IBM

IBM Text Analyzer Business Component

Installation Guide

Version 1.1

… A member of the WebSphere Business Components family

Before using this information and the product it supports, be sure to read the general information under "Notices" on page 25.

Second Edition (March 2001)

This edition applies to version 1.1 of IBM Text Analyzer Business Component (part number 20P4393 or 20P4395), and to all subsequent releases and modifications until otherwise indicated in new editions. Make sure you are using the correct edition for the level of the product.

Corrections and suggestions for future revisions of this document are appreciated. Mail your comments to:

IBM Canada Ltd. Laboratory
Information Development
2G/KB7/1150/TOR
1150 Eglinton Avenue East
Toronto, Ontario, M3C 1H7
Canada

When you send information to IBM, you grant to IBM a nonexclusive right to use or distribute the information in any way they believe appropriate without incurring any obligation to you.

Contents

IBM Text Analyzer Business Component Version 1.1 Installation Guide

The IBM Text Analyzer (TA) Business Component is installed as two distinct and separately deployable subcomponents:

- Categorization Engine (CE)
  - o installed and deployed as a WSBC Version 1.1 Advanced Component (TA AC)
  - o temporarily deployable as a Java API
- Training Tool (TT)
  - o installed and deployed as a Java API

Two samples are provided with the product:

- Text Analyzer AC usage sample
- Text Analyzer Training Tool GUI sample

This installation guide contains a section with installation instructions for Text Analyzer Business Component product and a separate section for the Text Analyzer AC usage sample.

Instructions are provided for installing the TA subcomponents and samples for each of the development and production environments. Your development and production environments may be on the same machine.

In this installation guide, installation means installing the required prerequisite and product files; deployment means configuring the product and environment to make the product usable.

Text Analyzer Installation Instructions

The Text Analyzer setup that runs from the IBM Text Analyzer Business Component Version 1.1 CD-ROM will copy all required Text Analyzer files to your WSBC Studio installation (See the installation prerequisite sections). The TA AC and the TA Training Tool can be deployed separately in both the development and production environments after the files are installed.

Online help documentation installation options

The default (Typical) setup option is to install online help documentation when installing Text Analyzer from the CD-ROM. The WSBC Studio documentation system is a prerequisite for installing Text Analyzer documentation installation (See the *WSBC Studio Installation Guide* for details on installing and deploying the help system. A troubleshooting section is included.)

You may choose to install the product documentation alone. Both the WSBC Studio documentation system and the TA documentation may be installed separately from other WSBC installable components by selecting the Custom installation option and selecting to install only the documentation.

Development environment installation prerequisites

The following are the prerequisites for installing TA product files to a development workstation from the CD-ROM.

WSBC Studio

The Text Analyzer installation procedure has a dependency on WebSphere Business Components Studio. You must install WSBC Studio to your development workstation before you will be able to install any Text Analyzer component files to the workstation. You should follow the recommended order of installation for any software you expect to use; this will facilitate automatic setup of the environment (See the *WSBC Studio Installation Guide*).

It is recommended that you do a complete WSBC Studio installation and setup as described in the *WSBC Studio Installation Guide*. The minimum setup for a TA AC development installation is summarized as follows:

- VisualAge for Java (VAJ)
    - o IBM EJB Development Environment feature for VAJ
    - o IBM WebSphere Test Environment feature for VAJ (Use the default database for the Persistent Name Server database driver if DB2 is not installed.)
- WSBC Studio AC Deployment Tool
- WSBC Studio Advanced Component Services for the VAJ WebSphere Test Environment.

If you need to work only with the Categorization Engine Java API and the Training Tool Java API, then VisualAge for Java is the only requirement. However, WSBC Studio must still be installed to allow the TA installation program to run successfully. In this case you may delay setting up the WSBC Studio development environment to the time when you migrate to the TA AC. (Support for the Categorization Engine Java API is temporarily provided to facilitate migration from earlier product versions; this API may not be supported in future releases.)

Hardware

The Text Analyzer installation procedure does not have any hardware prerequisites for installation beyond those required by WSBC Studio. A full installation of Text Analyzer, including documentation, requires 73539Kb of available hard drive space.

Installing Text Analyzer files to a development workstation

Text Analyzer files are installed to a Windows NT development workstation by running the setup program from the IBM Text Analyzer Business Component CD-ROM. The setup program file is located in the TAInstall folder and is named Setup.exe.

Run the Setup.exe file to start the Text Analyzer installer and follow the prompts. You will be warned if you have not installed WebSphere Application Server and Java API for XML Parsing (JAXP); you may continue with the installation and install these items at a later time.

You must accept the International Program License Agreement to start the installation.

Select the installation components you wish to install. Choose Typical to install both the Text Analyzer files and the Text Analyzer documentation. Choose Custom to select individual installation components (uncheck the components that should not be installed.)

If you selected installation of the Text Analyzer component itself, then the Text Analyzer files will be copied to a TextAnalyzer folder within your WSBC Studio installation structure. If you selected to install the Text Analyzer documentation, the WSBC online help system will be updated; you will find the Text Analyzer documentation as an item under the Topics heading in the table of contents.

A Text Analyzer folder named <Your WSBC Install directory>\1.1\ACFeatures\Components\TextAnalyzer\ is created. All Text Analyzer files for both the Text Analyzer AC (Categorization Engine) and the Training Tool are located here. You will need to copy various files from this location when developing a TA client, developing a customized TT, and when installing and deploying the CE and TT to a production environment.

Installed files

The following tables describe the Text Analyzer files and their locations after installation. The default installation folder C:\Program Files\IBM\WSBC\1.1\ACFeatures\Components\ contains the TextAnalyzer folder; all Text Analyzer files are located within this folder. The XML4Jand JAXP files are required to complete an installation as described in the deployment prerequisite sections.

