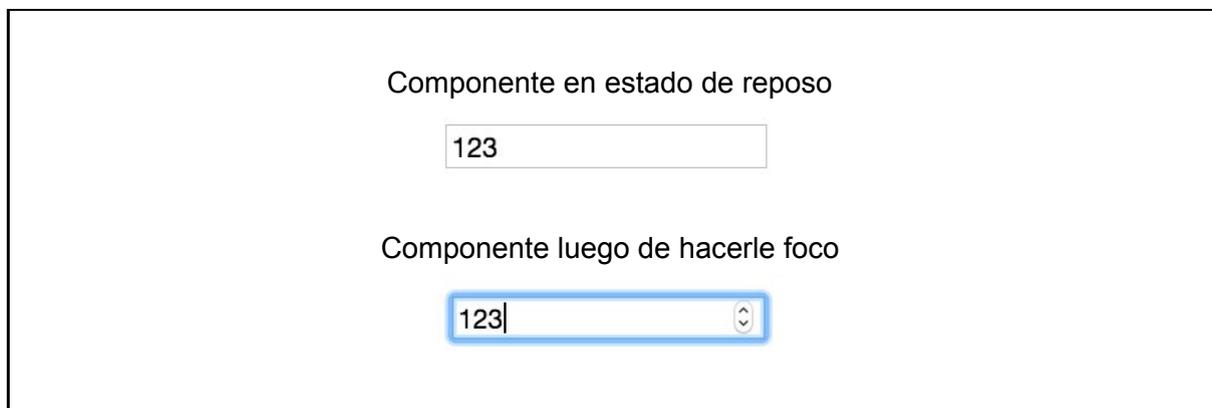


Figura 11.8: Representación visual del campo de inserción de números.

```
<input type="number" value="123" />
```

Sección de Código 11.10: Código html correspondiente a un campo de inserción de números

Operaciones manuales admitidas:

- Luego de hacer click sobre el campo utilizando el mouse o el teclado, se puede escribir el valor.
- Se puede hacer click sobre las flechas del componente hasta alcanzar el valor deseado.

Comandos de voz equivalentes a la operación manual:

1. “Click”  
Visualiza una referencia por cada elemento que puede ser presionado.
2. “{indicar referencia correspondiente al campo de introducción de números}”  
Solicita por pantalla mencionar el número deseado.
3. “{mencionar número deseado}”  
Introduce el número indicado.

## Elemento Password

El elemento de tipo `password` representa un campo de entrada cuyo valor es un texto que por cuestiones de seguridad se muestra en forma de asteriscos. Se puede hacer foco sobre el campo y escribir el valor.

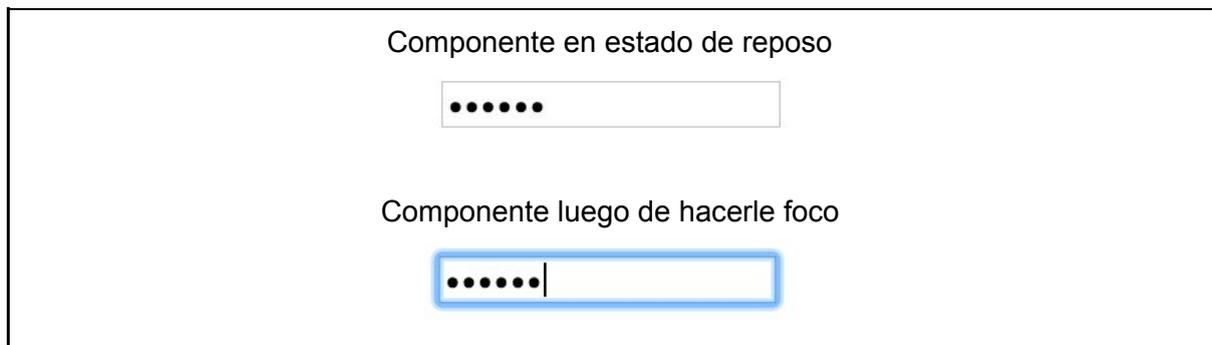


Figura 11.9: Representación visual de campo de inserción de contraseña.

```
<input type="password" value="psw123" />
```

Sección de Código 11.11: Código html correspondiente a un campo de inserción de contraseña.

Operaciones manuales admitidas:

- Luego de hacer click sobre el campo utilizando el mouse o el teclado, se puede escribir el valor.

Comandos de voz equivalentes a la operación manual:

1. “Click”  
Visualiza una referencia por cada elemento que puede ser presionado
2. “{indicar referencia correspondiente al campo de introducción de contraseña}”  
Solicita por pantalla mencionar el número deseado
3. “{indicar la contraseña deseada}”  
Introduce la contraseña dictada

## Elemento Radio

El elemento de tipo `radio` representa un campo de entrada que puede tener dos estados: habilitado o deshabilitado. A diferencia del tipo `checkbox` los elementos de tipo `radio` son de selección única. Es decir no puede haber más de uno seleccionado al mismo tiempo.

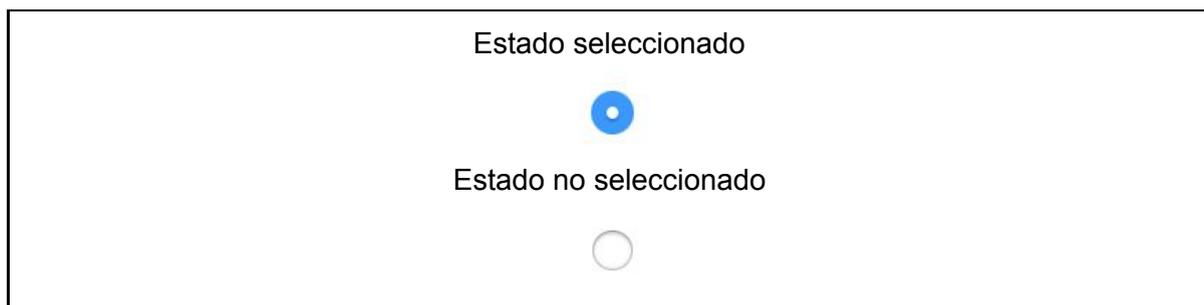


Figura 11.10: Representación visual de campo de selección de opciones.

```
<input type="radio" />
```

Sección de Código 11.12: Código html correspondiente a un campo de selección de opciones.

Rol Aria relacionado: `menuitemradio`, `radio`

Operaciones manuales admitidas:

- Click sobre el elemento utilizando el mouse.
- Luego de hacer foco en el elemento utilizando la tecla TAB del teclado se puede simular el click presionando la tecla ENTER del teclado.

Comandos de voz equivalentes a la operación manual:

1. "Click"  
Visualiza una referencia por cada elemento que puede ser presionado.
2. "{indicar referencia correspondiente a la opción que se desea presionar}"  
Ejecuta la acción de click sobre el elemento referenciado.

