

UIMA, entornos y librerías de PLN: Instrucciones de Instalación

Jordi Atserias

{jordi.atserias}@barcelonamedia.org

UIMA, entornos y librerías de PLN es un tutorial práctico de la SEPLN 2011 sobre los distintos entornos y librerías de PLN, especialmente enfocado en UIMA. Para poder seguir mejor este tutorial es recomendable poder asistir con un portátil configurado apropiadamente de manera que sea posible probar sobre el terreno los ejemplos que se exponen.

Esta mini guía explica como instalar UIMA (para los Sistemas operativos más habituales) y es una versión reducida en castellano de la guía en inglés de instalación de UIMA preparada junto con Bart Mellebeek para el ESSLLI 2010 (disponible en <http://esslli2010cph.info/?p=279>).

Si decide utilizar un portátil para seguir mejor el tutorial es **ESENCIAL** que este esté configurado correctamente **ANTES** de que comience el tutorial ya que obviamente no habrá tiempo para poder configurar cada uno de los portátiles de los participantes.

NO WARRANTY Estas instrucciones son una mera ayuda para instalar UIMA y no hay ninguna garantía de que funcionen en todos los casos ni de que no puedan desconfigurar su ordenador. Siga estas instrucciones bajo su propia cuenta y riesgo. No olvide realizar copias de seguridad antes de empezar cualquier tipo de instalación.

Las siguientes secciones proporcionan la información necesaria para configurar su entorno de trabajo. Si las instrucciones no son suficientes o encuentra alguna dificultad puede contactarnos por correo electrónico antes del curso. Una vez comience el tutorial no habrá tiempo para configurar su ordenador.

1 Requerimientos

Para poder instalar UIMA se recomienda un ordenador con:

- hardware
 - un mínimo de 1G memoria.
 - un mínimo de 3G de espacio disco libre ens disco duro.
- software
 - SO: Windows, Linux or Mac OS X.
 - Java versión 1.6.
 - UIMA v. 2.3.0.
 - Eclipse (recomendamos la versión v.3.6 Helios).
 - Navegador Firefox
 - Un editor de ficheros de texto.

Las siguientes secciones contienen instrucciones para configurar su entorno de trabajo.

2 Instrucciones de Instalación

2.1 Instalar Java 1.6

Compruebe que Java v1.6 es la versión de java utilizada por defecto en su ordenador:

```
java -version.
```

En caso de que Java no este instalado o la versión 1.6 no sea la utilizada por defecto, siga las instrucciones de la siguiente sección correspondientes a su sistema operativo.

2.1.1 Instalar java 1.6 en Windows

Descargue java v1.6 desde <http://www.java.com/en/download/manual.jsp>.

- Seleccione Manual download page
- Seleccione Windows 7/XP/Vista/2000/2003/2008 Online¹
- Aparecerá una ventana *File Download* preguntando si ejecutar (*run*) o guardar (*save*) el fichero descargado
 - Si prefiere ejecutar el instalador directamente, pulse en Run.
 - Si por el contrario prefiere guardar el fichero e instalarlo seguidamente, pulse en Save. Escoja la carpeta de su ordenador donde quiere salvar el fichero. Haga doble-click sobre el fichero guardado para empezar la instalación.
- Una vez empezado el proceso de instalación, el instalador muestra una opción para ver las condiciones de la licencia. Pulse el botón *Install* para aceptar los términos de la licencia y empezar la instalación.
- Al instalar java es posible que el instalador muestre la posibilidad de instalar otros productos relacionados. Después de comprobar que los programas que desea instalar están seleccionados, haga click en el botón de *Next* (Siguiente) para continuar la instalación.
- Siga confirmando en las ventanas de diálogo los siguientes pasos de la instalación; Pulse *Close* (Cerrar) en la última ventana.

2.1.2 Instalar java 1.6 en Linux Ubuntu

`sudo apt-get install sun-java6-jre` debería instalar la versión correcta de java.