Text Analyzer folder	
Filename	**Description**
Events.xsd	TextAnalyzer AC event data definition schema
Functions.xsd	TextAnalyzer AC functions definition schema
TextAnalyzerAC.jar	TextAnalyzer AC implementation
TextAnalyzerACEJB.jar	Text Analyzer deployable EJB implementation
TextAnalyzerACEJBClient.jar	Interface stubs for remote Text Analyzer clients
TextAnalyzerACEJBDeployed.jar	WebSphere deployed Text Analyzer AC EJB
TextAnalyzerBase.jar	Training Tool and Categorization Engine files

TextAnalyzer\is	
Filename	**Description**
InterfaceModels.xsd	WSBC TextAnalyzer AC type definition schema

TextAnalyzer\LPResources\AIX\bin	
Filename	**Description**
libeflnl27x.so	AIX language processing executable
libPoeJNI.so	AIX JNI for language processing

TextAnalyzer\LPResources\Solaris\bin	
Filename	**Description**
libeflnl27s.so	Solaris language processing executable
libPoeJNI.so	Solaris JNI for language processing

TextAnalyzer\LPResources\WIN\bin	
Filename	**Description**
eflnl27w.dll	Windows language processing executable
poejni.dll	Windows language processing JNI

TextAnalyzer\LPResources\dictionaries	
Filename	**Description**
Files named *.dic and *.abr	Language specific dictionaries and abbreviations. See Text Analyzer online help references for the complete list of supported languages and their associated files

TextAnalyzer\rose\TextAnalyzerAC\Interface	
Filename	**Description**
TextAnalyzerAC.cat	Rational Rose model of the Text Analyzer AC and its interface.

TextAnalyzer\TTGUIV1R1	
Filename	**Description**
data.xml	Training sample template
projectini.xml	Project configuration template
projectList.xml	List of existing projects
TTGUI.bat	V1R1 GUI Launcher

TextAnalyzer\TTGUIV1R1\Samples\FullReuters-21578	
Filename	**Description**
data.xml	Training Sample
projectini.xml	Training configuration
rules.xml	Categorization rule set

TextAnalyzer\TTGUIV1R1\Samples\MiniReuters-21578	
Column	**Column**
data.xml	Training Sample
projectini.xml	Training configuration
rules.xml	Categorization rule set

Development environment deployment prerequisites

Text Analyzer requires the following additional resources to be available in a development environment.

WSBC Studio Tools

The WSBC Studio AC Deployment Tool must be installed and deployable in the development environment for managing the Text Analyzer AC. Refer to the *WSBC Studio Installation Guide* for information on installing and using this tool.

WSBC Advanced Component services installation and deployment

To deploy the Categorization Engine as a TA AC in a development environment, the VAJ WebSphere Test environment be installed and configured as described in the *WSBC Studio Installation Guide*. Advanced Component services must then be installed and deployed in the development environment as described in that installation guide.

JAXP installation and deployment

The Text Analyzer environment must provide the Sun Java API for XML Parsing Version 1.0.1 Final Release. This is available by download from java.sun.com/xml and should be installed by following its installation instructions. (At the time of this writing, this version is being replaced by Java API for XML Processing Version 1.1. The Early Access releases are NOT recommended.)

Add all JAXP packages to your VisualAge for Java workspace and create a new VAJ project if necessary to do this.

XML4J installation and deployment.

IBM XML Parser for Java (XML4J). Version 2.0.15 is deployed as part of the IBM WebSphere Test Environment (WTE) feature in VisualAge for Java V3.5; this is done during WSBC Studio setup. The XML4J VAJ project will be added to your Workspace.

See "Set up a development environment" in the *WSBC Studio Installation Guide* for details if your VAJ does not have this and other required features installed. (XML4J is also listed as separate feature, but the WTE feature should still be installed.)

Deploying the TT Java packages in a VAJ workspace

The Training Tool requires deployment of Java packages, DLLs and language dictionaries.

1. If required, create a WSBCTextAnalyzerAC VAJ project in your VAJ workspace.
2. Import all TA base .class and resource files from the TextAnalyzerBase.jar file into your selected project.
3. Deploy the language processing resources by copying the following files to the <VAJRoot> \ide\program\ folder.
    o eflnl27w.dll
    o POEJNI.dll
4. Take note of where the Language dictionaries (*.dic and *.abr files) are located; this location will need to be set in the training configuration when running the Training Tool.

Starting the Training Tool in a development environment.

You must provide a customized front end to deploy the Training Tool and install the following files to your Training Tool test environment where they will be accessible to your customized GUI implementation (A sample reference GUI is provided. See the TA online help documentation for instructions on how to deploy and configure the V1R1 Training Tool reference GUI.)

- TextAnalyzerBase.jar
- eflnl27w.dll
- POEJNI.dll
- Language dictionaries as required (*.dic and *.abr files)

Note: The Training Tool API can be tested outside of the VAJ workspace if your customized front-end also supports this. The sample Training Tool reference GUI provides this standalone capability.

Deploying the TA AC (CE) Java packages in a VAJ workspace

1. Deploy the TA base Java packages and the language processing resources as described in Deploy the TT Java packages in a VAJ workspace.
2. Add the TA Java packages in the TextAnalyzerAC.jar files to your WSBCTextAnalyzerAC project.
3. Create the WSBCTextAnalyzerAC EJB group.
   a. Select the **EJB** tab in the VAJ Workbench.
   b. Open the Enterprise Beans context menu with a right-click on the Enterprise Bean pane (or select **EJB** from the menu bar)
   c. Select **Add > EJB Group….** Create a new group named WSBCTextAnalyzerAC in your TextAnalyzer project.
4. Import the TA AC EJB into the WSBCTextAnalyzerAC EJB group.
   a. Select the ACTextAnalyzer EJB group from the Enterprise Beans pane.
   b. Open the Enterprise Beans context menu with a right-click (or select **EJB** from the menu bar). Select **Import Enterprise Bean...** to open the SmartGuide.
   c. Set the filename by browsing to the TextAnalyzerACEJB.jar. Ensure that the beans, .classes, and resource checkboxes are checked and that the dialog reports that all files are selected.
   d. Click **Finish** to import the TA AC EJB and close the SmartGuide