## Elemento Range

El elemento de tipo `range` representa un campo de entrada cuyo valor es un número, el cual se encuentra delimitado entre dos valores. El campo permite ingresar el valor deseado haciendo uso de botón que se puede deslizar sobre un eje.

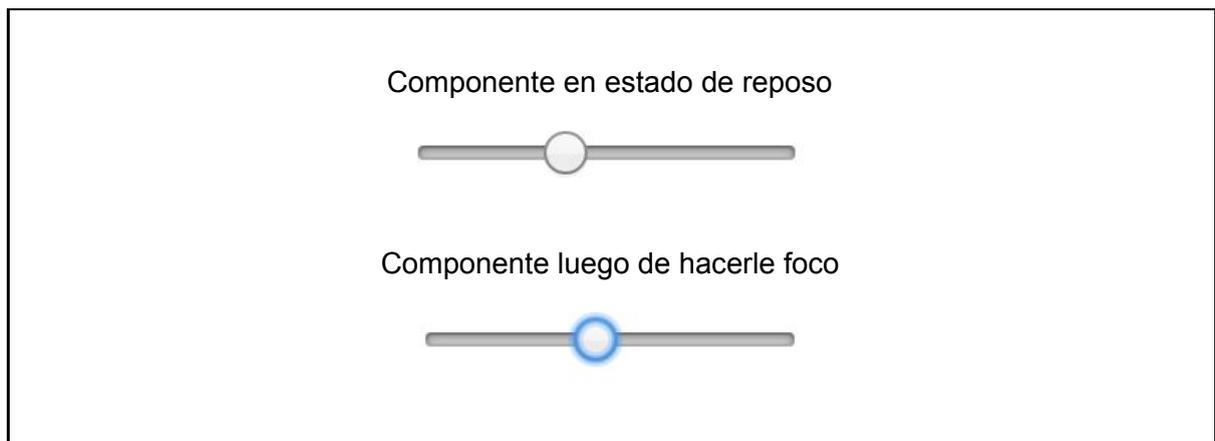


Figura 11.11: Representación visual de campo de selección de opciones.

```
<input type="range" min="1" max="200" value="90" />
```

Sección de Código 11.13: Código html correspondiente a un campo de selección de un valor de rango.

Rol Aria relacionado: `range`

Operaciones manuales admitidas:

- Click sobre el botón utilizando el mouse y luego arrastrar el botón hacia los extremos del eje.
- Luego de hacer foco en el botón utilizando la tecla TAB del teclado se puede deslizar el botón usando las teclas izquierda y derecha del teclado.

Comandos de voz equivalente a la operación manual:

1. "Click"  
Visualiza una referencia por cada elemento que puede ser sujeto de ser presionado.
2. "{indicar referencia correspondiente al campo de rango que se desea presionar}"  
Ejecuta la acción de click sobre el elemento referenciado.
3. "{indicar el valor deseado}"  
Selecciona el valor indicado.

## Elemento Reset

El elemento de tipo `reset` representa un botón sin estado, el cual es utilizado habitualmente para reiniciar los campos de entrada de un formulario.

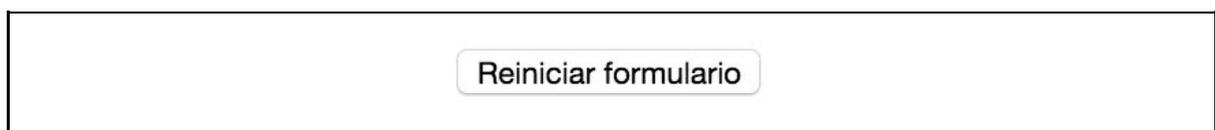


Figura 11.12: Representación visual de botón reset.

```
<input type="reset" value="Reiniciar formulario" />
```

Sección de Código 11.14: Código html correspondiente a botón reset.

Operaciones manuales admitidas:

- Click sobre el botón utilizando el mouse
- Luego de hacer foco en el botón utilizando la tecla TAB del teclado se puede simular el click presionando la tecla ENTER del teclado

Comandos de voz equivalentes a la operación manual:

1. "Click"  
Visualiza una referencia por cada elemento que puede ser presionado
2. "{indicar referencia correspondiente al botón que se desea presionar}"  
Ejecuta la acción de click sobre el elemento referenciado.

## Elemento Search

El elemento de tipo `search` representa un campo de entrada de solo texto. Actualmente este componente no muestra diferencias visuales o de comportamiento respecto al tipo `text`.

```
<input type="search" placeholder="Buscar" />
```

Sección de Código 11.15: Código html correspondiente a campo de búsqueda.

Rol Aria relacionado: `search`

Ver elemento de tipo `text` para más detalles.

## Elemento Submit

El elemento de tipo `submit` representa un botón sin estado, el cual es utilizado habitualmente para enviar los valores de los campos de entrada de un formulario.



Figura 11.13: Representación visual de botón submit.

```
<input type="submit" value="Enviar" />
```

Sección de Código 11.16: Código html correspondiente a botón submit.

Operaciones manuales admitidas:

- Click sobre el botón utilizando el mouse.
- Luego de hacer foco en el botón utilizando la tecla TAB del teclado se puede simular el click presionando la tecla ENTER del teclado.

Comandos de voz equivalentes a la operación manual:

1. "Click"  
Visualiza una referencia por cada elemento que puede ser presionado.
2. "{indicar referencia correspondiente al botón que se desea presionar}"  
Ejecuta la acción de click sobre el elemento referenciado.

## Elemento Tel

El elemento de tipo `tel` representa un campo de entrada de solo texto. Actualmente este componente no muestra diferencias visuales o de comportamiento respecto al tipo `text`.

```
<input type="tel"/>
```

Sección de Código 11.17: Código html correspondiente al campo de texto teléfono.

Ver tipo `text` para más detalles.

## Elemento Text

El elemento de tipo `text` representa un campo de entrada de solo texto. Luego de hacer foco en el campo se puede escribir libremente el valor deseado.

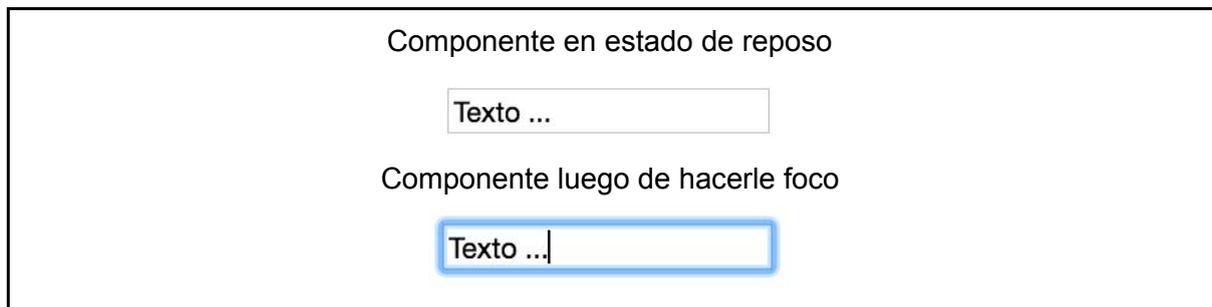


Figura 11.14: Representación visual de un campo de texto.

```
<input type="text" value="Texto ..."/>
```

Sección de Código 11.18: Código html correspondiente a un campo de texto.

