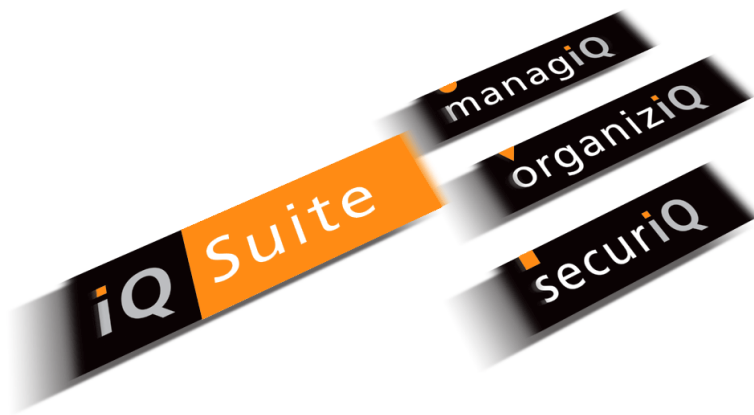


Using CORE in securiQ.Wall



Contents

1 Text Analysis Using CORE - Introduction	3
2 Setting Up Text Analysis Using CORE	3
2.1 Procedure.....	3
2.1.1 Creating the Reference Set of E-Mails.....	3
2.1.2 Collecting and Categorizing the Mails.....	4
2.1.3 Configuring the Analyzers.....	4
2.1.4 The "Teaching Job" – Creating a Classifier.....	8
2.1.5 The "Validation Job" – Checking/Testing the Classifier	11
2.1.6 Adjusting the Categorization	17
2.1.7 Activating CORE Analysis for Incoming Mails.....	18
2.2 Procedure in a Replicated Environment	22
2.2.1 Additional Settings in a Replicated Environment.....	23
2.3 Flow Chart.....	24
3 Company Profile - GROUP Technologies AG	27

1 Text Analysis Using CORE - Introduction

In text analysis based on **CORE Technology**, contents are not checked against wordlists, but "learned" as vector through a text representation, so that messages and documents with texts of the same category can later be identified without using wordlists at all. CORE is independent of wordlists and works with all European languages. Because the senders of **SPAM** tend to work with continually changing (even non-existing) e-mail addresses and content, this technology is especially useful for **blocking SPAM**.

As for a dictionary-based analysis, the **Text Analyzers** and **Converters** are used to categorize the texts and convert them into ASCII format.

To analyze messages and documents with CORE, a representative reference set of mails (SPAM, newsletters, business correspondence, etc.) is copied into a database. For this purpose, GROUP provides a database – **g_learn.nsf**, located in the **grptools** directory – in which documents can be stored and categorized. In training mode, a securiQ.Wall database job "learns" your categories and creates a classifier. In analysis mode, a securiQ.Wall database job then categorizes them for checking. If the checking procedure was successful, activate a securiQ.Wall mail or database job that applies this classifier to all documents, so that all messages and documents with content defined as undesirable are filtered.

2 Setting Up Text Analysis Using CORE

2.1 Procedure

1. Create a reference set of company-typical e-mails
2. Categorize these mails in a training database
3. Configure Analyzers
4. Create a "Teaching-Job"
5. Check the result with a "validation job"
6. Adjust categorization where required
7. If successful, set up final CORE job

2.1.1 Creating the Reference Set of E-Mails

The easiest way to create a reference set of e-mails is to set up a securiQ.Wall job that copies all incoming e-mails to a separate quarantine database without blocking them.



Please note that this job needs to be run prior to your standard securiQ.Wall job and therefore requires a higher priority.

Proceed as follows:

- Create a copy of the quarantine database: **g_arch_core.nsf**
- Activate a securiQ.Wall job to collect mails with the following settings:
 - For mails from Internet
 - Without deleting the documents
 - Without notification
 - To be stored in the previously created copy of the quarantine database:
g_arch_core.nsf



To collect the mails, use the pre-configured job **CORE: Collect Mails from Internet** and adjust it to your requirements.