Deploying the TA AC EJB in the VAJ WebSphere Test Environment

1. Generate WebSphere/WTE server-deployable EJB code:
   a. Open the VAJ Workbench and select the **EJB** tab.
   b. Select the ACTextAnalyzer EJB group in the Enterprise Beans pane.
   c. Open the Enterprise Beans context menu with a right-click (or select **EJB** from the menu bar).
   d. Select **Generate Deployed Code** from the context menu to generate the deployable EJB code. Wait for the Information window to close.
2. Create an EJB Server configuration for AC Services and the TA AC.
   a. Select the ACServices, Localizable Text and TextAnalyzerAC EJB groups in the Enterprise Beans pane.
   b. Open the Enterprise Beans context menu with a right-click (or select **EJB** from the menu bar).
   c. Open the **EJB Server Configuration** browser by selecting **AddTo > Server Configuration** from the context menu. Your new server configuration containing your selected EJB groups will be automatically labeled with a number added to the bottom of the list. (Take note of your TA server configuration label.)

d. Modify the server configuration's Transaction Timeout property.
   i. Select your server configuration.
   ii. Open the context menu with a right-click (or select **Servers** from the menu bar) and select **Properties** to open the Properties for EJB Server window
   iii. Take note of the value of the **Transaction Timeout** property and the value of the **Transaction Inactivity Timeout** property. If you experience timeout errors (often reported as CORBA errors) you should increase these values. The tested values resolving timeout errors is 600,000, but lower values may be adequate for your environment.
   iv. Click **OK** to set the properties and close the Properties for EJB Server window.

Starting the TA AC EJB in the VAJ WebSphere Test Environment

1. Start the VAJ WTE from the VAJ Workbench.
   a. Open the WTE Control Center by selecting **Workspace >Tools > WebSphere Test Environment** from the menu bar. Wait for the window to open.
   b. Start the Persistent Name Server
      i. Expand the **Servers** item and select the **Persistent Name Server** item to open the Persistent Name Server view.
      ii. Start the Persistent Name Server by clicking the **Start name server** button. Wait for the VAJ Console window to open and then wait for the "Server open for business." message.
2. Register the Text Analyzer AC with the WTE JNDI.
   a. Run the WSBC AC Deployment Tool and select VisualAge for Java as the EJB Server
   b. Load the TextAnalyzerAC.jar file.
   c. Deploy the TextAnalyzer AC instance
      Note: Refer to the *WSBC Studio Installation Guide* and the WSBC Studio online help documentation for AC Deployment Tool details.
3. Start the EJB Servers for AC Services and the TA AC.
   a. Select the **EJB** tab from the Workbench.
   b. Open the Enterprise Beans context menu with a right-click (or select **EJB** from the menu bar).
   c. Open the EJB Server Configuration browser by selecting **OpenTo > Server Configuration** from the context menu.
   d. Locate the server configuration containing the AC Services and Text Analyzer Enterprise Beans.
   e. Start the server configuration.
      i. Select the server configuration and right-click on it to open the Servers context menu (or select **Servers** from the EJB Server Configuration menu bar).
      ii. Select **Start** from the menu. Wait for the VAJ Console to display the "Server open for business." message.
         Note: To stop a server, open the EJB Server Configuration window, select the server and choose **Stop** from the server context menu.

Production environment installation prerequisites

There are no functional prerequisites for installing TA to a production environment other than being able to copy the required files from the development environment to the production environment. Note that some recommended file locations for the TA AC are within the WebSphere Application Server installation folder.

Locating installed Text Analyzer files

You will need to copy files from your development environment to the production environment. Refer to the Installed files list in the Install Text Analyzer files to a development workstation section of this document to locate the files called for in the production environment deployment instructions.

An alternative to copying the TA files from a development environment is to install WSBC Studio and Text Analyzer to your production environment. This will provide you with copies of the files in the Installed files list. (Refer to the *WSBC Studio Installation Guide* and the "Install Text Analyzer files to a development workstation" section for instructions.)

You must use the DLL versions specified for your production operating system. DLLs are provided for Windows NT, Windows 2000, AIX, and Solaris operating systems.

Hardware

The Training Tool and Categorization Engine memory requirements depend on the size of the documents sample. A very large data set such as the Reuters-21578 standard can be processed with 256Mb of memory available, however your memory requirements are likely to be substantially less.

A Text Analyzer AC deployment has no hardware requirements beyond those required for WSBC Studio installation.

Production environment deployment prerequisites

The Training Tool and Text Analyzer AC (Categorization Engine) environments may be on the same or on different machines. The following describes the deployment requirements for both the Training Tool and the Categorization Engine.

WSBC Studio Tools

The WSBC Studio AC Deployment Tool must be installed and deployed in the production environment for managing the Text Analyzer AC. Refer to *the WSBC Studio Installation Guide* for information on installing and using this tool.

WSBC Advanced Component services installation and deployment

To deploy the Categorization Engine as a TA AC in a production environment, WebSphere Application Server must be installed and configured as described in the *WSBC Studio Installation Guide*. Advanced Component services must then be installed and deployed in the production environment as described in that installation guide.

JDK 1.2.2 installation and deployment option

WebSphere Application Server provides JDK 1.2.2. If this is installed in your environment then you may use this JDK for the following purposes.

- To temporarily deploy the Categorization Engine as a Java API rather than as an AC
- To deploy the Training Tool Java API

If you are deploying either of the above APIs and do not have JDK 1.2.2 (or later), then you should obtain JDK from the Sun web site and follow the installation instructions provided with the product.