Operaciones manuales admitidas:

- Click sobre el campo de texto y luego escribir haciendo uso del teclado.
- Luego de hacer foco en el botón utilizando la tecla TAB se puede escribir haciendo uso del teclado.

Comandos de voz equivalentes a la operación manual:

1. “Click”  
Visualiza una referencia por cada elemento que puede ser presionado.
2. “{indicar referencia correspondiente al campo de introducción de texto}”  
Solicita por pantalla mencionar el número deseado.
3. “{mencionar el texto deseado}”  
Introduce la contraseña dictada

## Elemento Time

El elemento de tipo `time` representa un campo de entrada cuyo valor es un horario en formato `hh:mm am|pm`. El campo permite ingresar el horario de múltiples formas; se puede hacer foco sobre el campo y escribir el horario o interactuar con las flechas hasta alcanzar el horario deseado.

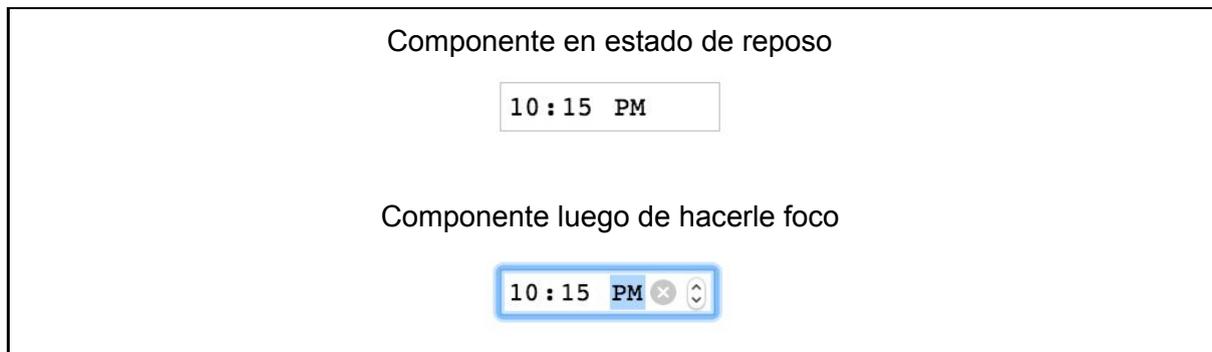


Figura 11.15: Representación visual de un campo de selección de hora.

```
<input type="time" value="22:15:00"/>
```

Sección de Código 11.19: Código html correspondiente a un campo de selección de horario.

Operaciones manuales admitidas:

- Luego de hacer click sobre el campo utilizando el mouse o el teclado, se puede escribir la hora y los minutos.
- Se puede hacer click sobre las flechas del componente hasta alcanzar el horario deseado.

Comandos de voz equivalente a la operación manual:

1. "Click"  
Visualiza una referencia por cada elemento que puede ser presionado.
2. "{indicar referencia correspondiente al campo de introducción horario}"  
Solicita por pantalla mencionar el horario deseado.
3. "{mencionar horario deseado}"  
Selecciona la hora y minutos deseado.

## Elemento Url

El elemento de tipo `url` representa un campo de entrada de solo texto. Actualmente este componente no muestra diferencias visuales o de comportamiento respecto al tipo `text`.

```
<input type="url" value="http://wikipedia.com"/>
```

Sección de Código 11.20: Código html correspondiente a un campo de inserción de dirección web.

Ver tipo `text` para más detalles.

## Elemento Week

El elemento de tipo `week` representa un campo de entrada cuyo valor es el número de una semana de un año determinado. El campo permite ingresar el número de semana y el año de múltiples formas; se puede hacer foco sobre el campo y escribir el número de semana y año, interactuar con las flechas hasta alcanzar la semana y año deseado o seleccionar una semana correspondiente a un año desde la ventana calendario.