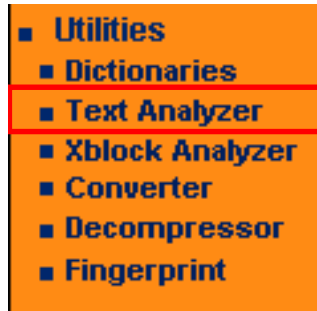
2.1.2 Collecting and Categorizing the Mails

- Collect mails in the **g_arch_core.nsf** quarantine database for at least one night.
- Select a typical set of mails from this database and copy them to the **g_learn.nsf** training database. Only select mails without attachments at this point.
- Categorize the mails manually. For details on how to categorize the mails, please refer to training database online help (Click on **Help** in the **g_learn.nsf** database).
 - Regarding the training database, observe the following:
 - **Copy** the documents to the training database; if they are not forwarded, the mail contents will be falsified!
 - Regarding the categories, observe the following:
 - Define more than 2 but not more than 10 categories, as fewer or more categories are useless.
 - A separation by language is recommended.
 - Category names must be entered in upper case letters, without blank, colon or backslash.
 - Select at least 5 and not more than 250 documents for each category:
Assigning more than 250 documents to a category will not improve the analysis – in that case defining further categories should be preferred

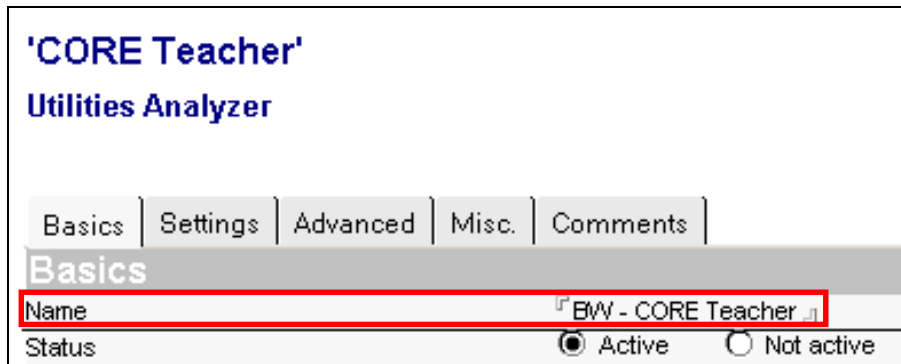
2.1.3 Configuring the Analyzers

Follow the instructions below to create the text analyzers (training and analysis) for the company-specific classifier:

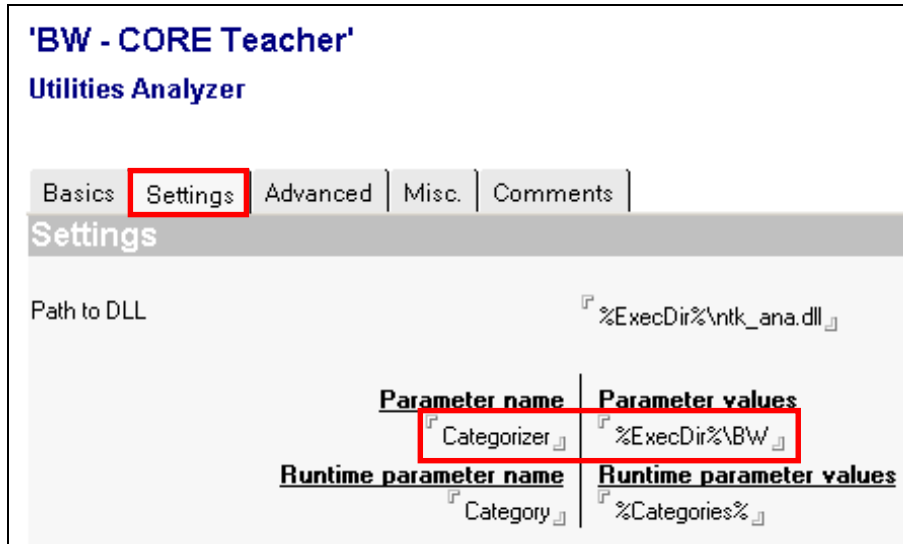
- Select: securiQ → Wall → Utilities → Text-Analyzer:



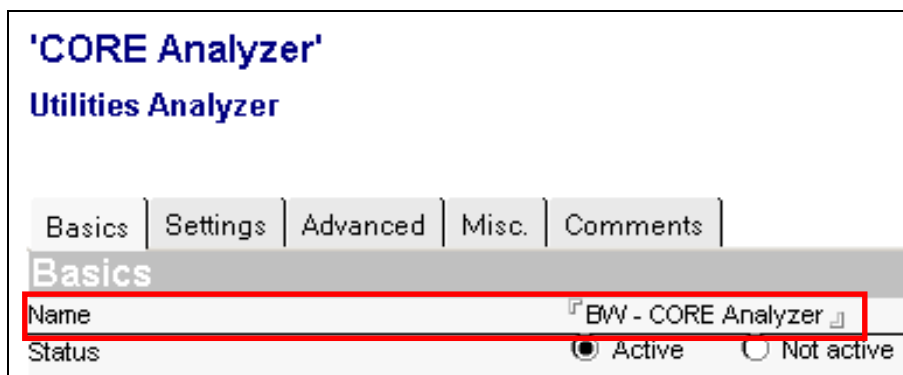
- Copy the existing **CORE Teacher** (be sure to select the correct operating system!) and adjust it.
 - Assign a new name
e.g. <company name or acronym> – CORE Teacher



