



XULIA



Antonio Losada González

Xestión Unificada da Linguaxe con Intelixencia Artificial.

Sistema de control integral del ordenador por voz especialmente diseñado para personas con movilidad muy reducida.

Pontevedra

4 de noviembre, 2015

Licencia GPL

Versión 1.0

Tabla de contenidos

Introducción a Xulia	3
Características principales.....	3
Pantalla de estado de Xulia	7
Zona del micrófono	7
Zona de barra de precisión.....	7
Cuadro de comando reconocido.....	7
Cuadro de clase de aplicación activa	7
Cuadro de modo activo	8
Cuadro de modificadores.....	8
Zona del multiplicador	8
Configuración de Xulia	9
Zona de variables	9
Zonas de Modos.....	10
Modos de operación	10
Modo dictado Windows.....	10
Modos dictado Google	11
Gramáticas de aplicación	11
Gramáticas	12
Edición de comandos	12
Listas de sustitución	16
Ayuda	18
Búsqueda de comandos	19
Entidad conversacional de I.A. Xulia	20

Introducción a Xulia

Xulia es un interface de control integral del ordenador por voz.

Está diseñado para interactuar tanto con el propio sistema operativo como con cualquier aplicación simulando el comportamiento del teclado y del ratón a bajo nivel

Su base de datos de comandos es totalmente configurable por el usuario así como el control individual de la precisión de reconocimiento de cada uno de sus comandos de voz

Es multilingüaje, de modo que cada usuario podrá configurar su base de datos de comandos de interacción en cualquier idioma

Xulia no ha sido diseñada para complementar o facilitar los interfaces estándar de ratón y teclado, sino que se ha diseñado para sustituir a estos en su totalidad. Las características incluidas han sido incorporadas para permitir que una persona tetrapléjica pueda manejar el ordenador con una productividad similar a cualquier otra persona sin limitaciones.

Características principales

Su base de datos está escrita en texto plano con formato XML con lo que puede ser compartida entre usuarios total o parcialmente

Todos sus comandos están agrupados en unidades lógicas llamadas gramáticas y estas a su vez se agrupan en modos de operación

La transición entre modos de operación se asocia a los comandos de voz y es totalmente configurable en su fichero de configuración.

En cada modo de operación hay un número de gramáticas activas, esto implica que sólo se podrán ejecutar los comandos contenidos en las gramáticas asociadas al modo activo. De esta forma discriminados y reducimos los falsos positivos, ya que sólo se encuentran activos los comandos de las gramáticas del modo activo.

Por otra parte el sistema soporta gramáticas de aplicación solapadas. Xulia identifica en todo momento la aplicación activa y busca si dentro de su configuración hay una gramática de aplicación para ella. En caso de existir, activa las gramáticas de aplicación de la aplicación activa que sólo permanecerán activas mientras la aplicación permanezca en primer plano.

Con esta funcionalidad, además de disponer de un sistema solapado de activación de comandos, podremos tener comandos de voz iguales que realicen tareas distintas dependiendo de la aplicación activa.

Por poner un ejemplo, las teclas aceleradores para abrir un fichero son distintas en Word y en notepad, pero el comando puede ser el mismo, "abrirfichero"

Xulia soporta la creación de modos estándar incompatibles con la activación automática de las gramáticas de aplicación.

En Xulia cada comando de voz se asocia a una macro. Una macro está representada por una secuencia de instrucciones. Para Xulia una instrucción puede ser un comando con o sin modificadores, un cambio de modo, o el envío de una secuencia de teclas, para lo cual tiene un lenguaje especial que permite enviar estas pulsaciones con combinación simultánea de las teclas modificadores shift, control o alt.

Adicionalmente Xulia dispone de más de 140 comandos que permiten colocar el ratón en cualquier lugar de la Pantalla, ejecutar aplicaciones, activar aplicaciones, simular pulsaciones de teclas de teclado y del ratón, etc.

El sistema de activación de comandos se ha diseñado para intentar evitar los falsos positivos, que son muy comunes al emplear el sistema de reconocimiento de voz de Windows con gramáticas abiertas de dictado. Con este sistema de activación selectivo conseguimos que el número de comandos activos en cada momento sea únicamente el necesario, de modo que no se lance la ejecución de ninguna instrucción no deseada. Para conseguir este fin, el sistema tiene establecidos unos límites de precisión de reconocimiento configurables de modo genérico separados para las gramáticas de dictado continuo y las gramáticas de comandos. La experiencia nos muestra que varios comandos son más difíciles de reconocer que otros e incluso que hay comandos que pueden ser confundidos fácilmente con ruido de fondo. Para evitar estos problemas, Xulia permite especificar de modo individual la precisión de reconocimiento mínimo de cada uno de los comandos de voz para que sean reconocidos.

Además del modo de reconocimiento de comandos, Xulia tiene dos modos avanzados de dictado continuo.

Uno de ellos emplea la gramática de dictado continuo de Windows y el otro emplea el API javascript de reconocimiento continuo de voz implementado en Google Chrome.

En el momento de creación de Xulia, esta es la única aplicación que emplea de modo indirecto el API javascript de reconocimiento continuo de voz para realizar el reconocimiento de voz dentro de Windows.

