

Technical Memo

Information Builders Two Penn Plaza New York, NY 10121-2898 (212) 736-4433

Copyright © 2012

WebFOCUS Sentiment Analysis September 2012 TM4726

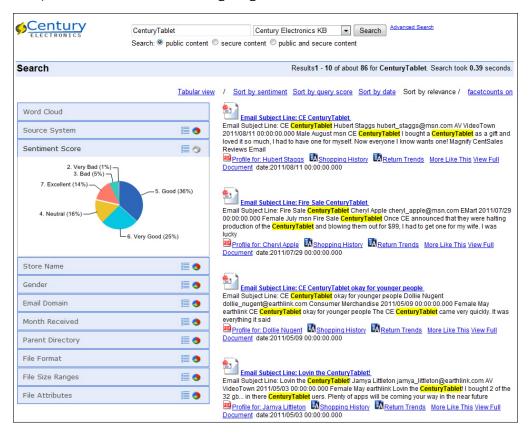
WebFOCUS Sentiment Analysis

WebFOCUS Sentiment Analysis scores structured and unstructured content by identifying positive, neutral, and negative meaning in information found in emails, documents, and database records.

Sentiment analysis translates both structured and unstructured data into emotional intelligence, providing companies with insight to facts with a point of view. Conventional methods, such as survey analysis or text mining, to gauge customer feelings and attitudes can be expensive, time-consuming, and difficult to measure in real-time. Transitioning to sentiment analysis facilitates churn reduction and increased sales opportunities by pinpointing customer perceptions and responding with appropriate strategies in product improvement, marketing, and public relations.

DN4501215.0912

Sentiment score can be applied to a record so that it can be stored in any database and leveraged by Business Intelligence for reporting, integration with an application, or part of a corporate dashboard. The following image shows the Sentiment score results for a search.



Expressions come in various modes, which makes recognizing sentiment difficult, especially when expressed in subtle and intricate ways. With Sentiment Analysis, users can take action on intelligence gained from knowing the emotions, opinions, and attitude of their audience in order to:

- Monitor tone of communications.
- Improve customer experience.
- Observe product and service perception.
- Identify what is important and valued.

All information has meaning, but identifying its tone provides a new dimension to truly understanding every aspect of enterprise data. When combined with Business Intelligence, the financial impact of emotion becomes measurable.

Overview

This document provides an overview of the WebFOCUS Sentiment Analysis options, including an in-house version called WAND DataFacet Taxonomy Server for Sentiment Analysis and a service version with Alchemy.

WebFOCUS Sentiment Analysis JOINs content in original data sources to the Sentiment Analysis synonym so that when a WebFOCUS procedure is executed, the value found in the field or fields designated for sentiment analysis are passed for evaluation by either the inhouse or service versions of WebFOCUS Sentiment Analysis, which returns the sentiment score. This information can be stored with the data and used for reporting, charts, or filtering search results.

Taxonomy Server for Web Services for Sentiment Analysis

The Adapter for Web Services creates a connection to the Taxonomy Server. This will be used as part of JOINs in WebFOCUS procedures to applicable fields for Sentiment Analysis evaluation. The field or fields assigned to a JOIN have their values passed to the Taxonomy Server, whose Master File has a field designated for the sentiment scoring result.

Adapter Prerequisites

The	Adapter	for	Web	Services	has t	the	following	prerequisites:

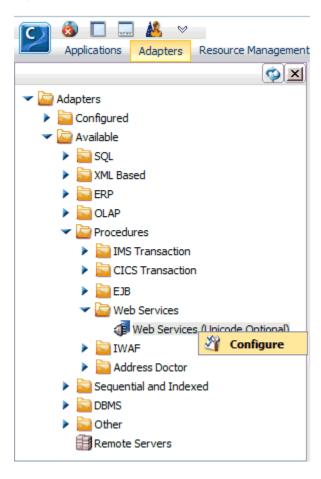
- ☐ The Taxonomy Server WSDL address.
- Version 7.7.04 Reporting Server and higher.
- USAGE length of the data being evaluated (optional for actual setup, needed for execution).

