



IBM Text Analyzer Business Component

Release Notes

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 2.

First Edition (November 2000)

This edition applies to version 1.1 of IBM Text Analyzer Business Component (part numbers 20P4393, and 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.

© Copyright International Business Machines Corporation 2000. All rights reserved.

Note to U.S. Government Users Restricted Rights — Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

Contents

- Text Analyzer Version 1.1 1**
 - Notes 1
 - Installation and deployment dependencies on WSBC products 1
 - Deployment dependencies on other products 1
 - Installation Guide deployment instructions supercede online documentation 1
 - Memory allocation for rule generation 1
 - FullReuters-21578 training sample contents 1
 - For current information 1
 - Text Analyzer sample 1

- Notices 2**
 - Trademarks and service marks 3

Text Analyzer Version 1.1

IBM Text Analyzer Business Component is a WSBC Advance Component Version 1.1. It consists of two subcomponents: the Categorization Engine which is deployed as the Text Analyzer AC and the Training Tool which is used to produce the Categorization Engine configuration.

The green text represents changes that were made to this document after the Text Analyzer Business Component Version 1.1 CD was created.

Notes

Installation and deployment dependencies on WSBC products

Text Analyzer has a development installation dependency on WSBC Studio and a deployment dependency on AC Services.

Deployment dependencies on other products

Text Analyzer has a deployment dependency on IBM XML4J Parser for Java and the Sun Java API for XML Parsing (JAXP).

Installation Guide deployment instructions supercede online documentation

The deployment instructions for the Text Analyzer AC given in the Installation Guide should be used where they differ from the instructions that appear in the online documentation. The changes are related to the installation and deployment of the Text Analyzer AC EJB.

Memory allocation for rule generation

When using the Training Tool API it may be necessary to allocate additional memory for the JVM when using a very large training sample. If you encounter an `OutOfMemoryError` try increasing the maximum memory size for the JVM heap with a command line option; for example:

```
java -Xmx256000000 YourRunnableClass
```

Consult your JDK documentation for more information on how to use the `Xms` and `Xmx` options.

FullReuters-21578 training sample contents

The FullReuters-21578 data.xml file contains the training subset of the Reuters-21578 data set. This sample contains 9603 training documents whereas the complete Reuters-21578 dataset also contains testing documents for a total of 21,578 documents.

For current information

The most current information for Text Analyzer is available at the WebSphere Business Components website at ibm.com/software/components/support.html.

Text Analyzer sample

The Text Analyzer sample is not currently available. See the WebSphere Business Components website for information on when the sample will be obtainable.

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.

Licenses 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