Empleando llamadas ajax periodicas, la información reconocida desde los servidores de Google y recuperada en el navegador con javascript es transferida a través de un servlet corriendo sobre un servidor tomcat que finalmente la comunica a xulia.

Mediante este conjunto de componentes es posible disfrutar de dictado continuo dentro de Windows empleando la tecnología de reconocimiento de voz de Google.

Comparando ambas tecnologías de reconocimiento, personalmente, considero que el funcionamiento de la tecnología de reconocimiento de Google es algo superior a la de Microsoft, sobre todo en lo relacionado con la inmunidad al ruido y a los falsos positivos, aunque el desempeño del sistema de reconocimiento continuo de voz de Microsoft es muy bueno.

El empleo de esta nueva tecnología resulta indispensable en los casos en que el idioma empleado por el usuario no forme parte de uno de los seis que soporta el reconocimiento de voz de Microsoft, que son el inglés, francés, alemán, español, chino y japonés.

El motivo de la creación de Xulia ha sido mejorar la productividad en el uso del PC de mi primo brasileño Ricardo que padece tetraplejia. Inicialmente teníamos el inconveniente de que Windows no soporta el portugués. Con esta nueva tecnología podemos emplear un navegador Google Chrome como un servidor de reconocimiento de voz en dictado continuo.

Como un añadido el último componente presente en Xulia es un sistema conversacional de inteligencia artificial que tiene como base un cerebro escrito en español empleando el lenguaje AIML. Esta entidad conversacional permite realizar una interacción en lenguaje natural.

Xulia es un sistema modular que ha sido posible por el esfuerzo desinteresado de mucha gente que ha puesto su trabajo en Internet a disposición de los demás, además de emplear los avances tecnológicos de empresas como Google o Microsoft.

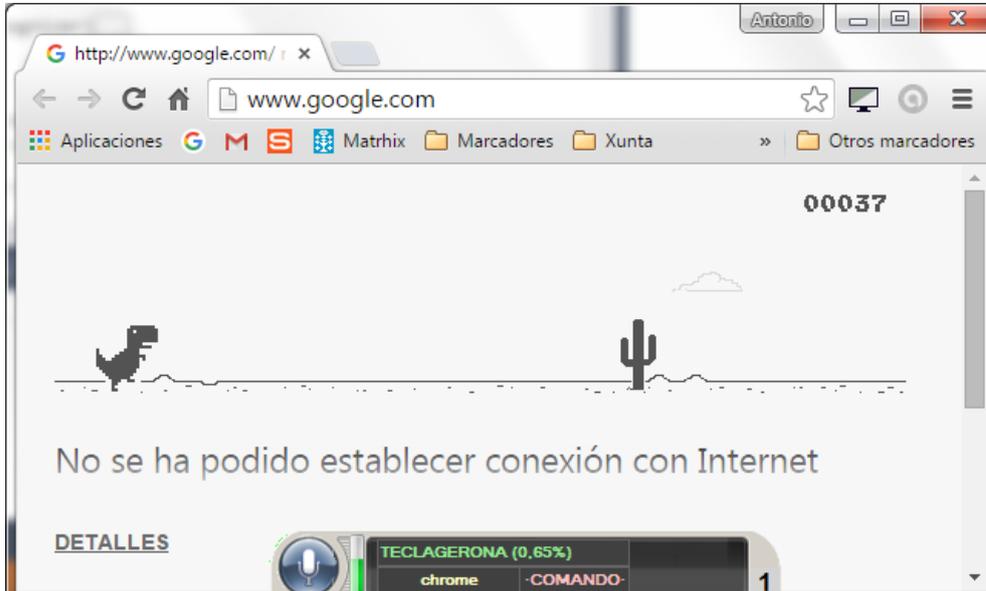
La suma de todos estos componentes ha dado vida a Xulia.

El acrónimo de Xulia corresponde con Xestión Unificada da Linguaxe con Intelixencia Artificial que hace referencia al sistema de reconocimiento y procesamiento del lenguaje del bot conversacional empleando el lenguaje AIML de inteligencia artificial para el procesamiento de secuencias de lenguaje natural.

El acrónimo está escrito en gallego porque el nombre original de Xulia es gallego y ha sido escogido para este proyecto porque es el nombre de mi hija.

Pantalla de estado de Xulia

La parte más visible de Xulia es su pantalla de estado. Esta pantalla está siempre visible en primer plano y muestra toda la información de estado de Xulia.



Esta pantalla está compuesta por varias zonas.

Zona del micrófono

Permite desactivar y activar Xulia con un click del ratón encima del micrófono.

También visualiza un reloj de arena mientras Xulia está ocupada procesando algún comando.

Zona de barra de precisión

La pantalla muestra una barra de estado vertical con efecto líquido dentro que visualiza la precisión del último comando reconocido. Para no tener problemas con los usuarios daltónicos o con deficiencias visuales, la barra no cambia de color cuando el comando no se reconoce, sino que desaparece.

Cuadro de comando reconocido

El cuadro de texto superior informa del último comando reconocido. Al lado del comando aparece el grado de precisión de reconocimiento del comando.

El comando se visualiza en color verde para indicar que ha sido reconocido con precisión suficiente para ser ejecutado, en rojo cuando ha sido reconocido un comando, pero la precisión es insuficiente y en amarillo claro cuando ha sido reconocido con precisión suficiente, pero no se ejecutará porque pertenece a una gramática cargada pero no activa.