For example, the COURSES synonym has three description fields that will be concatenated in a DEFINE field. The USAGE for each is A20V. The DEFINE field is set to A60. Therefore, the USAGE length needed is A60. If only one description field were to be used, the USAGE length needed is A20. This can always be changed later if the size is not known.

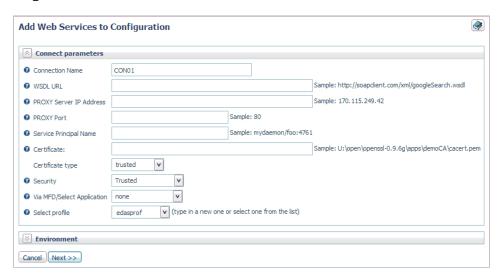
Procedure: How to Configure the Adapter for Web Services to Use Sentiment Analysis with the Taxonomy Server

- **1.** Open the Reporting Server Web Console.
- 2. From the Web Console menu bar, click Adapters.
- **3.** Expand the *Available* folder, if it is not already expanded.
- **4.** Expand the *Procedures* folder.
- **5.** Expand the Web Services folder.

6. Right-click Web Services (Unicode Optional) and select Configure.



The Add Web Services to Configuration dialog box opens, as shown in the following image.



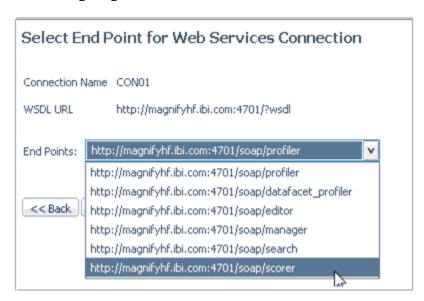
- 7. Enter a name in the Connection Name field.
- 8. Paste the WSDL URL into the WSDL URL field.

Note: For more information, see *How to Validate the WAND DataFacet Taxonomy Server Installation* on page 30.

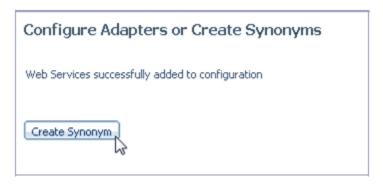
9. Click Next.

The Select End Point dialog box opens.

10. Select the http://<server>:<port>/soap/scorer End Point, and click Next, as shown in the following image.



11. Click Create Synonym.



The Select Operation for Web Services dialog box opens, as shown in the following image.

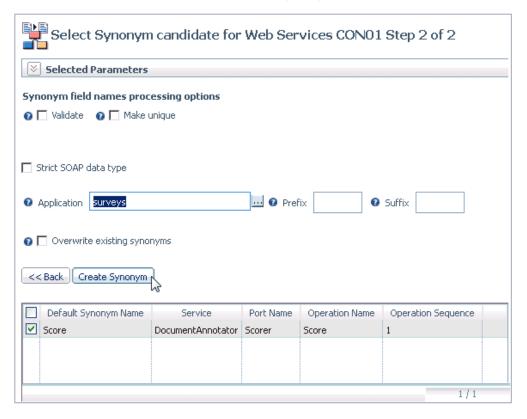


12 Click Select Operations.

The Select Synonym Candidates dialog box opens.

- **13.** Enter the target folder name in the Application field, or click the selector (...) button to navigate to one.
- 14. Optionally, add a prefix or suffix.
- **15.** Select the Score synonym.

16. Click *Create Synonym*, as shown in the following image.

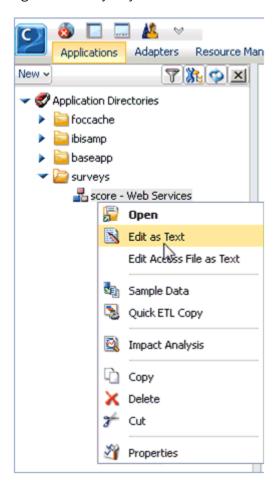


The synonym is created in the target application folder and a confirmation message is displayed.