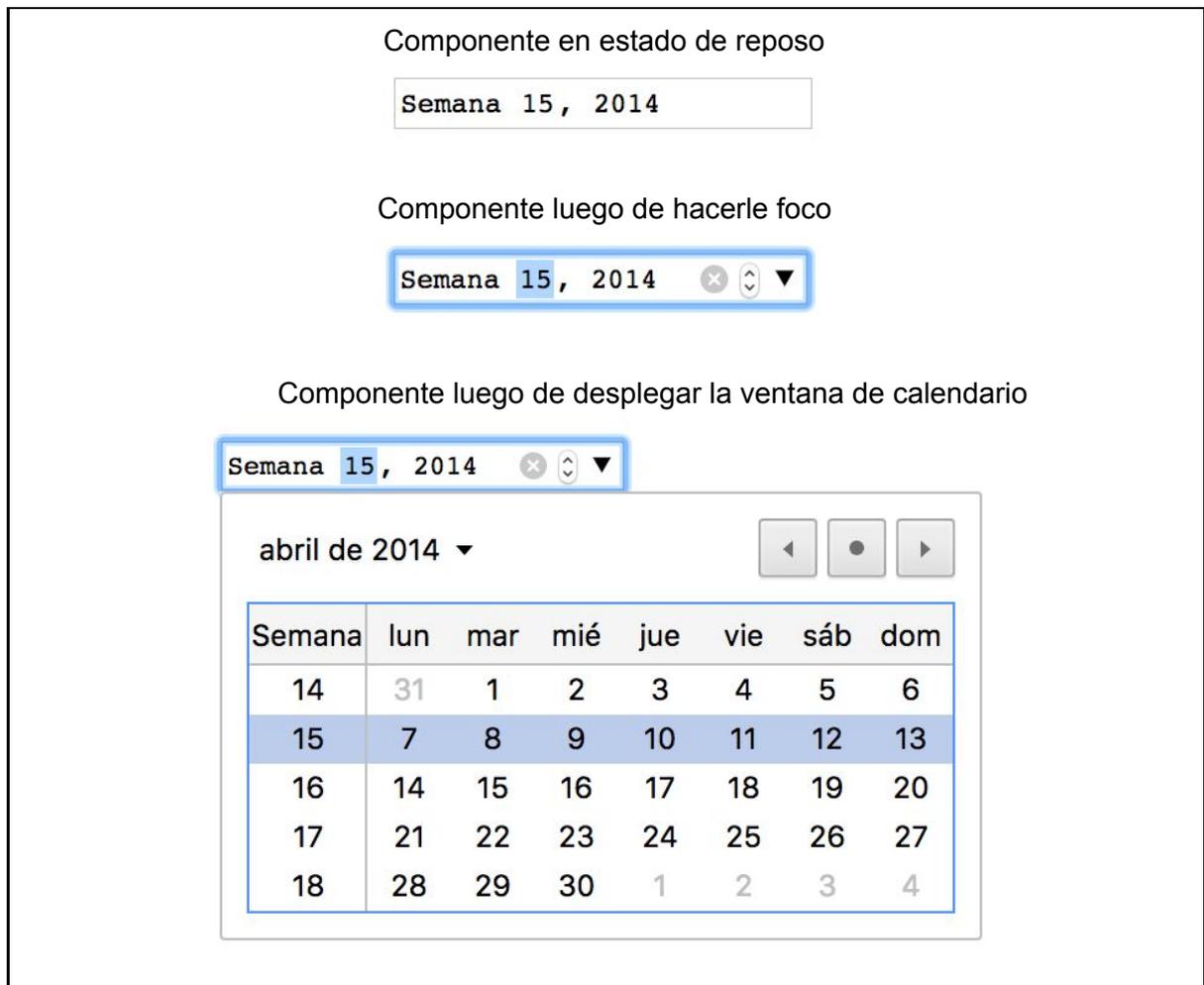


Figura 11.16: Representación visual de un campo de selección de semana de un año.

```
<input type="week" value="2014-W15"/>
```

Sección de Código 11.21: Código html correspondiente a un campo de selección de semana de un año.

Operaciones manuales admitidas:

- Luego de hacer click sobre el campo utilizando el mouse o el teclado, se puede escribir la semana y el año.
- Se puede hacer click sobre las flechas del componente hasta alcanzar la semana y año deseado.
- Se puede hacer click en la flecha del extremo derecho para desplegar la ventana calendario, luego, únicamente utilizando el mouse, se puede seleccionar la semana y el año deseado.

Nota: La ventana calendario donde se puede seleccionar la fecha no forma parte de la estructura del sitio web, es una ventana independiente administrada por el navegador.

Comandos de voz equivalentes a la operación manual:

1. "Click"  
Visualiza una referencia por cada elemento que puede ser presionado.
2. "{indicar referencia correspondiente al campo de introducción semana del año}"  
Solicita por pantalla mencionar la semana del año deseado.
3. "{mencionar semana del año deseado}"  
Selecciona la semana del año deseado.

## Componente Textarea

El componente visual `textarea` es un campo de entrada similar al `input` de tipo `text`, con la particularidad de que es multilínea.

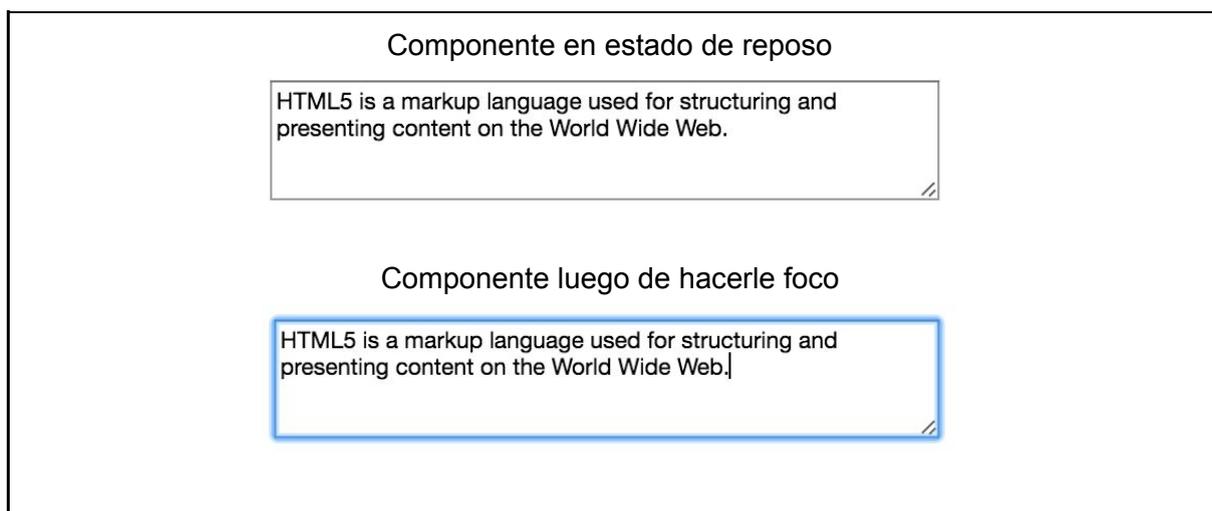


Figura 11.17: Representación visual de un campo de inserción de texto multilínea.

```
<textarea rows="4" cols="50">  
HTML5 is a markup language used for structuring and presenting content  
on the World Wide Web.  
</textarea>
```

Sección de Código 11.22: Código html correspondiente a un campo de inserción de texto multilínea

Rol Aria relacionado: `textbox`

Operaciones manuales admitidas:

- Click sobre el campo de texto y luego escribir haciendo uso del teclado.
- Luego de hacer foco en el botón utilizando la tecla TAB se puede escribir haciendo uso del teclado.

Comandos de voz equivalentes a la operación manual:

1. “Click”  
Visualiza una referencia por cada elemento que puede ser presionado
2. “{indicar referencia correspondiente al campo de introducción de texto}”  
Solicita por pantalla mencionar el número deseado.
3. “{mencionar el texto deseado}”  
Introduce la contraseña dictada.

## Componente Button

El componente visual `button` representa un botón sin estado.



Figura 11.18: Representación visual de un botón.

```
<button type="button">Botón</button>
```

Sección de Código 11.23: Código html correspondiente a un botón.

Operaciones manuales admitidas:

- Click sobre el botón utilizando el mouse
- Luego de hacer foco en el botón utilizando la tecla TAB del teclado se puede simular el click presionando la tecla ENTER del teclado

Comandos de voz equivalentes a la operación manual:

1. “Click”  
Visualiza una referencia por cada elemento que puede ser presionado.
2. “{indicar referencia correspondiente al botón que se desea presionar}”  
Ejecuta la acción de click sobre el elemento referenciado.

## Componente Select

El componente visual `select` representa una lista desplegable que admite la selección de un valor de una lista.