JAXP installation and deployment

Text Analyzer Categorization Engine has a runtime dependency on the Sun Java API for XML Parsing (JAXP) Version 1.0.1 (Version 1.1 is not in its final release at the time of this writing and is untested.) See Development environment deployment prerequisites for details of how to obtain and install JAXP. (Future releases of the Training Tool may also require JAXP.)

JAXP must be made accessible to the application server instance where you run the TA AC.

XML4J installation and deployment

Text Analyzer Categorization Engine and Training Tool API have a runtime dependency on IBM XML Parser for Java (XML4J) Version 2.0.15 or later. If you have WebSphere Application Server 3.5 installed in your production environment, then you already have the required version of XML4J (the default location is <WebSphereRoot>\AppServer\lib\xml4j.jar. You may obtain the parser separately from the IBM alphaWorks Web site (The current version is 3.1.1).

XML4J must be made accessible from the application server instance where you run the TA AC.

Deploying the Training Tool in a production environment.

Install the following files to your Training Tool production environment where they will be accessible to your customized GUI (or some other front-end) implementation. (See the online help documentation for building a customized GUI and for instructions on deploying and configuring the V1R1 Training Tool reference GUI.)

- TextAnalyzerBase.jar
- eflnl27w.dll
- POEJNI.dll
- Language dictionaries as required (*.dic and *.abr files)
  Note: The language dictionary requirements are solely determined by the languages actually used for text analysis. See the online help documentation for more information on the use of language dictionaries. There is no harm in installing all dictionaries if space is available.

Configure your Training Tool front-end implementation to use the TT API.

Configure the TT API settings according to the location of the above listed files (if required).

Installing a deployable TA AC in a production environment

The Text Analyzer AC is installed as an application-server-deployable EJB and its supporting resources. Install the following files to your production environment where they will be accessible to your application server instance:

- TextAnalyzerBase.jar
- TextAnalyzerAC.jar
- TextAnalyzerACEJBDeployed.jar
- XML4J.jar
- efln27w.dll
- POEJNI.dll
- Language dictionaries as required (*.dic and *.abr files) by the Categorization Engine configurations

Notes:

1. XML4J is supplied with WebSphere Application Server as WebSphereRoot>\AppServer\lib\xml4j.jar and should be configured as described in the "Production environment deployment prerequisites" section.
2. It is recommended that the DLL files be placed in <WebSphereRoot>\AppServer\bin\.
3. The language dictionary requirements are solely determined by the languages actually used for text analysis. See the online help documentation for more information on the use of language dictionaries. There is no harm in installing all dictionaries if space is available.

Deploying a TA AC instance using WebSphere Application Server

Before deploying Text Analyzer you should deploy AC services. (See "Deploy Advanced Component Services on WebSphere Application Server" in the *WSBC Studio Installation Guide* for details.)

Configuring WebSphere Application Server to deploy Text Analyzer AC instances

1. Start the WebSphere Application Server AdminServer service.
2. Open the WebSphere Administrator's Console.
3. Select the topology view and then navigate to and select the server instance on which you want to deploy the Text Analyzer AC (WSBC Server is recommended). A pane will appear on the right side.
4. Enter the following command line arguments and values into the **Command line arguments** text field on the **General** tab of the configuration.
    - -mx128m
    - -classpath

    Add the following paths to the classpath value.

    - <your full path>/TextAnalyzerBase.jar
    - <your full path>/TextAnalyzerAC.jar
    - <your full path>/TextAnalyzerACEJBDeployed.jar

    Notes:

    1. The memory allocation option may be adjusted as required to suit your needs.
    2. Be sure to specify the full file system paths for the .jar files in the classpath argument and to separate the entries with a semicolon.

3. If the classpath argument already exists then you should simply append the additional classpath values to the existing classpath value. Remember to use a semi-colon to separate paths.

5. On the **Advanced** tab, Take note of the value of the **Transaction Timeout** property and the value of the **Transaction Inactivity Timeout** property. If you experience timeout errors (often reported as CORBA errors) you should increase these values. The tested values resolving timeout errors is 600,000, but lower values may be adequate for your environment.

6. Navigate up to the node (machine) where your server is located. In the **Dependent classpath** field enter the same classpath value as specified in Step 4.

Deploying an instance of the Text Analyzer AC using WebSphere Application Server

1. Start WebSphere Application Server service.
2. Start the WebSphere Application Server Administrator's Console.
3. Select the topology view.
4. Navigate down the server tree to your EJB container (WSBC Container in WSBC Server is recommended).
5. Open the **Create EJB dialog** with a right-click on the container to open the context menu, select **Create** from the menu, and finally select **Enterprise Bean** from the submenu.
6. Click on the **Browse** button and navigate to the location of the TextAnalyzerACEJBDeployed.jar file and select the file to view the EJBs it contains.
7. Select the com.ibm.wsbc.ac.textAnalyzerInterface/TextAnalyzerInterface.ser EJB. If you are prompted to enable work load management, select **No**. (The navigator closes and you are returned to the Create EJB dialog which now has fields filled in.)
8. Ensure that the EJB has a unique JNDI HomeName within the container. Click **Edit** on the Create EJB dialog to open the Deployment Properties dialog. Edit the **JNDI HomeName** field; the default name is the fully qualified package name followed by the Session Bean identifier followed by the instance name (com/ibm/wsbc/ac/textanalyzer/TextAnalyzerInterface). Any name may be used provided that it is unique. Click **OK** to close the Deployment Properties Dialog.
9. Register the Text Analyzer AC functions with the WebSphere Application Server JNDI.
   a. Run the WSBC AC Deployment Tool and select **WebSphere** as the EJB Server
   b. Load the TextAnalyzerAC.jar file.
   c. Deploy the TextAnalyzer AC instance.
      Note: Refer to the *WSBC Studio Installation Guide* and the WSBC Studio online help documentation for AC Deployment Tool details.