En caso de que no funcione, descargue la versión autoextraíble de java 1.6 desde <http://www.java.com/en/download/manual.jsp> y ejecute los siguientes comandos:

```
sudo chmod +x jre-6-linux-i586.bin
sudo ./jre-6-linux-i586.bin
(if folder) $sudo mkdir /usr/lib/jvm
sudo mv jre1.6.0 /usr/lib/jvm
sudo update-alternatives --install
/usr/bin/java java /usr/lib/jvm/jre1.6.0/bin/java 1
sudo update-alternatives --set java /usr/lib/jvm/jre1.6.0/bin/java
```

Compruebe que java 1.6 es ahora su versión por defecto: `java -version`

¹Antes de proceder a la instalación, puede ser aconsejable bajar-reconfigurar su cortafuegos *Internet firewall*. Ya que en algunos casos, la configuración por defecto del cortafuegos puede bloquear la instalación en línea de software. Consulte las instrucciones de configuración de su cortafuegos

2.1.3 Instalar java 1.6 en Mac Os

Si existe la carpeta 1.6.* en `/System/Library/Frameworks/JavaVM.framework/Versions`, siga los siguientes pasos desde el punto 2 para configurar java 1.6.* como la versión establecida por defecto. En caso contrario, empiece por el punto 1 instalando java 1.6 desde cero. Si usted está usando una versión de Mac OS anterior a 10.4.11 (incluyendo Darwin) no podrá instalar java 1.6.

1. Instalación manual de Java 1.6:
 - Seleccione el menú apple y luego *Update Software*.
 - Instale Java para Mac Os X 10.5.
2. Para configurar java 1.6 como su versión de java por defecto siga los siguientes pasos:
 - Use `/Applications/Utilities/java-preferences` para cambiar sus preferencias
 - Abra una consola y vaya a `/System/Library/Frameworks/JavaVM.framework/Versions` luego ejecute, `sudo ln -fhsv la_carpeta_con_su_versión_1.6_de_java CurrentJDK`
 - Compruebe que java 1.6 es ahora su versión por defecto: `java -version`

2.2 Instalar UIMA 2.3.0

Para este curso, usaremos UIMA version 2.3.0. UIMA es un proyecto mantenido por *Apache Software Foundation*. Toda la información necesaria acerca de UIMA se encuentra en <http://uima.apache.org/>.

Para instalar UIMA version 2.3.0 siga los siguientes pasos:

- Descargue `uimaj-2.3.0-incubating-bin.tar.gz` desde <http://archive.apache.org/dist/incubator/uima/binaries/uimaj-2.3.0-incubating/uimaj-2.3.0-incubating-bin.tar.gz>
- Compruebe la integridad de su descarga (Ver <http://uima.apache.org/downloads.cgi#VerifyDownload>)
- Descomprima UIMA en una carpeta local `tar xzvf uimaj-2.3.0-incubating-bin.tar.gz`

2.2.1 Variables de Entorno

Es necesario crear y asignar correctamente dos variables de entorno: `UIMA_HOME` y `JAVA_HOME`. Si no está familiarizado con el manejo de las variables de entorno consulte la siguiente sección antes de continuar.

- Asigne a `JAVA_HOME` el directorio donde se encuentra su instalación de Java 1.6 (p.e. `/usr/lib/jvm/java-6-sun-1.6.0.20`).
- Asigne a `UIMA_HOME` el directorio donde instaló la distribución de Apache UIMA (p.e. `/home/$USER/apps/apache-uima`).
- Añada `UIMA_HOME/bin` a su `PATH`.
- Ejecute el comando `UIMA_HOME/bin/adjustExamplePaths.bat` (or `.sh`). Este comando actualizará los caminos (paths) de `UIMA_HOME/examples` para que se puedan ejecutar adecuadamente los ejemplos. Este comando ejecuta un programa en Java; Para eso java deberá estar en su variable `PATH` o la variable `JAVA_HOME` asignada correctamente a su java JRE.

Las siguientes secciones contienen más información sobre como establecer Variables de Entorno en los diferentes sistemas operativos:

Asignación de Variables de entorno en Windows Para ver o cambiar las variables de entorno:

1. Haga click con el botón derecho en Mi ordenador (*My Computer*) y luego seleccione Propiedades (*Properties*).
2. Haga click en Advanced tab.
3. Haga click en *Environment variables*.
4. Haga click en una de las siguientes opciones, *user* o *system variable*:
 - Haga click en *New* para añadir una nueva variable (nombre y valor).
 - Haga click en existing variable, luego haga click en *Edit* para cambiar su valor.
 - Haga click en existing variable, luego haga click en *Delete* para borrarla.

Asignación de Variables de entorno en Linux Existen diferentes tipos de shells (interpretes de comandos) en Linux: Bourne Shell, Bourne Again Shell, C Shell, Korn Shell, etc. Para saber qué tipo de shell está usando, ejecute: `echo $SHELL`

Dependiendo del tipo de shell que utilice el comando para asignar valores a las variables de entorno puede cambiar. Por ejemplo:

- **bash** Añada la siguiente línea a su `.profile` o `.profile_bash`: `export VARIABLE=valor`.
- **tcsh** Añada la siguiente línea a su `.login`: `setenv VARIABLE valor`.