Cuadro de clase de aplicación activa

El cuadro inferior izquierdo contendrá la clase de aplicación activa.

En caso de existir una gramática para esta clase de aplicación y el modo actual ser compatible con gramáticas de aplicación, esta gramática será activada de modo solapado con las gramáticas activas actuales.

Cuadro de modo activo

El cuadro inferior derecho indica el modo activo. Según la configuración de Xulia, cada modo tiene asociadas las gramáticas que se activarán y cada gramática tiene asociados sus comandos.

Cuadro de modificadores

La zona derecha de la pantalla de estado está reservada para las marcas de modificadores.

El sistema admite los siguientes modificadores



Primera fila de flags:

P: Se ha activado alguna pulsación de una tecla que no ha sido liberada

A: Activa la pulsación conjunta de ALT

C: Activa la pulsación conjunta de CONTROL

S: Activa la pulsación conjunta de SHIFT

Segunda fila de flags:

A: Aviso activo (Se activa por comandos)

A: Acento normal, grave o circunflejo

S: Selección activa

M: Mayúsculas activas

Zona del multiplicador

En la esquina derecha de la pantalla de estado siempre hay un número. Este número corresponde al multiplicador. Este dato se cambia con los comandos de voz de multiplicación. El siguiente comando de escritura o de movimiento se realizará tantas veces como indique el multiplicador. Por ejemplo, si decimos POR TRES y DERECHA, el cursor del ratón se moverá tres posiciones a la derecha.

Configuración de Xulia

La pantalla de configuración está dividida en varias solapas:

Variables: Zona de variables de configuración

Modos: Zona de modos, Definición de gramáticas de aplicación, Definición de gramáticas de idiomas y listas de sustitución.

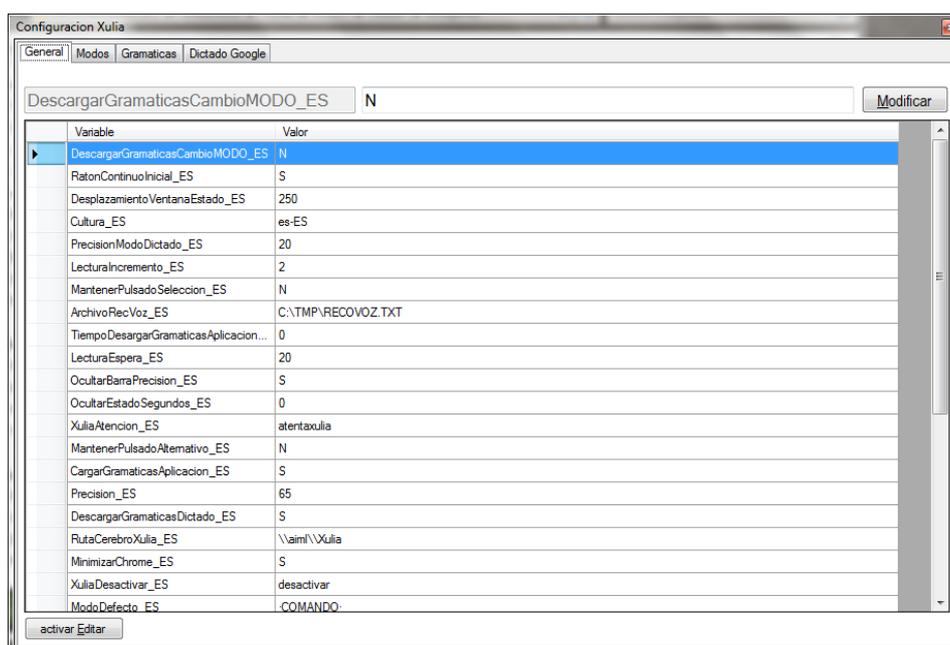
Gramáticas: Zona de especificación de comandos de gramáticas

Dictado Google: Zona de especificación de las listas de sustitución en modos de dictado continuo empleando tecnología Google.

Todas las solapas contienen un grid, varios cuadros de edición y varios botones. Todos los botones excepto los de eliminar secciones enteras o listas de sustitución enteras tienen aceleradores. La tecla aceleradora está señalada dentro del texto del botón con una letra mayúscula. Una vez dentro de la solapa para activar el botón con la tecla aceleradora, solo hay que pulsar ALT + la tecla aceleradora. De esta forma es muy sencillo navegar por la pantalla de configuración con el ratón mediante Xulia. Para cambiar de solapa podemos emplear los comandos “siguientcsolapa” y “anteriorsolapa”

En las siguientes líneas pasamos a describir cada una de las solapas.