Installing and deploying a Categorization Engine configuration in a production environment

The instructions below describe how to install and deploy CE configurations once they have been produced by the Training Tool (The TA AC Sample provides a ready sample CE configuration). CE configuration file sets may be installed and deployed as they become available. A Text Analyzer AC's Categorization Engine uses a CE configuration produced by the Training Tool. You may install the TA AC Sample CE configuration if you do not have any of your own that are ready.

Installing a CE configuration

Install the CE configuration files in location local to your application server instance:

- Training configuration file produced by the Training Tool
- Rule set file produced by the Training Tool

One consideration is that your TA AC client code must supply the location of the CE configuration files when it initializes the TA AC.

Deploying a CE configuration

The CE configuration must be modified for the local environment. Specifically, the training configuration file must be modified to indicate the location of the dictionary files.

1. Locate the training configuration file (or files) that are installed.
2. In each training configuration file that you deploy, locate the following XML element:

```
<ini>
        <section name="LanguageConfiguration">
                    <setting active="yes" attribute="PathForDictionary">
                            <--Insert your dictionaries path here-->
                    </setting>
        </section>
</ini>
```

3. Change the element data to be the file system path to the installed language dictionaries. This path must be the full absolute path.

Text Analyzer AC Sample Installation Instructions

The following sections describe how to install the TA AC usage sample in a development environment and a production environment.

Installation and deployment prerequisites

The sample has no installation or deployment prerequisites beyond those required for the Text Analyzer product (described in the Text Analyzer installation instructions).

You will need to complete the installation and deployment procedures in the *WSBC Studio Installation Guide* and the Text Analyzer installation instructions for the environment in which you wish to use this sample. See the environment-specific deployment instructions for a summary of these requirements.

Installing Text Analyzer AC sample files in a development environment

The TA AC sample files are installed to a Windows NT workstation by running the Text Analyzer AC Sample installation program.

Run the WSBCTextAnalyzerSample110_013101.exe file and follow the prompts. The TA AC sample's files are initially installed to the following directory:

> <WSBC install directory>\1.1\ACFeatures\Samples\TextAnalyzerSample

The default location is:

> C:\Program Files\IBM\WSBC\1.1\ACFeatures\Samples\TextAnalyzerSample

Some of these files will need to be copied to other locations to complete your installation.

Installed files

TextAnalyzerSample folder	
Filename	**Description**
TextAnalyzerSample.jar	Java class files for the TA AC sample (See also the description for TextAnalyzerSampleSource.jar file)

TextAnalyzerSample\com\ibm\wsbc\samples\ta\resources	
Filename	**Description**
AnalyzeText.cfg	Configuration file for registering the analyzeText AC command in JNDI by using the RegisterCommand class.
InitializeAnalyzer.cfg	Configuration file for registering the initializeText AC command in JNDI by using the RegisterCommand class.
RemoveAnalyzer.cfg	Configuration file for registering the removeAnalyzer AC command in JNDI by using the RegisterCommand class.

TextAnalyzerSample\resources	
Filename	**Description**
projectini.xml	Training configuration file that is the first part of the CE configuration for the TA AC sample
rules.xml	Rule set file that is the second part of the CE configuration for the TA AC sample

TextAnalyzerSample\source	
Filename	**Description**
TextAnalyzerSampleSource.jar	Java code source files for the TA AC sample. (These are the source files for the contents of the TextAnalyzerSample.jar file.)

TextAnalyzerSample\web	
Filename	**Description**
feedback.html	Input page for customer question.
main.html	Entry page to the demo web.
uc.html	"under construction" HTML page.

TextAnalyzerSample\web\image	
Filename	**Description**
banner.gif	Top-of-page banner for web pages
submit.gif	Graphic for the Submit button on the feedback page.

TextAnalyzerSample\web\jsp	
Filename	**Description**
distributions.jsp	Distribution information page linked from the investments page.
general.jsp	Response page to an uncategorized question or categorization error.
gic.jsp	Guaranteed investment certificate page linked from the investments page.
investments.jsp	Response page displayed when the customer question is categorized as an investment question.
loans.jsp	Response page displayed when the customer question is categorized as a loan question.
loc.jsp	Line of credit page linked from the loans page.
mortgage.jsp	Mortgage page linked from the loans page.
multiple.jsp	Response page for a customer question that is categorized as both an investment and loan question.
mutualfund.jsp	Mutual fund page linked from the investments page.
personalloan.jsp	Personal loan page linked from the loans page.
uc.jsp	"under construction" JSP page

TextAnalyzerSample\web\theme	
Filename	**Description**
Master.css	Cascading style sheet for HTML presentation

Deploying Text Analyzer AC sample in a development environment

Installation and deployment prerequisites

If you have followed the instructions in the WSBC Studio and Text Analyzer installation guides then the prerequisites are in place. The following is a summary of the critical requirements.

To use the TA AC sample in a development environment you must meet the installation and deployment prerequisites for the following procedures and then complete these procedures:

- *WSBC Studio Installation Guide*, "Configure VisualAge for Java"
  - o Installing AC Services (for VAJ)
  - o Configuring AC Services (for VAJ)
- Text Analyzer installation instructions
  - o Install Text Analyzer files to a development workstation
  - o Deploy the TA AC (CE) Java packages in a VAJ workspace

See the "Development environment installation prerequisites" in the Text Analyzer installation instructions for the minimum requirements of a TA AC development environment.

Installing the TA AC sample files into the VAJ environment

The following instructions direct you to copy files from the following installation directory:

> <WSBC install directory>\1.1\ACFeatures\Samples\TextAnalyzerSample

Refer to the Installed files list to locate the files.