Para comprobar si las variables están asignadas correctamente abra una nueva consola shell y ejecute `env` para ver todas las variables de entorno y sus valores.

Asignación de Variables de entorno en Mac Los usuarios de Mac pueden también seguir las instrucciones para Linux sobre como asignar variables de entorno. Para MacOS se tiene que asignar la variable JAVA HOME a `/Library/Java/Home`

2.2.2 Comprobar su instalación de UIMA

Para comprobar que su instalación de UIMA funciona perfectamente, ejecute `$UIMA_HOME/bin/documentAnalyzer.bat` (or `.sh`)².

Una nueva ventana ‘Document Analyzer’ debería aparecer (ver Figura 1).

Asigne los valores en la ventana de la siguiente manera:

- Input Directory: `UIMA_HOME/examples/data`
- Output Directory: `UIMA_HOME/examples/data/processed`
- Location of Analysis Engine XML Descriptor:
`UIMA_HOME/examples/descriptors/analysis_engine/PersonTitleAnnotator.xml`

Sustituya en el ejemplo anterior `UIMA_HOME` por la carpeta donde instaló Apache UIMA (p.e. `/home/yourusername/apps/apache-uima`) Luego, presione el botón ‘Run’. En unos instantes debería aparecer una ventana ‘Analyzed Results’. Haga doble-click en alguno de los documentos que aparecen en la ventana para ver los resultados de ese documento.

Puede encontrar más información sobre cómo instalar UIMA en <http://uima.apache.org/doc-uima-examples.html>.

²En algunos Notebooks puede ser necesario seleccionar Ubuntu Gnome al iniciar la interfaz gráfica para que Document Analyzer funcione correctamente

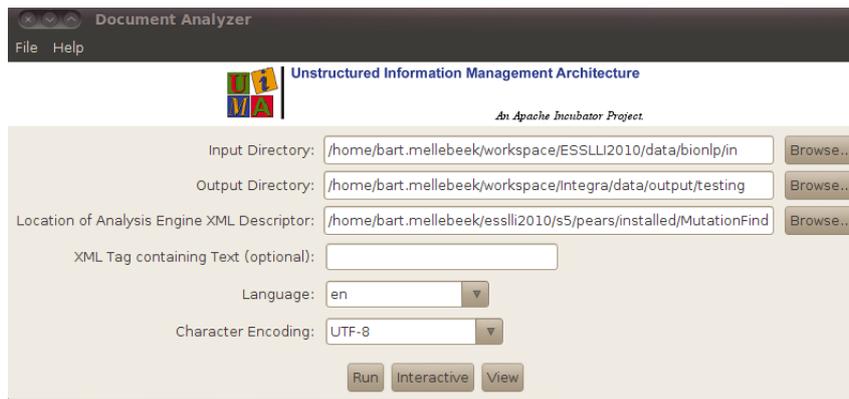


Figure 1: Comprobando la instalación de UIMA.

2.3 Instalando Eclipse

Recomendamos el uso de la versión 3.6 de eclipse Eclipse (Helios). Seleccione esa versión del Eclipse IDE para Java Developers en

<http://www.eclipse.org/downloads/packages/eclipse-ide-java-developers/heliosr>

2.3.1 Instalación de los plugins de eclipse para UIMA

Una vez Eclipse esta correctamente instalado, es necesario descargar e instalar los plugins de eclipse para UIMA. La mejor manera de hacerlo es mediante Eclipse Update site. Seleccione *Help Menu* y luego *Install New Software*. Aparecerá una nueva ventana como muestra la Figura 2. Pulse en el botón *add* de la parte superior derecha e introduzca <http://www.apache.org/dist/uima/eclipse-update-site/>. Eclipse mostrara dos clases de plugins. Seleccione *'Apache UIMA Eclipse tooling and runtime support'* y luego pulse en *install*.