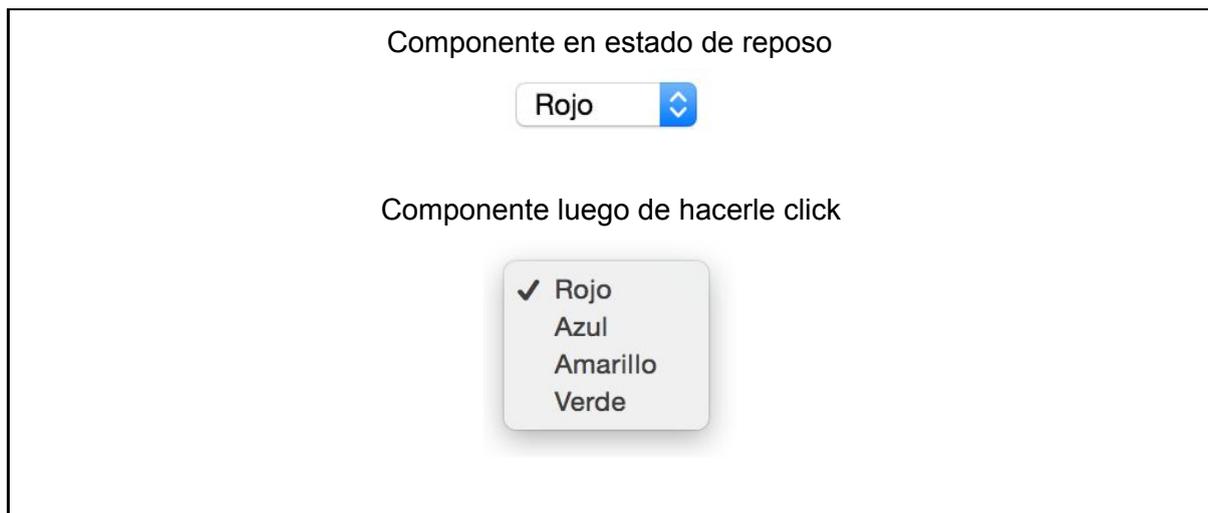


Figura 11.19: Representación visual de un campo de selección de opción.

```
<select>
  <option value="rojo">Rojo</option>
  <option value="azul">Azul</option>
  <option value="amarillo">Amarillo</option>
  <option value="verde">Verde</option>
</select>
```

Sección de Código 11.24: Código html correspondiente a un campo de selección de opción.

Rol aria relacionado: `option`, `listbox`, `listitem`

Operaciones manuales admitidas:

- Click sobre el campo utilizando el mouse, luego hacer click sobre el valor deseado
- Luego de hacer foco en el campo utilizando la tecla TAB del teclado se puede simular el click presionando la tecla ENTER del teclado y seleccionar el valor deseado haciendo uso de las teclas arriba y abajo del teclado.

Comandos de voz equivalente a la operación manual:

1. "Click"  
Visualiza una referencia por cada elemento que puede ser presionado.
2. "{indicar referencia correspondiente al campo de selección de opciones}"  
Solicita por pantalla mencionar la opción deseada.
3. "{mencionar opción deseada}"  
Selecciona la opción deseada.

## Componente Select multiple

El componente visual `select` con el atributo `multiple` es una variante de `select` que es utilizado para permitir seleccionar múltiples valores.

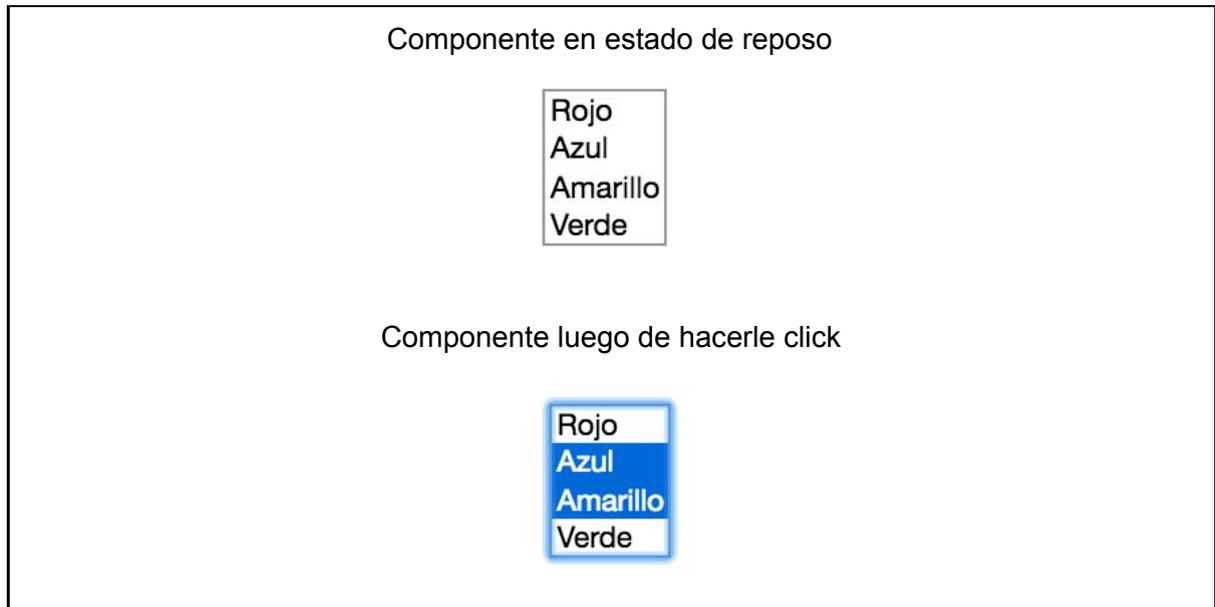


Figura 11.20: Representación visual de un campo de selección de opciones.

```
<select multiple>
  <option value="rojo">Rojo</option>
  <option value="azul">Azul</option>
  <option value="amarillo">Amarillo</option>
  <option value="verde">Verde</option>
</select>
```

Sección de Código 11.25: Código html correspondiente a un campo de selección de opciones múltiples.

Operaciones manuales admitidas:

- Click sobre el campo utilizando el mouse, luego hacer click sobre las opciones deseadas.
- Luego de hacer foco en el campo utilizando la tecla TAB del teclado se puede simular el click presionando la tecla ENTER del teclado y seleccionar la opción deseada haciendo uso de las teclas arriba y abajo del teclado.

Nota: Para seleccionar múltiples opciones es necesario presionar la tecla CONTROL o SHIFT del teclado mientras se selecciona las opciones.

Comandos de voz equivalentes a la operación manual:

1. "Click"  
Visualiza una referencia por cada elemento que puede ser presionado.
2. "{indicar referencia correspondiente al campo de selección de opciones}"  
Solicita por pantalla mencionar la opción deseada.

3. "{mencionar las opciones deseadas}"  
Mencionar las opciones deseadas.

## Componente Datalist

El componente visual `datalist` es un complemento para los campos de entrada `input` de tipo `text`. `Datalist` permite visualizar posibles valores predefinidos que se pueden seleccionar para completar el campo de texto de una forma más segura y ágil.