- Under **Settings**, assign a name to the **Categorizer**:
e.g. <company name or acronym>



- Copy the existing **CORE Analyzer** and adjust it:
 - Assign a new name
e.g. <company name or acronym> – CORE Analyzer



- Under **Settings**, assign a name to the **Categorizer**:
e.g. <company name or acronym>. The name has to correspond to the name assigned to the **CORE Teacher**.

'CORE Analyzer'
Utilities Analyzer

Basics **Settings** Advanced Misc. Comments

Settings

Path to DLL

Parameter name	Parameter values
<input type="text" value="Categorizer"/>	<input type="text" value="%ExecDir%\BW"/>

- Under **Misc.**, enter all categories defined in the training database in the **Supported Categories** field:
e.g.:

'BW - CORE Analyzer'
Utilities Analyzer

Basics Settings Advanced **Misc.** Comments

Misc.

Mode .Xblock image analysis
 Text analysis
 Text training

Categories from dictionaries No
 Yes

Supported categories



If new categories are added in a subsequent training run, be sure to update the Analyzer accordingly! Only the categories entered in this field can be selected from a list when the job is created.

2.1.4 The "Teaching Job" – Creating a Classifier

- Copy the securiQ.Wall database job **Teaching CORE Categories (Body-Subject)** and adjust it:
 - Assign a new name
 e.g. <company name or acronym> – Teaching CORE Categories (Body-Subject)

'BW - Teaching CORE Categories (Body-Subject)'
 securiQ.Wall Database Job

Text training

Basics

Job name

Status Active Not active

Priority

Execution mode **Scheduled** **Event driven**

Start time

Interval Days
 Hours
 Minutes

Database selection +

- Under **Operations**, enter the previously defined text analyzer for the teaching process in the **Text training** field:
<company name or acronym> CORE Teacher

'BW - Teaching CORE Categories (Body-Subject)'
securiQ.Wall Database Job

Text training

Operations

Mode

.Xblock image analysis
 Text analysis
 Text training

Text trainer BW - CORE Teacher ▾

Conversion ▾



Within the teaching job, the CORE Teacher is applied to all categories defined in the training database. It is not possible (and not useful) to "learn" only some the categories of the training database.

- In the **Analyse Elements** field, select the e-mail fields to be analyzed.

Analyse Elements

- Attachments
- Inline pictures
- Text in subject item
- Text in body item
- Merge text items for analysis
- Other text items



To identify SPAM, it is not necessary to analyze attachments as this kind of mail is currently not sent with attachments. Excluding attachments avoids unnecessarily increasing the server load.

If you want to use CORE for more than blocking SPAM and therefore wish to analyze attachments as well, you need a (further) classifier that was specifically trained to analyze mails with attachments, as their structure is different from those without attachments. In that case, you need to create two training databases, one for mails with attachments and the other for mails without attachments. You also need two text analyzers for the training process and two for the analysis process, two securiQ.Wall database jobs for teaching and two securiQ.Wall mail Jobs for analysis – one for mails with attachments and one for mails without attachments. One job only processes mails without attachments and analyzes the subject line and the body text, while the other only processes mails with attachments and processes the attachments in addition to the subject line and the body text.



In the job, specify the analyzer designed to process mails with attachments. In addition, specify three converters (in the order below) in the **Conversion** field:

1. File to XML Extractor
2. XML to Text Converter
3. Text Normalizer

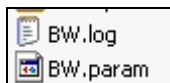
■ Check if the teaching job was completed correctly:

Two documents are created in the **Job Log** for a completed database job:

<ul style="list-style-type: none"> ■ Selection Rules ■ Database <ul style="list-style-type: none"> ■ All Jobs ■ Selection Rules ■ Job Log 	<ul style="list-style-type: none"> ▼ WILLIDOMINO01/WILLIONLINE.DE <ul style="list-style-type: none"> ▼ 1- TEACHING CORE CATEGORIES (BODY-SUBJECT) <ul style="list-style-type: none"> ▼ Jobinfo <ul style="list-style-type: none"> --Jobinfo: next run-- 20.09.2003 12:00:00 ▼ WILLIDOMINO01/WILLIONLINE.DE <ul style="list-style-type: none"> grptools\lg_learn.nsf 19.09.2003 17:13:57 - 19.09.2003 17:14:09
---	---

- Jobinfo: next run--
- run from ... to ...