Procedure: How to Update the Master File Description

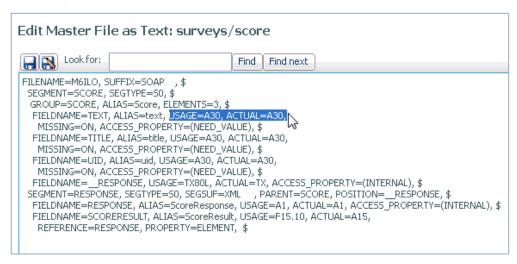
The Sentiment Analysis synonym is preconfigured to accept 30-character strings and scores to a precision of two decimal places. The string length must match the input string length found in the original Master File, otherwise content will be truncated as it is passed to the Taxonomy Server. It is highly recommended that you set the score result precision to 10 decimal places. This will provide various levels of sentiment granularity.

- **1.** Navigate to the Application directory with the Sentiment Analysis synonym.
- 2. Right-click the synonym and select *Edit* as *Text*, as shown in the following image.



DN4501215.0912

3. Change the USAGE and ACTUAL values for the TEXT FIELDNAME to be the same as the field being evaluated (see *Adapter Prerequisites* on page 3).



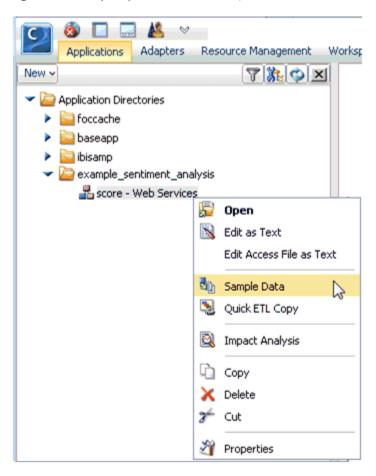
- 4. Click the Save icon.
- 5. Change the USAGE value for the SCORERESULT field to F15.10 (from F15.2).

```
Edit Master File as Text: surveys/score
Look for:
                                         Find
                                               Find next
FILENAME=M6ILO, SUFFIX=SOAP
 SEGMENT=SCORE, SEGTYPE=S0, $
  GROUP=SCORE, ALIAS=Score, ELEMENTS=3, $
  FIELDNAME=TEXT, ALIAS=text, USAGE=A30, ACTUAL=A30,
   MISSING=ON, ACCESS_PROPERTY=(NEED_VALUE), $
  FIELDNAME=TITLE, ALIAS=title, USAGE=A30, ACTUAL=A30,
   MISSING=ON, ACCESS PROPERTY=(NEED VALUE), $
  FIELDNAME=UID, ALIAS=uid, USAGE=A30, ACTUAL=A30,
   MISSING=ON, ACCESS_PROPERTY=(NEED_VALUE), $
  FIELDNAME=__RESPONSE, USAGE=TX80L, ACTUAL=TX, ACCESS_PROPERTY=(INTERNAL), $
 SEGMENT=RESPONSE, SEGTYPE=S0, SEGSUF=XML , PARENT=SCORE, POSITION=__RESPONSE, $
  FIELDNAME=RESPONSE, ALIAS=ScoreResponse, USAGE=A1, ACTUAL=A1, ACCESS_PROPERTY=(INTERNAL), $
  FIELDNAME=SCORERESULT, ALIAS=ScoreResult, USAGE=F15.10, ACTUAL=A15,
   REFERENCE=RESPONSE, PROPERTY=ELEMENT, $
```

6. Click the Save icon.

Procedure: How to Validate the Sentiment Analysis Synonym and the Adapter for Web Services

- **1.** Navigate to the Application directory with the Sentiment Analysis synonym.
- 2. Right-click the synonym and select Sample Data, as shown in the following image.



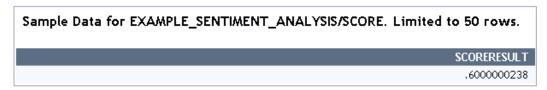
- 3. Select the SCORE.TEXT check box.
- **4.** Enter text in the *Input Value* field.

5. Click Sample Data, as shown in the following image.



The TEXT field sets the content passed from WebFOCUS to the Taxonomy Server.

The SCORERESULT field stores the resulting sentiment score.