Figura 11.21: Representación visual de un campo `datalist`.

```
<input list="browsers" name="browser">
<datalist id="browsers">
  <option value="Internet Explorer">
  <option value="Firefox">
  <option value="Chrome">
  <option value="Opera">
  <option value="Safari">
</datalist>
```

Sección de Código 11.26: Código html correspondiente a un campo `datalist`.

Rol ARIA relacionado: `combobox`

Operaciones manuales admitidas:

- Click sobre el campo de texto utilizando el mouse y luego escribir el texto deseado. En caso de visualizar sugerencias, se puede hacer click sobre alguna de las opciones para autocompletar el campo de texto.
- Luego de hacer foco en el campo utilizando la tecla TAB del teclado, se puede simular el click presionando la tecla ENTER del teclado e introducir el texto deseado. En caso de visualizar sugerencias, se puede seleccionar una opción haciendo uso de las teclas arriba y abajo del teclado y luego confirmar la opción presionando ENTER.

Comandos de voz equivalente a la operación manual:

1. "Click"  
Visualiza una referencia por cada elemento que puede ser presionado.
2. "{indicar referencia correspondiente al campo de búsqueda de opciones}"  
Solicita por pantalla mencionar la opción deseada.
3. "{indicar los parámetros de búsqueda}"  
Filtra las opciones disponibles.
4. "{indicar la opción deseada}"  
Selecciona la opción deseada.

## Componente Keygen

El componente visual `keygen` similarmente a un `select` representa una lista desplegable que admite la selección de un valor de la lista. Las opciones que presenta este componente son opciones de encriptamiento con las que será enviado el formulario.

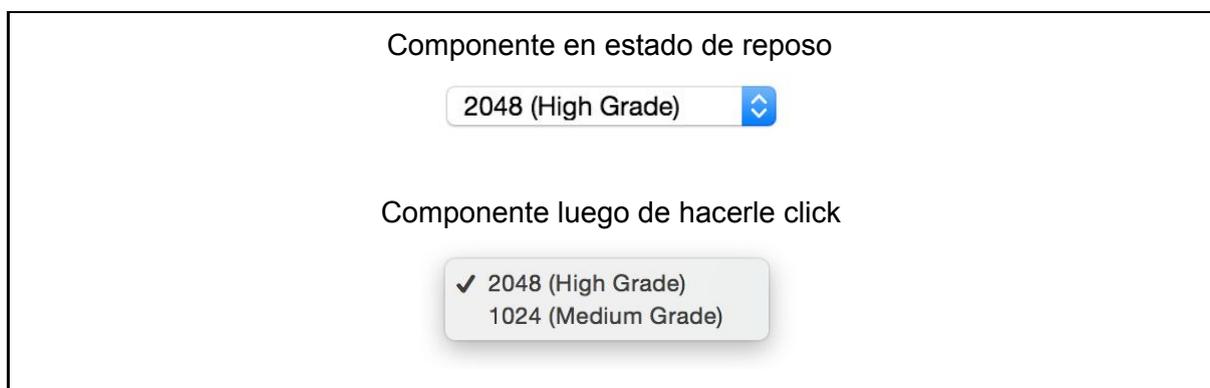


Figura 11.22: Representación visual de un campo de selección de opciones de encriptación.

```
<keygen name="security"/>
```

Sección de Código 11.27: Código HTML correspondiente a un campo de selección de opciones de encriptación.

Operaciones manuales admitidas:

- Hacer click sobre el campo utilizando el mouse, luego hacer click sobre el valor deseado
- Luego de hacer foco en el campo utilizando la tecla TAB del teclado se puede simular el click presionando la tecla ENTER del teclado y seleccionar el valor deseado haciendo uso de las teclas arriba y abajo del teclado.

Comandos de voz equivalentes a la operación manual:

1. "Click"  
Visualiza una referencia por cada elemento que puede ser presionado.
2. "{indicar referencia correspondiente al campo de selección de opciones}"  
Solicita por pantalla mencionar la opción deseada.
3. "{mencionar opción deseada}"  
Selecciona la opción deseada.

## Componente Map

El componente visual `map` representa un grupo de áreas sensibles a las que se les puede hacer click.

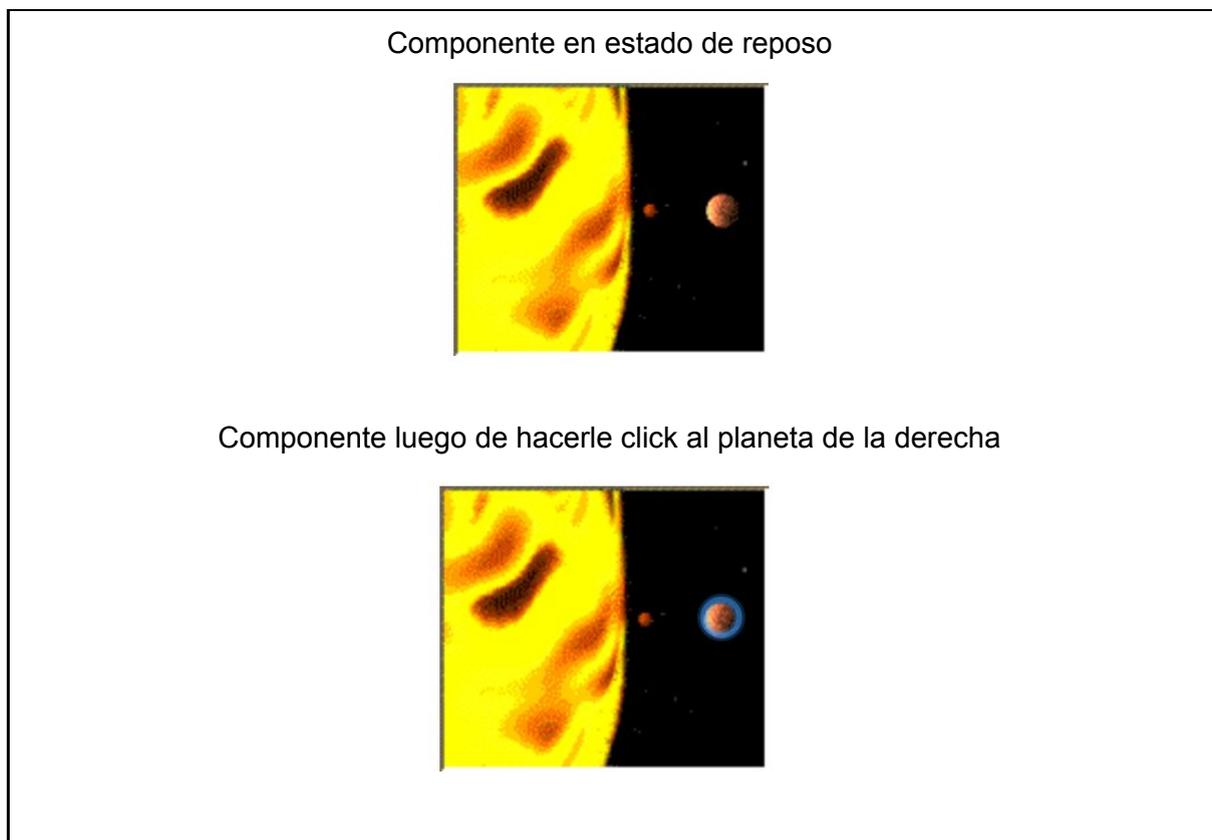


Figura 11.23: Representación visual de un mapa de área sensibles.

```


<map name="planetmap">
  <area shape="rect" coords="0,0,82,126">
  <area shape="circle" coords="90,58,3">
  <area shape="circle" coords="124,58,8">
</map>