Installing the Categorization Engine configuration files

1. Locate the following directory:
   <VAJ install directory>\ide\program\
2. In the program directory, create a directory named TextAnalyzerSample\resources
3. Copy the following files into the new resources directory:
   - projectini.xml
   - rules.xml
4. Set the location for the language processing error log in the projectini.xml file. Replace the data in the following XML element:

   > <setting active="yes" attribute="ErrorLogPath" > d:\errorLog</setting>

   > with:

   > <VAJ install directory>\ide\program\TextAnalyzerSample\resources

5. Set the location of the language processing dictionaries in the projectini.xml file. Replace the data in the following XML element:
   <setting active="yes" attribute="PathForDictionary" >j:\vaj35\ide\program</setting>
   with your dictionary installation path.
   (See the Text Analyzer installation instructions for details; the default is C:\Program Files\IBM\WSBC\1.1\ACFeatures\Components\TextAnalyzer\LPResources\dictionaries)

Installing the HTML presentation files

The sample uses an HTML interface. The following describes how to complete the installation of the HTML presentation files.

1. Install the HTML files.
    a. Locate the following directory:
        <VAJ install directory>\ide\project_resources\IBM WebSphere Test Environment\hosts\default_host\default_app\web\
    b. Create a subdirectory webapp\TextAnalyzerSample under the ..\default_host\default_app\web\ directory.
    c. Copy the following installed files into the directory you just created:
        - All installed HTML files
            o   feedback.html
            o   main.html
            o   uc.html
2. Install the cascading style sheet (theme) file.
    a. Create a directory named theme in ..\webapp\TextAnalyzerSample.
    b. Copy the installed cascading style sheet file Master.css into this directory.
3. Install the graphics files
    a. Create a directory named image in ..\webapp\TextAnalyzerSample.
    b. Copy all installed image files to this image directory
        - banner.gif
        - submit.gif
4. Install the JSP files
    a. Locate the following directory:
        <VAJ install directory>\ide\project_resources\IBM WebSphere Test Environment\hosts\default_host\default_app\web\jsp\
    b. Copy all installed JSP files to the above-mentioned jsp directory
        - distributions.jsp
        - general.jsp
        - gic.jsp
        - investments.jsp
        - loans.jsp
        - loc.jsp
        - mortgage.jsp
        - multiple.jsp
        - mutualfund.jsp
        - personalloan.jsp
        - uc.jsp

Deploying the Java code files in a VAJ workspace

The following instruction describe how to import the Java source code for the TA AC Sample from the installed TextAnalyzerSample.jar. file

1. Open the VAJ Workbench and select the **Projects** tab.
2. Add a WSBCTextAnalyzerACSample project to the workspace.
    a. Open the Smart Guide Add Project view by right-clicking in the **All Projects** pane and selecting **Add > Project** from the context menu (or select **Selected > Add > Project** from the Workbench menu bar).
    b. Select the **Create a new project named** option and enter WSBCTextAnalyzerACSample in the associated text box.
    c. Click **Finish** to create the project and close the SmartGuide

3. Add the sample's source code files to the WSBCTextAnalyzerSample project.
   a. Open the SmartGuide Import view by selecting **File > Import** from the Workbench menu bar.
   b. Choose to import from a jar file import source by selecting the **Jar file** option on the Import view.
   c. Advance to the Import from jar/zip file view by clicking **Next**.
      i. Select the installed TextAnalyzerSampleSource.jar file as the source file. Enter (or browse to) its full path and filename in the **Filename:** textbox. (The default installation location is: C:\Program Files\IBM\WSBC\1.1\ACFeatures\Samples\TextAnalyzerSample\source\TextAnalyzerSampleSource.jar)
      ii. Choose to import files of the Java source code type. Check the box next to .java files.
      iii. Choose to import all Java source code files found in the source file.
         1. Open the Source file import window by clicking the **Details** button next to the .java check box.
         2. Click **Select all** (all listed files should now be checked). Click **OK** to close the window. (The Import from jar/zip file view should now report that you have selected 0 .class files, 8 .java files and 0 resource files for import.)
      iv. Select the WSBCTextAnalyzerSample project as the project to import into. Enter its name in the Project text box.
   d. d. Click **Finish** to import the source files and close the SmartGuide.

   Note: The Java class files that correspond to the Java code files imported from the TextAnalyzerSampleSource.jar are contained in the TextAnalyzerSample.jar file. There is no need to regenerate these classes to run the sample in WTE or WebSphere Application Server.

Deploying the JNDI registration resources

You must complete the deployment of the Java code files before the JNDI registration resources can be deployed.

1. Locate the following directory tree in the <WSBC install directory>\1.1\ACFeatures\Samples\TextAnalyzerSample directory:
   \com\ibm\wsbc\samples\ta\resources*.cfg
   The default location is:
   C:\Program Files\IBM\WSBC\1.1\ACFeatures\Samples\TextAnalyzerSample
2. Copy this tree (and the three configuration files) to the following location:
   <VAJ install directory>\ide\project_resources\WSBCTextAnalyzerACSample
3. Verify that the following path exists
   <VAJ install directory>\ide\project_resources\WSBCTextAnalyzerACSample\com\ibm\wsbc\samples\ta\resources\
   Note: The directory name WSBCTextAnalyzerACSample matches the name of the name of the project in which the package com.ibm.wsbc.samples.ta is located.
4. Verify that the following files exist in the above mentioned path
   - o AnalyzeText.cfg
   - o InitializeAnalyzer.cfg
   - o RemoveAnalyzer.cfg
5. Place the JNDI registration resources in the classpath of the TA AC sample RegisterCommand class.
   a. Open the VAJ Workbench to the **Projects** tab.
   b. Expand the WSBCTextAnalyzerACSample project and then expand the com.ibm.wsbc.samples.ta package.
   c. Edit the RegisterCommand class properties
      i. Open the Properties for RegisterCommand window by right-clicking the RegisterCommand class item and selecting **Properties** from the pop-up menu.
      ii. Select the **Class Path** tab in the Properties for RegisterCommand window.
      iii. Open the Class Path window by clicking the **Edit** button next to the Project path text box.

iv.   Select the WSBCTextAnalyzerACSample project path for inclusion in the classpath. (The resources were placed in the project path in an earlier step.)

v.   Click **OK** to set the path and close the Class Path window (The project path should now be included in the Complete classpath: text area on the Properties for RegisterCommand window.)

d.   Click **OK** to save the properties and close the Properties for RegisterCommand window.