Zona de variables

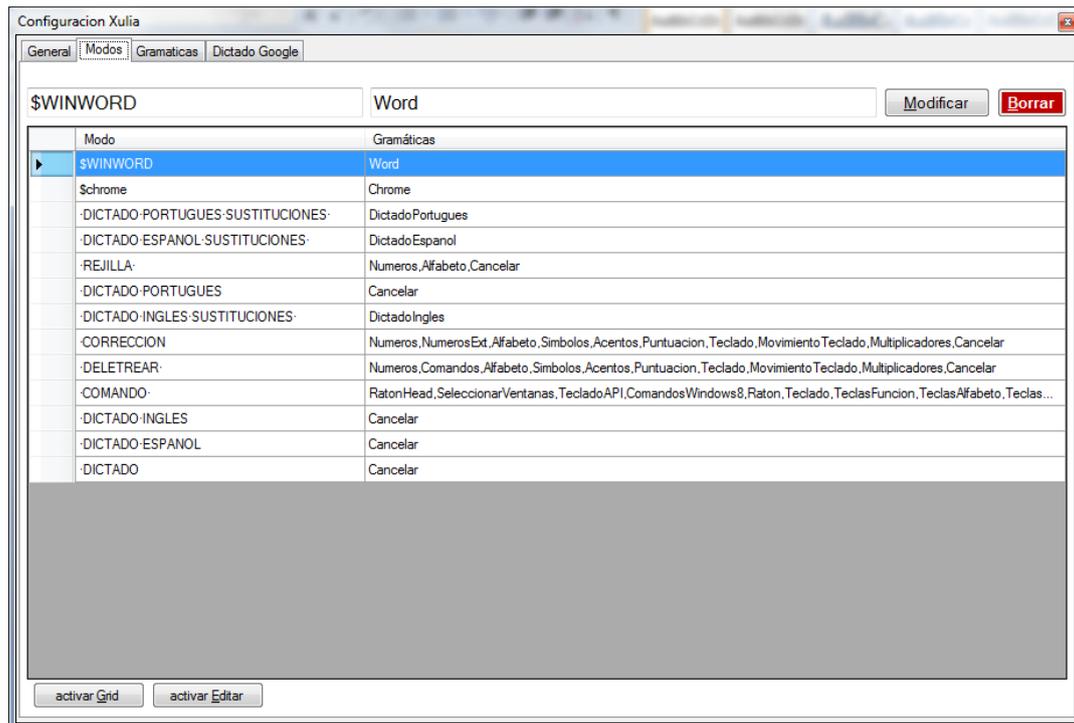


Esta zona contiene todas las variables globales de configuración del sistema.

Para modificar una variable se pulsa encima de la misma para que suba a los cuadros de edición. Después se modificará y se pulsa en Modificar.

Cada vez que el cursor avanza y cambia de línea, la variable seleccionada sube a los cuadros de edición

Zonas de Modos



Modos de operación

Esta solapa contiene los modos de operación, los nombres de las listas de sustitución, las gramáticas de aplicación y los modos de dictado.

Todos los modos deben comenzar por el símbolo ‘.’ (el punto medio). De forma estándar para los modos definidos por el usuario, en caso de que cuando se active un modo queramos que se activen automáticamente las gramáticas de aplicación, el modo debe terminar por ‘.’. Si no termina por este símbolo, aunque se active una aplicación con gramática de aplicación, ésta no se activará. Podemos ver que el modo ·DICTADO no termina por ‘.’. Esto quiere decir que no soporta gramáticas de aplicación, ya que estos comandos serían reconocidos dentro del dictado.

Modo dictado Windows

El modo de dictado con tecnología Windows debe llamarse ·DICTADO y contendrá las gramáticas de comandos activas en modo dictado Windows.

Modos dictado Google

Los modos de dictado con tecnología Google tendrán dos submodos, el que define las gramáticas activas en el modo y el que define el nombre de la lista de sustitución. Estos modos serán:

·DICTADO·idioma·SUSTITUCION·

·DICTADO·idioma

La primera contiene el nombre de la listas de configuración y sustitución para el idioma indicado. La palabra idioma debe sustituirse por el nombre del idioma, p.e. ESPANOL, PORTUGUES, etc. Solo podemos emplear letras del alfabeto inglés, sin ñ, ç o acentos.

Normalmente en estos modos la única gramática activa es Cancelar que nos permite salir del modo. Podríamos tener activas otras gramáticas de comandos, pero esto implicaría que probablemente estos comandos sean reconocidos por el procesador de lenguaje de Google Chrome, con lo que tendríamos que jugar con las listas de sustitución para eliminarlos del dictado.