Reference: Example of a Single Field Procedure

PAGE 1		
EMAILID	EMAILSUBJECT	SCORERESULT
1	Why did CE's developer relations exec Scott Addams go to a competitor?	.400000006
2	A good Option for Tablets	.399911582
3	A good Option for Tablets (Updated 1)	.400000006
4	A good Option for Tablets (Updated 2)	.200000003
5	Great Device, with option to make it faster!	.599955797
6	From a dedicated apple fan This IS a good device	1.000000000
7	great great	.400000006
8	worst tablet	.000000000
9	Limited usage	.200000003
10	CenturyTabletgood	.200000003
11	it made me hurt my husband's feelings	.400000006
12	Great for college students on a budget!	.392857134
13	CE CEOS	.199955791
14	iPad clone (sort of)	.000000000
15	CE CenturyTablet 9.7in Tablet	.400000006
16	Bad seller,bad product	.000000000
17	DOA	.200000003
18	thank you	.400000006
19	Very Good Product!!!	.400000006
20	Just love it	.200000003
21	Great tablet for this price	.600000023
22	Super Product Although Discontinued Their Loss our Gain	.599955797

JOIN EMAILBODY IN CENTURYREVIEWS TO TEXT IN SCORE

TABLE FILE CENTURYREVIEWS
PRINT EMAILSUBJECT
SCORERESULT
BY EMAILID
END

The above procedure joins the single EMAILBODY field in the source Master File (CENTURYREVIEWS) to the TEXT field in the Sentiment Analysis synonym (SCORE).

The SCORERESULT field will contain the sentiment score returned by the Taxonomy Server based on the joined EMAILBODY field from the source Master File (CENTURYREVIEWS).

In this case, the USAGE and ACTUAL values for the TEXT field of the Sentiment Analysis synonym are set to match the EMAILBODY field in the CENTURYREVIEWS synonym.

Reference: Example of a Multiple Field Procedure

This method concatenates multiple fields to create one single body of text to be evaluated for Sentiment Analysis. This requires one pass to concatenate all values, which can then be held for a second pass joining the concatenated fields to the Sentiment Analysis synonym.

PAGE 1				
EMAILID	EMAILDATE	EMAILSUBJECT	SCORERESULT	STORENAME
1	2011/10/27 00:00:00.000	Why did CE's developer relations exec Scott Addams go to a competitor?	.4000000060	EMart
2	2011/07/31 00:00:00.000	A good Option for Tablets	.3999115825	Audio Expert
3	2011/09/09 00:00:00.000	A good Option for Tablets (Updated 1)	.6000000238	AV VideoTown
4	2011/12/15 00:00:00.000	A good Option for Tablets (Updated 2)	.4000000060	City Video
5	2011/08/10 00:00:00.000	Great Device, with option to make it faster!	.5999557972	EMart
6	2011/02/02 00:00:00.000	From a dedicated apple fan This IS a good device	1.0000000000	Audio Expert
7	2011/12/30 00:00:00.000	great great	.4000000060	AV VideoTown
8	2011/12/30 00:00:00.000	worst tablet	.0000000000	Consumer Merchandise
9	2011/12/30 00:00:00.000	Limited usage	.2000000030	EMart
10	2011/11/11 00:00:00.000	CenturyTabletgood	.2000000030	TV City
11	2011/12/29 00:00:00.000	it made me hurt my husband's feelings	.4000000060	AV VideoTown
12	2011/12/29 00:00:00.000	Great for college students on a budget!	.3928571343	City Video
13	2011/12/29	CE CEOS	1000557012	Consumer

SET ASNAMES=ON

DEFINE FILE CENTURYREVIEWS
BODY/A9000=EMAILSUBJECT|' '|EMAILBODY;
END

```
TABLE FILE CENTURYREVIEWS
PRINT BODY
                  AS BODY
   EMAILID
                   AS EMAILID
  EMAILSUBJECT
                  AS EMAILSUBJECT
   SENDERCOMPANY
                  AS STORENAME
                  AS EMAILDATE
  EMAILDATE
                  AS EMAILMONTH
  EMAILMONTH
   EMAILFROMGENDER AS EMAILFROMGENDER
  EMAILDOMAIN
                  AS EMAILDOMAIN
ON TABLE HOLD
END
JOIN BODY IN HOLD TO TEXT IN SCORE
TABLE FILE HOLD
PRINT EMAILDATE
  EMAILSUBJECT
   SCORERESULT
   STORENAME
BY EMAILID
END
```

Note: With this method, the Sentiment Analysis synonym should be edited so that the DEFINE field length matches the length of the TEXT field.