- In the **grptools** directory on the server, you will find two new files:
 - <company name or acronym>.param: the classifier for the analysis using CORE
 - <company name or acronym>.log: the log file of the teaching process



- If run successfully, deactivate the teaching job.

2.1.5 The "Validation Job" – Checking/Testing the Classifier

The initial check of the classifier created by the training job should be performed using a "validation job" or securiQ.Wall database checking job. A further possibility is to have a securiQ.Wall mail job use the classifier to analyze the incoming mail and check the results obtained.

- Checking the classifier with Wall database "validation" or checking job:
 - Before running the validation job on the training database, set the following ToolKit parameter in the **Notes.ini** file:
`Toolkit_DBGripperOpenExpanded=YES`
The parameter can be set using the "set config" console command or in the iQ.Suite Global Parameters.
 - Copy the securiQ.Wall database job **Validate CORE Categories Training Result (Body-Subject)**, then adjust and enable it:

- Assign a new name
e.g. <company name or acronym> – Validate CORE Categories Training Result (Body-Subject)

'BW - Validate CORE Categories Training Result (Body-Subject)'
securiQ.Wall Database Job

Text analysis

Basics | Operations | Advanced | Misc. | Comments

Basics

Job name	▾ BW - Validate CORE Categories Training Result (Body-Subject) ▾
Status	<input type="radio"/> Active <input checked="" type="radio"/> Not active
Priority	▾ 100 ▾
Execution mode	<input checked="" type="radio"/> Scheduled <input type="radio"/> Event driven
Start time	▾ 04.04.2003 12:41 ▾
Interval	▾ 0 ▾ Days ▾ 0 ▾ Hours ▾ 0 ▾ Minutes
Database selection	+ ▾ a-z ▾ ▾ grptools\q_learn.nsf ▾

- Under **Operations**, enter the **text analyzer** previously defined:
<company name or acronym> CORE Analyzer

'BW - Validate CORE Categories Training Result (Body-Subject)'
securiQ.Wall Database Job


Text analysis

Basics | **Operations** | Advanced | Misc. | Comments

Operations

Mode

.Xblock image analysis
 Text analysis
 Text training

Analyzer	▾ BW - CORE Analyzer ▾
Conversion	▾  Text Normalizer

- In the **Analyse Elements** field, select the same settings as for the teaching job:

Analyse Elements	<input type="checkbox"/> Attachments
	<input type="checkbox"/> Inline pictures
	<input checked="" type="checkbox"/> Text in subject item
	<input checked="" type="checkbox"/> Text in body item
	<input checked="" type="checkbox"/> Merge text items for analysis
	<input type="checkbox"/> Other text items

- In the **Categories** field, enter the analyzer categories with the threshold set to 1:

Categories		Threshold for categories
<input type="button" value="New"/> <input type="button" value="Edit"/> <input type="button" value="Remove"/>		
SPAM-DE	1	<input type="button" value="▲"/>
SPAM-EN	1	<input type="button" value="▼"/>
NEWS	1	
BUSINESS	1	

Click on **New** to add further categories. To change an existing category, select it and click on **Edit**.

- Run the job. The job will now check all mails in the training database and automatically reassign them to the categories.

- The results determined by the job are shown in the training database as further categories (red):

▼ BUSINESS	9	25,7%
▶ BUSINESS	9	100,0%
▼ NEWS	8	22,9%
▶ NEWS	8	100,0%
▼ SPAM-DE	8	22,9%
▶ SPAM-DE	8	100,0%
▼ SPAM-EN	10	28,6%
▶ SPAM-EN	10	100,0%
	35	100,0%

In this example above, the validation job has correctly identified (100%) all categorized mails. Now test the classifier with a Wall mail job that checks the mail traffic.

- Testing the classifier with a securiQ.Wall advanced mail job:

- Set up a Wall advanced job.

- Assign a name, e.g.