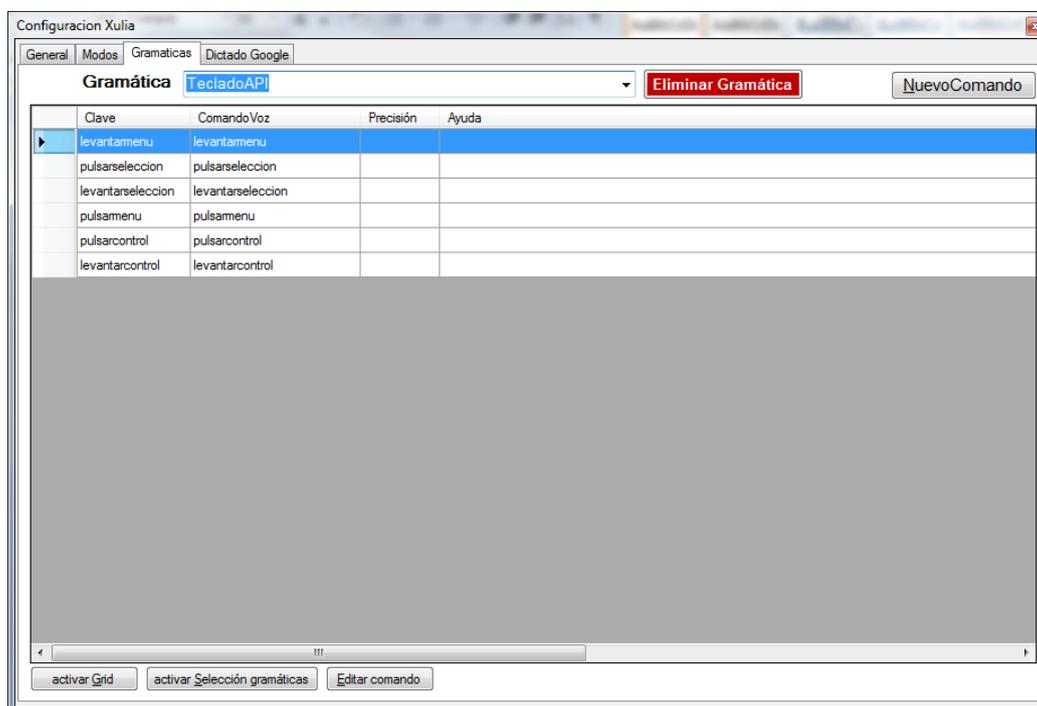
Cada modo representa una agrupación de gramáticas y puede ser activado desde un comando. Cuando un modo está activo, solo se reconocen los comandos de las gramáticas indicadas en el modo.

Todos los modos deben estar en mayúsculas y deben comenzar por el carácter “.” (Punto medio)

Gramáticas de aplicación

Las gramáticas de aplicación deberán comenzar por el símbolo \$ y contendrán la lista de gramáticas que se activarán cuando Xulia detecte que una determinada aplicación se está utilizando en primer plano. El nombre de la gramática está compuesta por el símbolo \$ seguido de la clase de aplicación.

Gramáticas



La solapa de gramáticas permite visualizar los comandos de cada una de las gramáticas y permite modificar, añadir o quitar estos comandos. También permite eliminar gramáticas completas.

Para visualizar una gramática, primero debemos seleccionar la gramática en el desplegable. Para abrir el desplegable puede emplearse la combinación ALT-<Tecla abajo>. Después de realizar este paso la gramática se carga en la tabla. Podemos emplear las teclas aceleradoras de los botones para acceder a la tabla. Una vez dentro podemos movernos hacia arriba y hacia abajo para buscar el comando que queremos editar.

Una vez seleccionado el comando, para editarlo se puede teclear “e” o enviar el comando “teclaespaña” o bien pulsar con el ratón.

Edición de comandos

La pantalla de edición de comandos tiene los siguientes campos:

Clave: Clave asignada al comando, normalmente será igual al propio comando

Comando de voz: Comando de voz que reconocerá Xulia para ejecutar la macro

Precisión: Precisión mínima de reconocimiento del comando. Si lo dejamos vacío cogerá la precisión estándar

Ayuda: Ayuda o descripción del comando para la pantalla de ayuda.

Macro: Macro que ejecutará Xulia al escuchar el comando

La macro está compuesta por una secuencia de instrucciones. Los botones Subir y Bajar nos permiten reordenar esta secuencia.

Zona de comando:

La zona de comando está compuesta por un desplegable y un cuadro de texto.

Cada instrucción podrá estar compuesta por un comando con modificadores o por una secuencia de pulsaciones de teclado en caso de no haber seleccionado ningún comando.

Para borrar el comando seleccionado pulsaremos el botón Limpiar.

En caso de escoger un comando, el sistema mostrará la ayuda del comando y los parámetros del mismo que deberemos cubrir para incorporar el comando a la macro. El cuadro de modificadores de comando solo podrá contener +, ^, %, \$ con el siguiente significado:

+ El comando solo se ejecutará si está seleccionado SHIFT

^ El comando solo se ejecutará si está seleccionado CONTROL

% El comando solo se ejecutará si está seleccionado ALT

\$ El comando solo se ejecutará si Xulia está en modo ATENCION

Con los modificadores podemos incrementar la seguridad para impedir la ejecución de ciertos comandos.

En caso de no haber seleccionado un comando podremos emplear el cuadro de modificadores para introducir la secuencia de teclas a enviar.

Podemos introducir cualquier letra o símbolo. Cuando Xulia reconozca el comando de voz asociado simulará la pulsación de cada una de las letras y símbolos como si se hubieran tecleado.

Tendremos las siguientes funcionalidades:

Para especificar caracteres que no se muestran al presionar una tecla, como ENTRAR o TAB y claves que representan acciones en lugar de caracteres, utilice los códigos en la tabla siguiente.