In this case, the USAGE and ACTUAL values for the TEXT field of the Sentiment Analysis synonym are set to match the BODY DEFINE field.

```
Edit Master File as Text: example_sentiment_analysis/score
🔒 🔊 Look for:
                                         Find
                                               Find next
FILENAME=M6ILO, SUFFIX=SOAP , $
 SEGMENT=SCORE, SEGTYPE=S0, $
  GROUP=SCORE, ALIAS=Score, ELEMENTS=3, $
  FIELDNAME=TEXT, ALIAS=text, USAGE=A9000, ACTUAL=A9000
   MISSING=ON, ACCESS PROPERTIX=(NEED VALUE), $
  FIELDNAME=TITLE, ALIAS=title, UkrGE=A30, ACTUAL=A30,
   MISSING=ON, ACCESS PROPERTY=(NEED VALUE), $
  FIELDNAME=UID, ALIAS=uid, USAGE=A30, ACTUAL=A30,
   MISSING=ON, ACCESS_PROPERTY=(NEED_VALUE), $
  FIELDNAME=__RESPONSE, USAGE=TX80L, ACTUAL=TX, ACCESS_PROPERTY=(INTERNAL), $
 SEGMENT=RESPONSE, SEGTYPE=S0, SEGSUF=XML , PARENT=SCORE, POSITION=__RESPONSE, $
  FIELDNAME=RESPONSE, ALIAS=ScoreResponse, USAGE=A1, ACTUAL=A1, ACCESS_PROPERTY=(INTERNAL), $
  FIELDNAME=SCORERESULT, ALIAS=ScoreResult, USAGE=F15.10, ACTUAL=A15,
   REFERENCE=RESPONSE, PROPERTY=ELEMENT, $
```

Alchemy Server for Sentiment Analysis

Alchemy, formally known as Alchemy API, is a sentiment analysis tool that is used to extract and analyze data about people, facts, places, and other topics. Using algorithms and other language processing technologies, Alchemy returns sentiment data that allows you to determine the positive or negative feelings about a particular item.

This section presents the configuration for Alchemy.

Alchemy Configuration

Alchemy is configured in the Adapters section on your WebFOCUS Server.

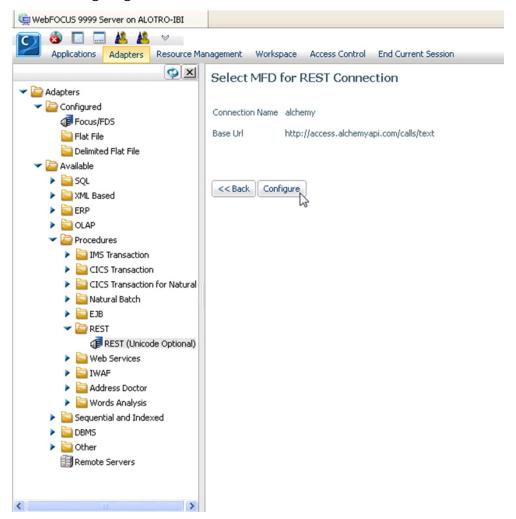
Notes:

- Configuration of and access to a WebFOCUS 7705 Reporting Server is a prerequisite.
- ☐ The REST Adapter that is used for sentiment analysis with Alchemy is currently limited to 2000 characters.