Deploying the servlet-run classes in the VAJ WTE

The sample uses the WebSphere Test Environment servlet engine. Add the TA AC sample classes to the servlet engine's classpath.

1. Open the VAJ Workbench
2. Open the WebSphere Test Environment Control Center by selecting **Workspace > Tools > WebSphere Test Environment…**
3. Open the Servlet Engine view by expanding the **Servers** node and selecting the **Servlet Engine** item.
4. Open the Servlet Engine Class Path window by clicking on the **Edit Class Path…** button.

   a.   Select the WSBCTextAnalyzerACSample and WSBCTextAnalyzerAC to be included in the classpath by checking the box next to the project name. You may also click Select All to select all projects, but in this case, you should still confirm that the project exists in the list.
   Note: See "Import Java code files into a VAJ workspace" if this project is not in the scrollable list.

   b.   Click **OK** to set the classpath and close the Servlet Engine Class Path window.

Deploying the HTML presentation files for the VAJ WTE

The HTML presentation files (as installed) are configured for WebSphere Application Server. The VisualAge for Java WebSphere Test Environment has slightly different file system path requirements. Complete deployment of these files by modifying them as described below.

1. For the Feedback.html, investments.jsp, loans.jsp and multiple.jsp files, use a text editor to find all occurrences of the following string:

   /webapp/TextAnalyzer

2. If the occurrence of this string is a call to servlet code (which will look like /webapp/TextAnalyzerSample/servlet/com.ibm.wsbc.samples...) remove the string /webapp/TextAnalyzerSample.

3. For all other occurrences, do not make any changes.

You can find the HTML files in the directory directory <VAJ install directory>\ide\project_resources\IBM WebSphere Test Environment\hosts\default_host\default_app\web\webapp and the JSP files in ..\web\jsp.

Starting the TA AC Sample in the VAJ WebSphere Test Environment

The following instructions for the sample differ slightly from the Text Analyzer installation instructions. The sample requires that the WTE servlet engine is started and that the JNDI registration be done manually. You must complete all deployment steps before attempting the following instructions.

1. Start the VAJ WTE from the VAJ Workbench.

   a.   Open the WTE Control Center by selecting **Workspace >Tools > WebSphere Test Environment** from the menu bar. Wait for the window to open.

   b.   Start the Servlet Engine
       i.   Expand the **Servers** item and select the **Servlet Engine** item to open the Servlet Engine view.
       ii.   Start the servlet engine by clicking the **Start Servlet Engine** button. Wait for the VAJ Console window to open and then wait for the "Server open for business." message.

   c.   Start the Persistent Name Server
       i.   Expand the **Servers** item and select the **Persistent Name Server** item to open the Persistent Name Server view.

<blockquote>

ii.   Start the Persistent Name Server by clicking the **Start name server** button. Wait for the VAJ Console window to open and then wait for the "Server open for business." message.
Note: The status of the servers is also reported at the bottom of the WTE Control Center.
</blockquote>

2. Start the EJB Server for AC Services and the TA AC.
    a.   Select the **EJB** tab from the VAJ Workbench.
    b.   Open the Enterprise Beans context menu with a right-click (or select **EJB** from the menu bar).
    c.   Open the EJB Server Configuration browser by selecting **OpenTo > Server Configuration** from the context menu.
    d.   Locate the server configuration containing the AC Services and Text Analyzer Enterprise Beans. (Instructions for creating the server configuration and its associated EJB groups are contained in the *WSBC Studio Installation Guide* and the Text Analyzer installation instructions).
         Start the TA server configuration.
         i.   Select the server configuration and right-click on it to open the Servers context menu (or select **Servers** from the EJB Server Configuration menu bar).
         ii.   Select **Start** from the menu. Wait for the VAJ Console to display the "Server open for business." message.

         Note: To stop a server, open the EJB Server Configuration window, select the server and choose **Stop** from the server context menu.

3. Register the Text Analyzer AC with the WTE JNDI.
    a.   Open the VAJ Workbench to the Projects tab.
    b.   Expand the WSBCTextAnalyzerACSample project and then expand the com.ibm.wsbc.samples.ta package.
    c.   Run the main method of the RegisterCommand class.
         Note: The AC Deployment tool is normally used to handle this task. This sample accomplishes the same task with a manual process that runs the RegisterCommand class (provided with the sample). All AC commands are bound to a TA AC instance named TextAnalyzerInterface using the settings found in the JNDI registration resource files.

Using the TA AC Sample Application

Open the following in a Web browser:

http://localhost:8080/webapp/TextAnalyzerSample/main.html

On the feedback page submit a comment related to loans or investments. Click the **Submit** button to have Text Analyzer return a page appropriate to your comment.

Installing the Text Analyzer Sample in a production environment

Installing and deploying the Categorization Engine configuration files

The sample provides a CE configuration (produced by the Training Tool) that must be installed as described below. This sample's training configuration file is named projectini.xml and the rule set file is named rules.xml; these files are contained in the resources directory mentioned below.

1. Create the following directory:

    > <WebSphere Application Server install directory>\AppServer\hosts\default_host\TextAnalyzerSa mple\

2. Locate the following installed files directory; copy the directory (and its contents) into the TextAnalyzerSample directory create in the above step:

    > <WSBC install directory>\1.1\ACFeatures\Samples\TextAnalyzerSample\resources\

3. Deploy the sample's CE configuration by updating the training configuration file (projectini.xml) with the path to your dictionary location. Refer *to Installing and deploying a Categorization Engine configuration in a production environment* for more details.

Installing the Java class archive

If necessary, copy the following installed file to your production environment where it will be accessible to your application server instance.