Tecla	Código
RETROCESO	{BACKSPACE}, {BS}, o {BKSP}
INTERRUMPIR	{BREAK}
BLOQ MAYÚS	{CAPSLOCK}
SUPR o SUPRIMIR	{DELETE} o {DEL}

FLECHA ABAJO	{DOWN}
FIN	{END}
ENTRAR	{ENTER} o ~
ESC	{ESC}
AYUDA	{HELP}
INICIO	{HOME}
INS o INSERT	{INSERT} o {INS}
FLECHA IZQUIERDA	{LEFT}
BLOQ NUM	{NUMLOCK}
AV PÁG	{PÁG}
RE PÁG	{PÁG}
IMPRIMIR PANTALLA	{IMPR PANT} (reservado para uso futuro)
FLECHA DERECHA	{RIGHT}
BLOQ DESPL	{BLOQ DESPL}
TAB	{TAB}
FLECHA ARRIBA	{UP}
F1	{F1}
F2	{F2}
F3	{F3}
F4	{F4}
F5	{F5}
F6	{F6}
F7	{F7}
F8	{F8}
F9	{F9}
F10	{F10}
F11	{F11}
F12	{F12}
F13	{F13}
F14	{F14}
F15	{F15}
F16	{F16}
Sumar en el teclado numérico	{AGREGAR}
Restar en el teclado numérico	{RESTAR}
Multiplicar en el teclado numérico	{MULTIPLICAR}
Dividir en el teclado numérico	{DIVISION}

Para especificar teclas con cualquier combinación de las mayúsculas, CTRL y ALT (teclas), preceda el código de tecla con uno o varios de los siguientes códigos.

Tecla	Código
SHIFT	+
CTRL	^

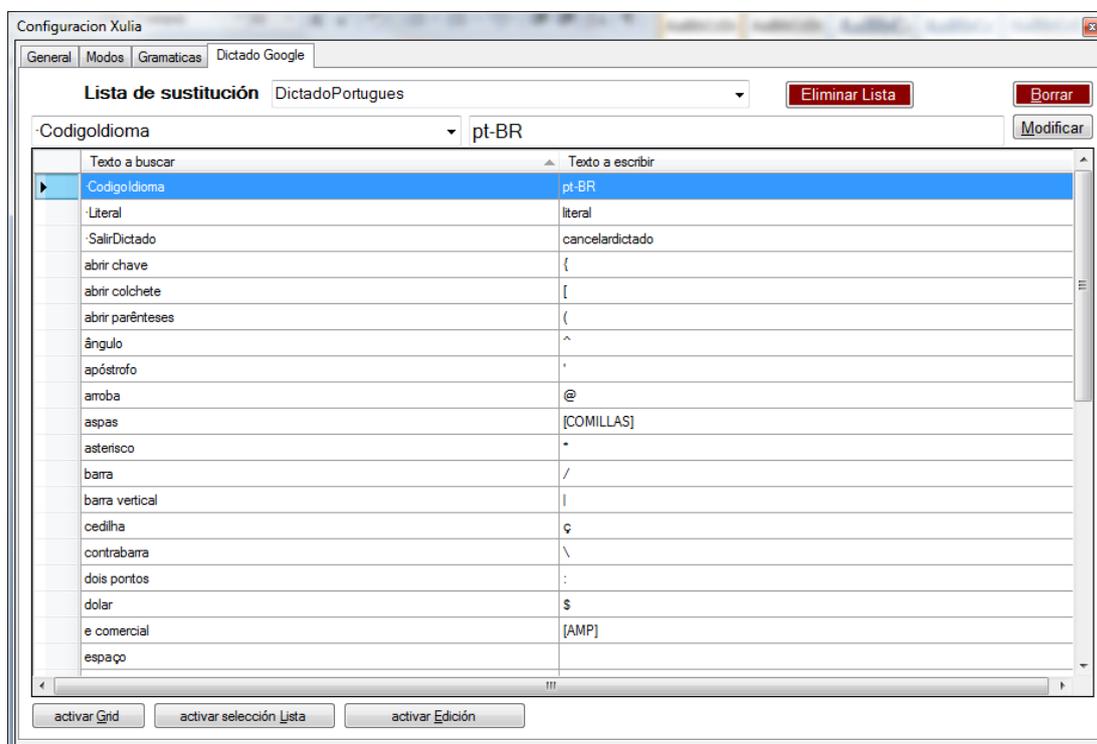
ALT	%
------------	----------

Para especificar que se debe mantener presionada cualquier combinación de mayúsculas, CTRL y ALT mientras se presionan otras teclas, incluya el código de esas claves entre paréntesis. Por ejemplo, para especificar que se mantiene MAYÚS mientras se presione E y C, use “+(EC)”. Para especificar que se mantiene MAYÚS mientras se presiona E, seguido de C sin MAYÚS, use “+EC”.

Para especificar pulsaciones repetidas, utilice el formato {tecla número}. Debe incluir un espacio entre la clave y el número. Por ejemplo, {LEFT 42} significa presionar la tecla FLECHA IZQUIERDA 42 veces; {h 10} significa presione H 10 veces.

Para indicar la pulsación de caracteres reservados, como +, %, ^, etc debe introducirlos entre llaves, por ejemplo para simular la pulsación del símbolo + introduciremos {+}

Listas de sustitución



En modo dictado con tecnología Google, un navegador oculto Google Chrome se encarga del reconocimiento de voz. No es posible interactuar con esta aplicación, por lo que todo lo que se diga será interpretado y Xulia solo puede actuar tras el reconocimiento sustituyendo secuencias de caracteres.

Dentro del destino de una sustitución podemos emplear todos los comandos de las tablas que hemos visto en el apartado de secuencias de pulsaciones de teclas del apartado anterior.