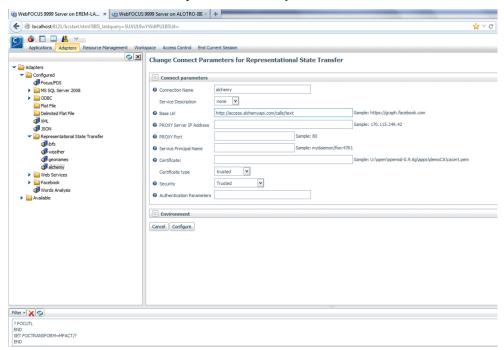
Procedure: How to Configure the REST Adapter to Use Sentiment Analysis with Alchemy

- **1.** Open the Reporting Server Web Console.
- 2. From the Web Console menu bar, click Adapters.
- **3.** Expand the *Available* folder, if it is not already expanded.
- **4.** Expand the *Procedures* folder and then the *REST* folder.

5. Select the REST (Unicode Optional) entry and in the right pane, click *Configure*, as shown in the following image.



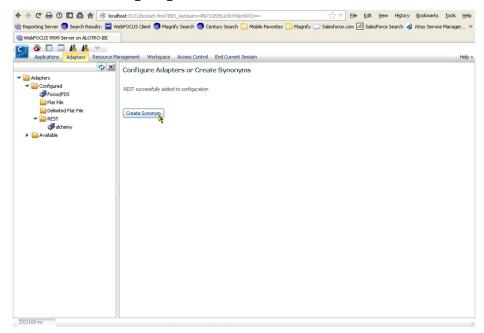
- **6.** Click *Configure* and then click *REST* to expand the list of available adapters.
- **7.** In the *Connection Name* field, enter a name that uniquely identifies the connection, as shown in the following image.



Note: It is recommended that you enter *alchemy* for the connection name.

- **8.** Verify that the Base URL is set to http://access.alchemyapi.com/calls/text.
- **9.** Look in the EDASPROF.PRF to see that the connection string was created.
- 10. Click Adapters and then open the Available folder.
- 11. Click the Procedures folder and then the REST folder.

12. Once REST has been successfully added to the configuration, click *Create Synonym*, as shown in the following image.



- **13.** Using the values shown on the following screen, populate values for the synonym, including:
 - ☐ **Synonym Name.** The name of the synonym you are creating.
 - □ **Select Operation.** The operation to perform (for example, Get).
 - ☐ Service URL Extension. An extension to the service URL.

The Service URL Extension value to be used is:

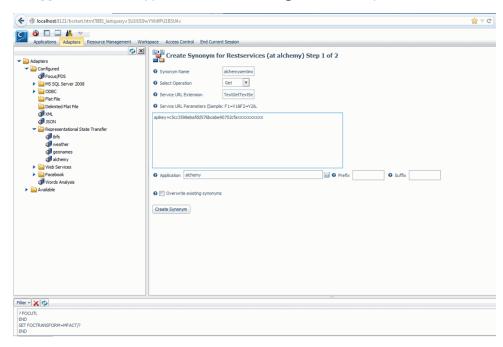
TextGetTextSentiment

□ **Service URL Parameters.** Space for the API key. Contact your WebFOCUS administrator to receive the API key that should be used.

The Service URL Parameters value to be used is:

apikey=xxxxxxxxxxxxxxxxxxxxxxxxxxxxxtext=test

DN4501215.0912



■ Application. The application that this configuration will impact.

Note: The values set to Service URL Parameters must include the valid API key and sample text in order to correctly generate a master file based on the response received. The text value can be any string.

Procedure: How to Update the Master File Description

Note: You must update the Master File description so that you have the right key.

- **1.** Open the Master File and make the following changes to the description:
 - Change the FIELDNAME=APIKEY as indicated below.

Change the FIELDNAME=TEXT field as indicated below (or as dictates by data being used).

```
FIELDNAME=TEXT, ALIAS=text, USAGE=A30, ACTUAL=A30, $
```

to

```
FIELDNAME=DOC, ALIAS=text, USAGE=A300, ACTUAL=A300, $
```

Note: The USAGE and ACTUAL for FIELDNAME DOC will need to be at least the same size or larger than the incoming text being evaluated, otherwise this text will be only evaluated up to the point of truncation.