> <WSBC install directory>\1.1\ACFeatures\Samples\TextAnalyzerSample\TextAnalyzerSample.jar

Take note of where you place this file because it will be need in the *Deploying the TA AC sample using WebSphere Application Server* section.

Installing the sample's TA AC function JNDI registration resources

If necessary, copy the following installed directory (and its contents) to your production environment where it will be accessible to your application server instance.

> <WSBC install directory>\1.1\ACFeatures\Samples\TextAnalyzerSample\com\
>
> **Note**: You need to copy the tree only from the *com* directory down.

Take note of the directory where you place this tree because it will be needed in the *Deploying the TA AC sample using WebSphere Application Server s* section.

Installing the HTML presentation files

The sample uses an HTML interface. The files for this interface must be installed as described below.

1. Locate the following directory (created in the preceding section).

   <WebSphere Application Server install directory>\AppServer\hosts\default_host\TextAnalyzerSample

2. Locate the following installed files directory; copy it (and its contents) into the TextAnalyzerSample directory mentioned above:

   <WSBC install directory>\1.1\ACFeatures\Samples\TextAnalyzerSample\web\

Deploying the TA AC sample using WebSphere Application Server

The following instructs describe how to complete the deployment of the sample.

Expected environment

The following instructions assume the following are installed and deployed:

- AC Services as described and recommended in the WSBC Studio Installation Guide.
- Text Analyzer AC as described and recommended in Deploying a TA AC instance using WebSphere Application Server.

Configuration

The following section describes how to configure WebSphere application server to run the sample.

1. Start the WebSphere Admin Server
2. Start the WebSphere Administrator's Console.
3. Set the node's dependent classpath
   a. In the Administrator's Console tree view, click the node with your server name.
   b. On the right, append the following path value to the **Dependent classpath** field:
   c. <your full path>\TextAnalyzerSample.jar
   d. Click **Apply**.

4. Set the server command line arguments
    a. Expand the node with your server name. Click the **WSBC Server** node.
    b. In the **General** tab, in the **Command line arguments** field, append the following path to the -classpath value (remember to add a semi-colon first).
    c. <your full path>\TextAnalyzerSample.jar
5. Create the web application
    a. Open the **Create Web Application** window by selecting **Console > Tasks > Create a Web Application** from the **WebSphere Advanced Administrative Console** menu bar.
    b. Set the web application name to *TextAnalyzerSample* in the **Web Application Name** field.
    c. Enable serving servlets by classname by checking the **Serve Servlets by Classname:** checkbox.
    d. Click **Next** to open the next view. Set the parent servlet engine by expanding the **Nodes** tree. Expand the WSBC Server node. Select the **Default servlet engine** node.
    e. Click **Next** to open the next view. Accept the default values; click **Finish** to create the web application and close the window. (The web application should now appear under the WSBC Server in the Administrative Console.)
6. Set the server's working directory. In the Administrative Console, find and select the WSBC Server node. In the right pane, enter the following in the **Working directory** field:
    <WebSphere installation directory>\AppServer\host\default_host

Starting the Application Server

Right-click the server node (WSBC Server if the recommendations were followed) and select **Start** from the pop-up menu. (This should start AC Services and the Text Analyzer AC.)

Registering the sample's TA AC functions

From a command prompt, run the Java class RegisterCommand from the TextAnalyzerSample.jar file. This will register the three AC function commands that the sample will use.

To run the java class, use the following command:

```
Java –classpath <classpath value> com.ibm.wsbc.samples.ta.RegisterCommmand
```

Where <classpath value> contains the following
- <your full path>\TextAnalyzerSample.jar
- <your full path to the JNDI registration resources>
- <WebSphere installation directory>\lib\ujc.jar

Notes:
1. <your full path to the JNDI registration resources> is determined by where you placed the registration resources directory tree. It is the path down to and including the parent of the *com* directory, but not the *com* directory itself. *See Installing the sample's TA AC function JNDI registration resources* for more details. The default value is:
    C:\Program Files\IBM\WSBC\1.1\ACFeatures\Samples\TextAnalyzerSample\TextAnalyzerSample.jar;C:\Program Files\IBM\WSBC\1.1\ACFeatures\Samples\TextAnalyzerSample;C:\WebSphere\AppServer\lib
2. Remember to separate the individual path values with semi-colons.

Using the TA AC sample

Open the following in a Web browser:

> http://<machine name>/webapp/TextAnalyzerSample/main.html

The variable <machine name> is the name of the sever on which you have deployed this sample.

On the feedback page submit a comment related to loans or investments. Click the **Submit** button to have Text Analyzer return a page appropriate to your comment.

Notices

IBM may not offer the products, services, or features discussed in this document in all countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law:

INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

IBM Canada Ltd.,
Department 071,
1150 Eglinton Avenue East
Toronto, Ontario, M3C 1H7
Canada

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement, or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurement may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrates programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. You may copy, modify, and distribute these sample programs in any form without payment to IBM for the purposes of developing, using, marketing, or distributing application programs conforming to IBM's application programming interfaces.

Trademarks and service marks

The following terms are trademarks of International Business Machines Corporation in the United States, or other countries, or both:

AIX

CICS

DB2

DB2 Universal Database

e-business

IBM

LANDP

MQSeries

OS/2 Warp

OS/390

RS/6000

SanFrancisco

VisualAge

Visual Banker

WebSphere

Lotus, Domino, Lotus Notes, and Notes Mail are trademarks of the Lotus Development Corporation in the United States, other countries, or both.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

MMX, Pentium, and ProShare are trademarks or registered trademarks of Intel Corporation in the United States, other countries, or both.

UNIX is a registered trademark in the United States, other countries, or both and is licensed exclusively through X/Open Company Limited.

Rational Rose is a registered trademark of Rational Software Corporation.

Other company, product, and service names may be trademarks or service marks of others.

End of document