También podemos emplearla esta funcionalidad para colocar mi email, mi dirección, mi teléfono o para colocar símbolos de puntuación en idiomas reconocidos por Google en los que todavía no está incorporada esta funcionalidad como el portugués.

Adicionalmente tenemos dos comandos:

[NUEVA_LINEA], [NUEVA_FRASE] Cuando las colocamos como sustitución de una palabra siempre obligan a que lo siguiente que se dicte se muestre con la primera letra en mayúsculas.

En el caso de [NUEVA_LINEA] incorpora un retorno de línea al último texto dictado.

Todas las listas de sustitución tienen tres cadenas claves que comienzan por ‘.’.

•Codigoldioma

Representa el código de idioma que debe reconocerse y se envía a Google Chrome, los distintos códigos son:

es-ES: Español

en-US: Inglés americano.

pt-BR: Portugués brasileño.

·SalirDictado

Identifica una palabra que al ser reconocida en el texto provoca la salida del modo de dictado continuo

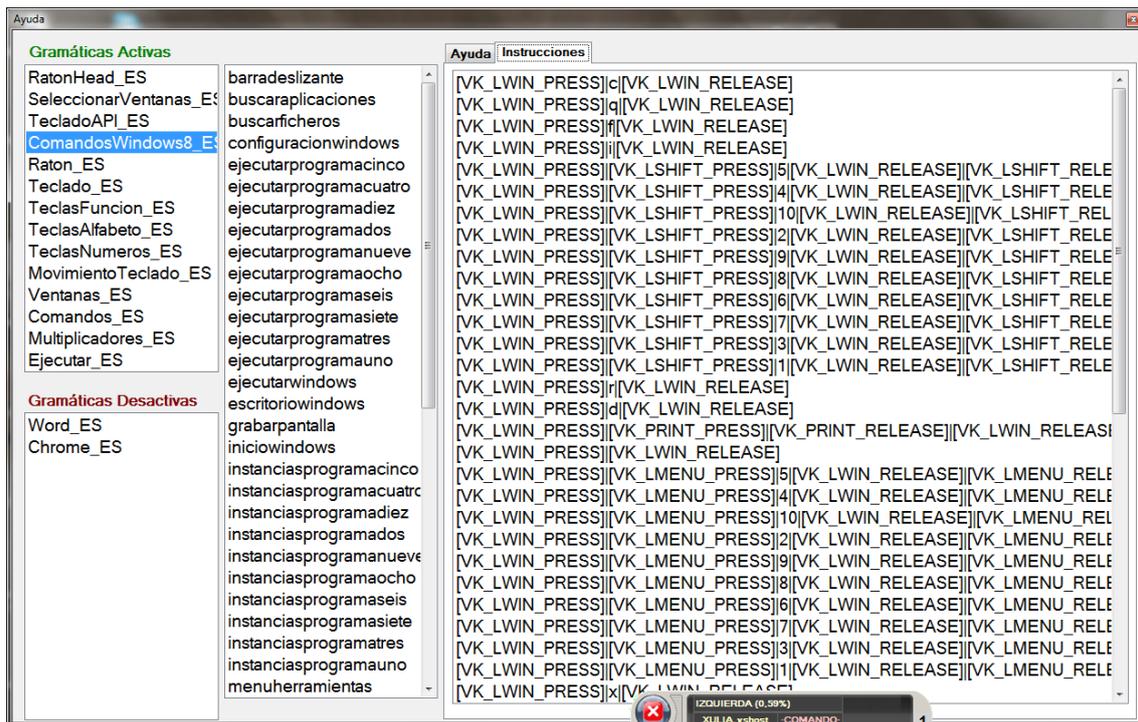
·Literal

Representa el comando para evitar que se sustituyan las secuencias de caracteres de las listas de sustitución. Si la secuencia va precedida de la palabra definida como literal, esta no será sustituida.