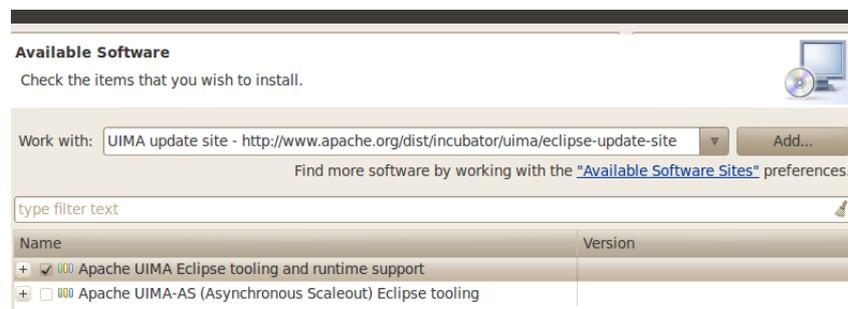


Figure 2: Instalando los plugins de eclipse para UIMA.

Para más información acerca de la instalación de los plugins de Eclipse para UIMA consulte la documentación oficial disponible en http://uima.apache.org/downloads/releaseDocs/2.2.2-incubating/docs/html/overview_and_setup/overview_and_setup.html#ugr.ovv.eclipse_setup

2.3.2 Comprobación de los eclipse plugins

Abrir eclipse, seleccionar *File*, *New* y luego *Others*. Aparecerá una ventana de diálogo, como muestra la Figura 3. Si puede seleccionar los diferentes UIMA artifacts, agrupados en el menú UIMA, la instalación del plugin es correcta.

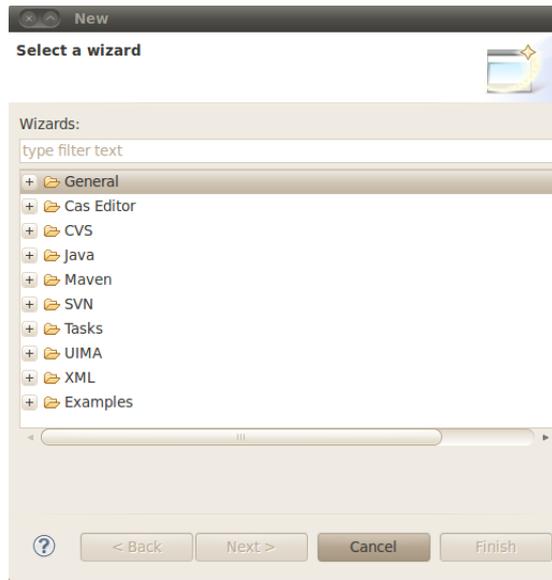


Figure 3: Comprobación de la instalación de los plugins de Eclipse para UIMA.

2.3.3 Configuración del proyecto ejemplos de UIMA

El proyecto `UIMA_HOME/apache-uima/examples` contiene ejemplos que permiten empezar a usar de manera sencilla las aplicaciones del UIMA SDK dentro de Eclipse.

Para importar y configurar correctamente este proyecto siga las siguientes instrucciones:

- En Eclipse, si la perspectiva-vista Java no esta abierta, abra la perspectiva-vista Java mediante *Window* → *Open Perspective* → *Java*.
- Asigne la variable `UIMA_HOME` al directorio donde instaló UIMA SDK, para ello:
 - Vaya a *Window* → *Preferences* → *Java* → *Build Path* → *Classpath Variables*.
 - Haga click en *New*.
 - Introduzca `UIMA_HOME` (all capitals, exactly as written) en el campo *Name*.
 - Introduzca su directorio de instalación (p.e. `/home/$USER/apps/apache-uima`) en el campo *Path*.
 - Haga click *OK* en el diálogo *New Variable Entry*.
 - Haga click *OK* en el diálogo *Preferences*.
 - Si pregunta si quiere realizar una compilación completa (*full build*), haga click en *Yes* (Sí).
- Seleccione *File* → *Import* en el menú.
- Seleccione *General/Existing Project into Workspace* y haga click en el botón *Next* (Siguiente).
- Click *Browse* y seleccione su directorio `UIMA_HOME/`.
- Click *Finish*. Esto creará un nuevo proyecto ‘uimaj-examples’ en su Eclipse workspace. Eclipse no debería mostrar ningún error.

2.3.4 Configuración Opcional para acceder al código fuente de UIMA

Para poder acceder al código fuente de UIMA dentro de Eclipse (p.e. para poder depurar) es necesario ejecutar vía un *shell script* (la variable UIMA_HOME tiene que apuntar a la carpeta donde instaló la distribución binaria de UIMA):

- Descargue y descomprima la distribución del código fuente de UIMA correspondiente a la versión instalada de UIMA.
- Ejecute el comando `addSourceToJars` que encontrará en el directorio raíz de la distribución del código fuente UIMA.

Una vez el código fuente de las librerías de UIMA estará disponible automáticamente dentro de Eclipse.