Creating a Simple FEX with Alchemy

Using the following examples, you can create a simple FEX for use with Alchemy.

```
TABLE FILE ALCHEMYSENTIMENT
PRINT STATUS
SCORE
TYPE
WHERE DOC EQ 'Happy';
END
```

You can then create a FEX using a simple Single Pass JOIN.

```
JOIN EMAILBODY IN CENTURYREVIEWS TO DOC IN ALCHEMYSENTIMENT AS J
TABLE FILE CENTURYREVIEWS
PRINT EMAILSUBJECT
SCORE
STATUS
TYPE
BY EMAILID
END
```

You can also create a Multiple Pass JOIN, with other types of JOINS, including the following example:

```
SET ASNAMES=ON

DEFINE FILE CENTURYREVIEWS

BODY/A9000=EMAILSUBJECT|''|EMAILBODY;

END

TABLE FILE CENTURYREVIEWS

PRINT BODY AS BODY

EMAILID AS EMAILID

EMAILSUBJECT AS EMAILSUBJECT

SENDERCOMPANY AS STORENAME

EMAILDATE AS EMAILDATE

EMAILMONTH AS EMAILMONTH

EMAILFROMGENDER AS EMAILFROMGENDER

EMAILDOMAIN AS EMAILDOMAIN

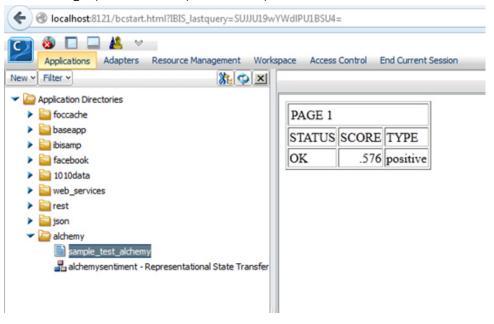
ON TABLE HOLD

END
```

```
JOIN BODY IN HOLD TO DOC IN ALCHEMYSENTIMENT AS J
TABLE FILE HOLD
PRINT EMAILDATE

EMAILSUBJECT
SCORE
STATUS
TYPE
STORENAME
BY EMAILID
END
```

The following report is an example of the output.



Best Practices

The following are best practices for Sentiment Analysis:

- Sentiment Analysis is best suited for user-generated or business-based information. This type of data tends to be found in mixed content, the combination of structured and unstructured information.
 In order to improve efficiency in the overall process of scoring content as it is generated and updated, it is recommended that the score be stored with the data, rather than checked on demand each time a report is executed or a search is performed.
 Unstructured content tends to be larger than data found in typical reports. Machine memory will be affected since information is passed between and evaluated by WebFOCUS and the WAND DataFacet Taxonomy Server for Sentiment Analysis.
 The sentiment score returned in SCORERESULT is a decimal value that should have a precision of 10. Therefore, Dialogue Manager can be used (for example, IF THEN ELSE, DECODE, and so on) to provide labels for sentiment scores tailored by any given range, thereby providing sentiment at any level of granularity.
- as a Category value to filter results. Result sets can also be sorted based on the numerical sentiment score.

Magnify search benefits from Sentiment Analysis information, as Sentiment can be used

WAND DataFacet Taxonomy Server Installation and Configuration

This section reviews the prerequisites and installation information.

This is third-party software provided by WAND Inc. that is resold by Information Builders. The Taxonomy Server evaluates content and returns a sentiment score based on a default Sentiment Taxonomy. The Sentiment Taxonomy can be customized upon request through WAND Professional Services.

Prerequisites

- ☐ WAND DataFacet Taxonomy Server for Sentiment Analysis (Windows only, open on port 4701).
- WebFOCUS Adapter for Web Services (Version 7.7.04 Reporting Server).
- Data (typically user generated content, such as Facebook comments, surveys, or other types of communication-based content. Data can also include business content or other unstructured content.)

WAND DataFacet Taxonomy Server for Sentiment Analysis

The Taxonomy Server evaluates content that it receives from web service calls and returns a score between -1 and 1, where -1 is the worst and 1 is the best.

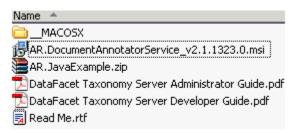
WAND DataFacet Taxonomy Server Requirements

The Taxonomy Server has the following requirements.