```

Sección de Código 11.28: Código html correspondiente a un mapa de área sensibles.

Operaciones manuales admitidas:

- Click sobre el área deseada.
- Presionar la tecla TAB del teclado hasta alcanzar el área deseada y luego se puede simular el click presionando la tecla ENTER del teclado.

Comandos de voz equivalentes a la operación manual:

1. "Click"  
Visualiza una referencia por cada elemento que puede ser presionado.
2. "{indicar referencia correspondiente al mapa de opciones}"  
Solicita por pantalla mencionar la opción deseada.
3. "{mencionar opción deseada}"  
Selecciona la opción deseada.

## Componente Audio

El componente visual `audio` representa un control de reproducción de un archivo de audio.



Figura 11.24: Representación visual de un reproductor de audio.

```

<audio controls>
  <source src="horse.ogg" type="audio/ogg">
  <source src="horse.mp3" type="audio/mpeg">
</audio>

```

Sección de Código 11.29: Código html correspondiente a un reproductor de audio.

Operaciones manuales admitidas:

- Click sobre el botón de play para reproducir
- Click sobre el botón de pausa para pausar la reproducción
- Click sobre la barra de progreso de reproducción para adelantar o retroceder el inicio de reproducción

- Deslizar el botón de estado de reproducción para adelantar o retroceder el inicio de reproducción
- Click sobre el botón de sonido para silenciar
- Click sobre la barra de volumen para aumentar o disminuir el volumen
- Deslizar el botón de estado de volumen para aumentar o disminuir el volumen

Comandos de voz equivalentes a la operación manual:

1. “Click”  
Visualiza una referencia por cada elemento que puede ser presionado.
2. “{indicar referencia correspondiente al reproductor de audio}”  
Solicita por pantalla mencionar la opción deseada.
3. Mencionar alguno de los siguientes comandos:
  - a. “Reproducir”
  - b. “Detener”
  - c. “Pausar”
  - d. “Adelantar {tiempo}”
  - e. “Retroceder {tiempo}”
  - f. “Silenciar”
  - g. “Aumentar Volumen {porcentaje}”
  - h. “Disminuir Volumen {porcentaje}”

## Componente Video

El componente visual `video` representa un control de reproducción de un archivo de video.



Figura 11.25: Representación visual de un reproductor de video.

```
<video controls>
  <source src="movie.mp4" type="video/mp4">
  <source src="movie.ogg" type="video/ogg">
</video>
```

Sección de Código 11.30: Código html correspondiente a un reproductor de video.

Operaciones manuales admitidas:

- Click sobre el botón de play para reproducir
- Click sobre el botón de pausa para pausar la reproducción
- Click sobre la barra de progreso de reproducción para adelantar o retroceder el inicio de reproducción
- Deslizar el botón de estado de reproducción para adelantar o retroceder el inicio de reproducción
- Click sobre el botón de sonido para silenciar
- Click sobre la barra de volumen para aumentar o disminuir el volumen
- Deslizar el botón de estado de volumen para aumentar o disminuir el volumen

Comandos de voz equivalentes a la operación manual:

1. "Click"  
Visualiza una referencia por cada elemento que puede ser presionado.
2. "{indicar referencia correspondiente al reproductor de audio}"  
Solicita por pantalla mencionar la opción deseada.
3. Mencionar alguno de los siguientes comandos:
  - a. "Reproducir"
  - b. "Detener"
  - c. "Pausar"
  - d. "Adelantar {tiempo}"
  - e. "Retroceder {tiempo}"
  - f. "Silenciar"
  - g. "Aumentar Volumen {porcentaje}"
  - h. "Disminuir Volumen {porcentaje}"

## Componente Vínculo

El componente visual `a` representa un link al cual se le puede hacer click. El contenido del vínculo puede ser cualquier otro componente visual, por ejemplo una imagen o un párrafo.



Figura 11.26: Representación visual de un vínculo.

```
<a href="http://www.wikipedia.com">Wikipedia</a>
```

Sección de Código 11.31: Código html correspondiente a un vínculo.

Rol Aria relacionado: `link`, `menuitem`

Operaciones manuales admitidas:

- Click sobre el link utilizando el mouse
- Luego de hacer foco en el link utilizando la tecla TAB del teclado se puede simular el click presionando la tecla ENTER del teclado

Comandos de voz equivalentes a la operación manual:

1. "Click"  
Visualiza una referencia por cada elemento que puede ser presionado
2. "{indicar referencia correspondiente al vínculo que se desea presionar}"  
Ejecuta la acción de click sobre el elemento referenciado.

# Anexo 3

## Evaluación de usabilidad de la herramienta

### Instrucciones iniciales para el evaluador

Es importante que el entrevistado complete toda la entrevista, la cual lleva un tiempo estimado de cuarenta y cinco minutos.

Además de ingresar la información requerida por la evaluación, mientras se entrevista al usuario, es necesario que se grabe la pantalla de la computadora donde se realiza la prueba y el sonido de la sala.

Con el objetivo de introducir a la persona en la temática de la herramienta y reducir las dudas del mismo acerca de la evaluación, se solicita informar al entrevistado la problemática de interacción manual que presenta la navegación actual y la alternativa propuesta por la herramienta.

### Ficha de Evaluación

#### Datos de la evaluación

- Fecha
- Nombre del entrevistador

#### Datos del entrevistado

- Nombre y apellido
- Ocupación
- Edad
- Idioma nativo
- Idioma secundario

#### Preguntas Iniciales

¿Con qué frecuencia accede a internet?

Cuando accede a internet, ¿qué sitios web visita regularmente?

¿Qué navegador web utiliza para acceder a internet?

¿En algún momento percibió limitaciones o restricciones físicas a la hora de interactuar con la computadora?