Ayuda

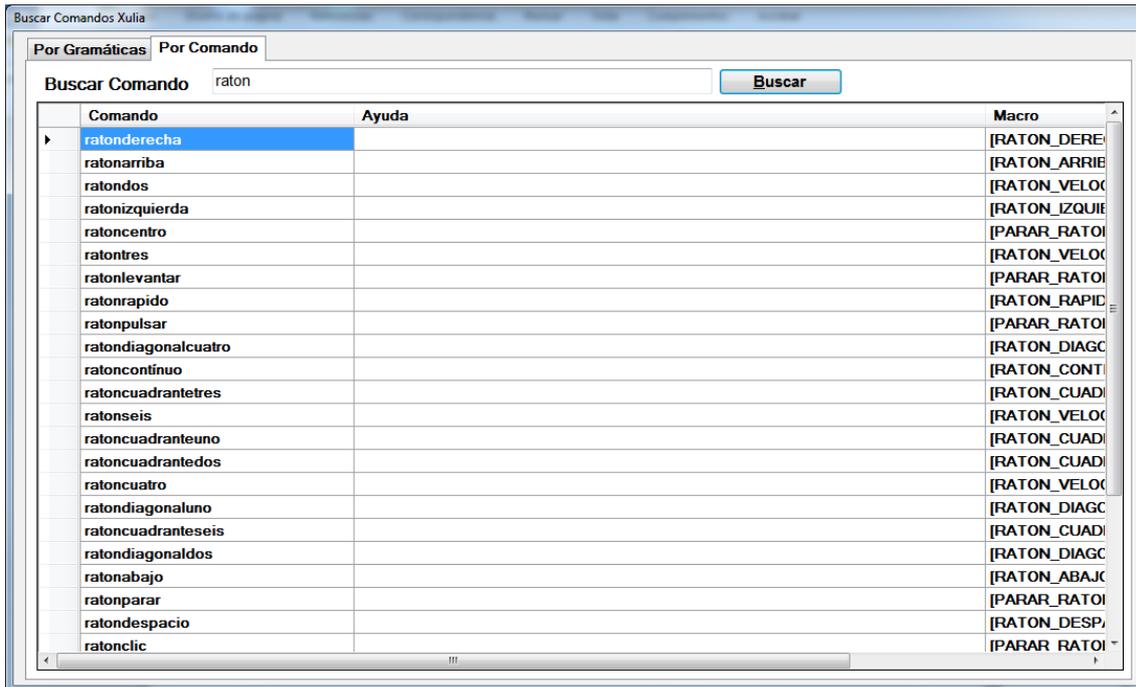
En la pantalla de ayuda se visualizan todas las gramáticas cargadas separadas por activas y no activas.

Al seleccionarlas se visualizan todos los comandos con la información de ayuda y macro asociada



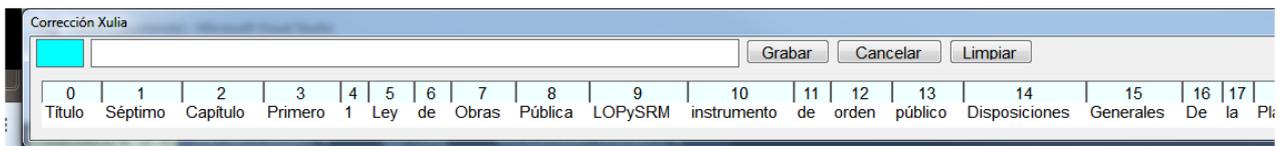
Búsqueda de comandos

Dada que Xulia puede disponer de centenares de comandos, en ocasiones puede ser complicado localizar el comando adecuado. Para ayudar en esta tarea, Xulia tiene la pantalla de búsqueda de comandos, con opción de buscar los comandos de una gramática o buscar los comandos de cualquier gramática que contienen ciertas letras o palabras.



Corrección de errores tipográficos

Xulia tiene una funcionalidad de corrección rápida de errores tipográficos. Al activar el comando que muestra la pantalla de corrección, Xulia recupera la información seleccionada y la distribuye en palabras numeradas, de esta forma es muy fácil identificar la palabra que se quiere corregir.



Al entrar en la pantalla podemos ver la frase dividida en palabras.

Tecleando el número de la palabra y pulsando `<enter>`, la palabra aparece en el cuadro de edición, tras lo que podemos proceder a corregir la palabra.

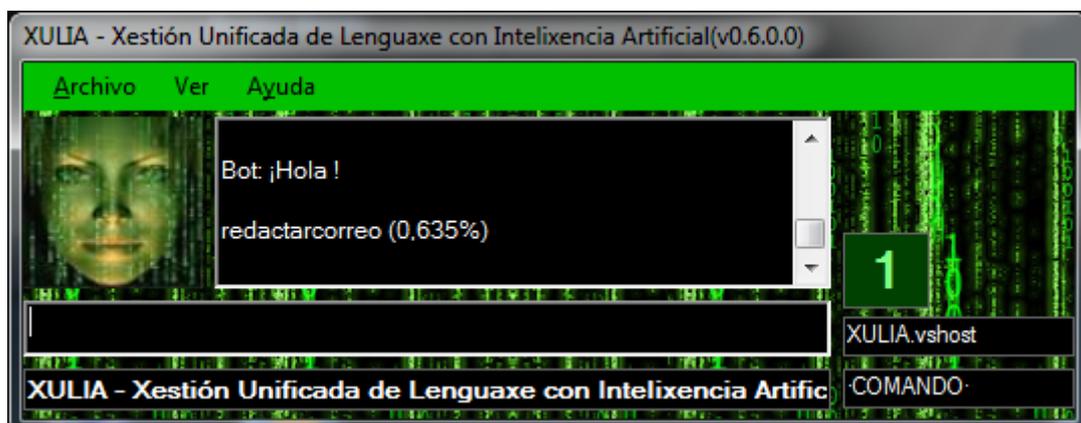
Una vez corregido, si pulsamos `<enter>` la palabra se modifica.

Pulsando el botón Limpiar o `ALT-L` borramos la palabra y pasamos a seleccionar otra,

Pulsando `<ESC>` o `ALT-G` guardamos la frase corregida y la pegamos de nuevo en la misma aplicación desde la que se recuperó

Pulsando el botón cancelar o con `ALT-C` cerramos el cuadro sin realizar ninguna corrección

Entidad conversacional de I.A. Xulia



Este es el módulo de Xulia menos evolucionado. Actualmente presenta una capacidad de conversación muy baja, si bien, esperamos poder ampliar su cerebro en próximas versiones.