Requirement	Taxonomy Server for Windows		
Operating System	Microsoft Windows Server 2008 or higher		
Framework	Windows .Net 3.5 SP1 or higher		
Web Server	Not Required		
CPU Cores	Two or more		
Free RAM	1 GB		
Minimum Storage	100 MB		

Procedure: How to Install the WAND DataFacet Taxonomy Server

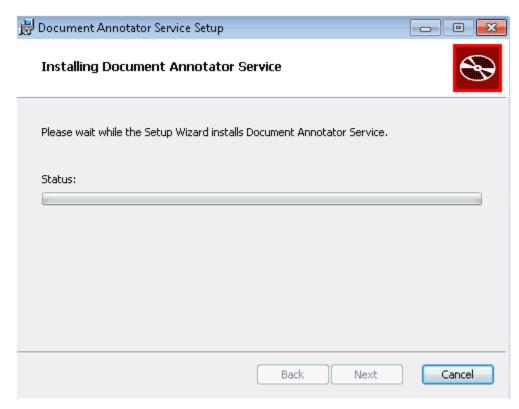
The Taxonomy Server installation contains the following files:



Double-click the AR.DocumentAnnotatorService_v2.1.1323.0.msi file.
 The Software License Agreement dialog box opens, as shown in the following image.



2. Select the *I accept the terms in the License Agreement* check box, and click *Install*. The Setup Wizard installs the Document Annotator Service.





3. When the installation is complete, click *Finish*.

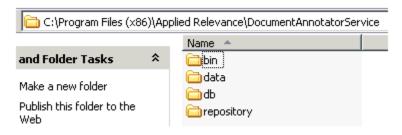
Procedure: How to Configure the WAND DataFacet Taxonomy Server

By default, the Taxonomy Server is configured to localhost. You should modify the configuration to reflect the machine name or IP address so it can be called from remote WebFOCUS installations.

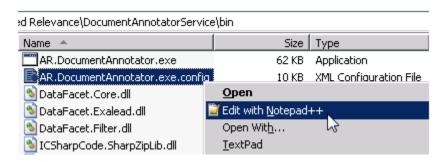
1. Navigate to the default installation directory:

c:\Program Files (x86)\Applied Relevance\DocumentAnnotatorService

2. Navigate to the bin directory.



3. Open the AR.DocumentAnnotator.exe.config file in a text editor.



4. Navigate to line 107 and locate the *baseAddress* attribute.



5. Change the baseAddress value from localhost to the current server name or IP address.



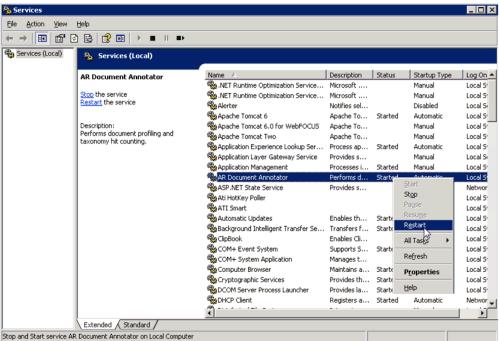
6. Copy the newly configured Taxonomy Server address.



- **7.** Store this value for later reference, since it will be needed to validate the installation and configure the Adapter for Web Services to use Sentiment Analysis.
- 8. Open Services.
- 9. Configure the AR Document Annotator service accordingly.

DN4501215.0912

10. Restart the application server for the changes to take effect.



Procedure: How to Validate the WAND DataFacet Taxonomy Server Installation

1. Open a browser and navigate to the Taxonomy Server URL. You should see a screen similar to the following.

Procedure: How to Uninstall the WAND DataFacet Taxonomy Server for Sentiment Analysis

- 1. Open the Windows Control Panel, then Add/Remove Programs.
- 2. Select Document Annotator Service.
- 3. Click Remove.



DN4501215.0912

Procedure: How to Get the WSDL

By default, the WSDL is http://<hostname>:4701/?wsdl. Navigate to the Taxonomy Server URL to confirm.

- 1. Open a browser and navigate to the Taxonomy Server URL.
- 2. Copy the WSDL location.