## Ejercicios de prueba

### Instalación manual de la herramienta

1. Abrir el navegador Google Chrome
2. Acceder al sitio web [www.handsfreeforweb.com](http://www.handsfreeforweb.com)
3. Instalar la aplicación Handsfree for Web

### Configuración de la aplicación

1. Identificar el icono correspondiente a la aplicación instalada en la barra de navegación.
2. Abrir el panel de administración de la herramienta.
3. Elegir el idioma correspondiente.
4. Seleccionar el modo de interacción “Presionar Ctrl antes de decir cada comando” .

### Uso de la herramienta utilizando comandos de voz

*Prueba de comandos que asisten al usuario en caso de necesitar ayuda*

1. Expresar el comando de voz “Ayuda” y visualizar todos los comandos de voz disponibles.
2. Indicar el comando “Salir”.
3. Mencionar el comando de voz “Barra de ayuda”.

*Prueba de navegación web*

Acceder a los diversos sitios web con los que regularmente interactúa.

Opcionalmente se proponen las siguientes interacciones:

Reproducción de contenido multimedia

1. Abrir un sitio de videos en una nueva pestaña.
2. Realizar la búsqueda de un video de interés.
3. Abrir el video.
4. Seleccionar el video.
5. Pausar el video.
6. Volver a reproducir el video.
7. Silenciar el video.
8. Agregar a favoritos el sitio web actual.

Prueba de interacción con campos de formulario

1. Ir a la pestaña izquierda.
2. Abrir el sitio web [www.handsfreeforweb.com](http://www.handsfreeforweb.com)
3. Invocar el comando “sección evaluación”.

4. Completar el formulario con tus datos personales.

#### Prueba de interacción con funcionalidad del navegador

1. Acceder a la lista de favoritos.
2. Abrir el sitio de videos visitado anteriormente.
3. Ver el historial.
4. Abrir el sitio de noticias al cual se accedió recientemente.
5. Ver las descargas realizadas hasta el momento.
6. Consultar los sitios web más visitados.
7. Abrir el sitio web más visitado.
8. Cerrar pestaña.

#### Preguntas acerca de la experiencia de usuario

¿Cómo calificaría los pasos de instalación de la herramienta?

- Simples
- Complejos
- Extremadamente complejos

¿Cómo valoraría la ayuda brindada por la herramienta a la hora de navegar la web?

- Muy buena
- Buena
- Regular
- Mala

¿Cómo considera que la información fue presentada por la herramienta?

- Apropiada
- Confusa
- Insuficiente
- Innecesaria
- Invasiva

¿Qué opinión tiene de la velocidad con que la aplicación ejecuta el comando de voz?

- Adecuada
- Lenta
- Extremadamente lenta

¿Estaría interesado en ser informado acerca de los mecanismos por los cuales usted puede expandir el conjunto de comandos de voz que pueden ser ejecutados?

¿Considera que la herramienta puede ser una alternativa viable a los problemas de interacción manual?

¿Cómo consideraría la experiencia de uso de la aplicación?

# Bibliografía

- [1] Introduction to Web Accessibility. <http://webaim.org/intro/>. Consultado el 28/05/2017
- [2] INDEC. Población con dificultad o limitación permanente. Censo Nacional de Población, Hogares y Viviendas 2010. Octubre 2014
- [3] KAREN WARD, DAVID NOVICK. Hands-Free Documentation. University of Texas at El Paso; 2003
- [4] Internet Users. <http://www.internetlivestats.com/internet-users/>. Consultado el 28/05/2017
- [5] JULIO BORRÁS, PABLO BERNABEU SOLER. Accesibilidad de usuarios con discapacidad motora severa a un sitio web educativo. 2006
- [6] Speaking in Context: Designing Voice Interfaces. <http://punchcut.com/perspectives/speaking-context-designing-voice-interfaces/>. Consultado el 28/05/2017
- [7] J. H. M. DANIEL JURAFSKY. Speech and Language Processing, An introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Prentice Hall, Upper Saddle River, New Jersey 07458, 2000.
- [8] BERTHA MÁRQUEZ, JOSÉ ZULAICA. Implementación de un reconocedor de voz gratuito a el sistema de ayuda a invidentes Dos-Vox en español. Universidad de las Américas Puebla. Enero 2004.
- [9] MARKUS FORSBERG. Why is Speech Recognition Difficult?. Chalmers University of Technology. Febrero 2003.
- [10] ANDREW WILLIAMS. Cortana vs Google Now vs Siri: Which is best?. <http://www.trustedreviews.com/opinions/siri-vs-google-now-vs-cortana>. Consultado el 28/05/2017
- [11] IWAMIDA, YAMATO, SAKIYAMA, KITAO, TAKASHIMA. Voice Recognition Technology for Car-Mounted Devices. Fujitsu Ten Tech. 1999
- [12] How to use Speech Recognition in Windows. <https://support.microsoft.com/en-us/help/14213/windows-how-to-use-speech-recognition>. Consultado el 28/05/2017
- [13] Use your voice to enter text on your Mac. <https://support.apple.com/en-us/HT202584>. Consultado el 28/05/2017

- [14] Dragon NaturallySpeaking, Get more done by voice. <https://www.nuance.com/dragon/dragon-for-pc/premium-edition.html>. Consultado el 28/05/2017
- [15] Using Opera with your voice. <http://www.opera.com/help/tutorials/voice/>. Consultado el 28/05/2017
- [16] Type with your voice. <https://support.google.com/docs/answer/4492226?hl=en>. Consultado el 28/05/2017
- [17] What are extensions?. <https://developer.chrome.com/extensions>. Consultado el 28/05/2017
- [18] Web Speech API Specification. <https://dvcs.w3.org/hg/speech-api/raw-file/tip/speechapi.html>. Consultado el 28/05/2017
- [19] The Speech Synthesis Interface. <https://dvcs.w3.org/hg/speech-api/raw-file/tip/speechapi.html#tts-section>. Consultado el 28/05/2017
- [20] Google Cloud Speech API, Language Support. <https://cloud.google.com/speech/docs/languages>. Consultado el 28/05/2017
- [21] Lawrence Philips' Metaphone Algorithm. <http://aspell.net/metaphone/>. Consultado el 28/05/2017
- [22] RODRIGUEZ, ÁLVAREZ, BRAVO. Coeficientes de Asociación. Plaza y Valdés. Marzo 2001
- [23] Font Awesome. <http://fontawesome.io/icons/> Consultado el 28/05/2017
- [24] Mobile and tablet internet usage exceeds desktop for first time worldwide. <http://gs.statcounter.com/press/mobile-and-tablet-internet-usage-exceeds-desktop-for-first-time-worldwide> Consultado el 28/05/2017
- [25] Speech Recognition API. <http://caniuse.com/#feat=speech-recognition>. Consultado el 28/05/2017
- [26] Visual Disabilities. <http://webaim.org/articles/visual/blind#limitations>. Consultado el 28/05/2017