<company name or acronym> - TEST Analysis with CORE

Job name	『TEST Analysis with CORE』
Status	<input checked="" type="radio"/> Active <input type="radio"/> Not active
Priority	『10000』
Runs on	<input checked="" type="radio"/> All mails <input type="radio"/> Selected mails

Assign a higher priority than for your standard content checking job and run the job on all mails.

- Under **Operations**, enter the CORE Analyzer:
<company name or acronym> CORE Analyzer

Mode	<input type="radio"/> .Xblock image analysis <input checked="" type="radio"/> Text analysis
Analyzer	BW - CORE Analyzer
Conversion	Text Normalizer
Analyse Elements	<input type="checkbox"/> Attachments <input type="checkbox"/> Inline pictures <input checked="" type="checkbox"/> Text in subject item <input checked="" type="checkbox"/> Text in body item <input checked="" type="checkbox"/> Merge text items for analysis <input type="checkbox"/> Other text items

- Enter **all** categories with the threshold set to 1, in order to check whether each mail is assigned to the correct category:

Categories	Threshold for categories
<div style="display: flex; justify-content: space-between; margin-bottom: 5px;"> New Edit Remove </div> SPAM-DE SPAM-EN NEWS BUSINESS	1 1 1 1

- Let all documents be normally delivered and not deleted. Also disable all notifications, except for those to the Administrator in case of an error (in the **System Errors** tab).

Alarm System Errors	
Alarm	
Delete document	<input checked="" type="radio"/> No <input type="radio"/> Yes
Document in Quarantine?	<input type="radio"/> No <input checked="" type="radio"/> Yes
Category in Quarantine report	TEST CORE
Write analysis details to an e-mail field	<input checked="" type="radio"/> No <input type="radio"/> Yes
Notify administrator	<input checked="" type="radio"/> No <input type="radio"/> Yes
Administrator subject	[SPAM: Denied Content]
Administrator body	This mail contain threshold.
Add analysis details to notification message	<input checked="" type="radio"/> No <input type="radio"/> Yes
Notify recipient	<input checked="" type="radio"/> No <input type="radio"/> Yes
Recipient subject	[Denied Content]
Recipient body	This mail contain threshold.
Add analysis details to notification message	<input checked="" type="radio"/> No <input type="radio"/> Yes
Notify sender	<input checked="" type="radio"/> No <input type="radio"/> Yes
Sender subject	[Denied Content]
Sender body	This mail contain threshold.
Add analysis details to notification message	<input checked="" type="radio"/> No <input type="radio"/> Yes

- In the **Misc.** tab, select the separate quarantine database (**g_arch_core.nsf**) already used for the **CORE Collect Mails from Internet** job, so that all mails are copied to that database:

⌵ * ⌵
⌵ %Admin% ⌵
⌵ Loglevel 0 (value from NOTES.INI) ⌵
⌵ %TEMP% ⌵
⌵ %DataDir%\g_wdog.nsf ⌵
⌵ %DATADIR%\g_arch_core.nsf ⌵
⌵ '%SERVER%' - <product></product> Demon' ⌵

- In the separate quarantine database (**g_arch_core.nsf**), delete all documents, reports and originals in order to check the results.
- Activate the job.

- After some time, check the result in the quarantine database (View **Originals – With Body**): If the result is unsatisfactory (e.g. business mails classified as SPAM), adjust the categorization (see section below).

2.1.6 Adjusting the Categorization

Repeat the following steps as often as required, i.e. until the classification performed by the test job is satisfactory. The goal should be that no business mail is classified as SPAM.

1. Copy the mails wrongly categorized to the training database and assign the correct category.
2. Disable the Wall mail test job.
3. Delete the job log of the database jobs.
4. In the quarantine database, delete all documents for the test job.
5. Run the Wall database teaching job.
6. Check the job log to make sure the teaching job has been completed correctly.
7. After having run the teaching job, disable it and re-run the Wall mail test job.
8. After some time, check the result in the quarantine database (View **Originals – With Body**): If the result is unsatisfactory (e.g. business mails classified as SPAM), go back to Step 1.
9. After several runs, you should check whether the documents in each category are still fairly "similar", e.g.:

- BUSINESS: short text mails
- NEWS: long text mails with many links
- SPAM: MIME Mails

- To obtain a better view of the test job results, use the Designer to adjust the quarantine database as follows:
 In the Originals views, insert an additional column (after the categories) to display the analysis result: (@Word(Check_Details;"=;"1)).

2.1.7 Activating CORE Analysis for Incoming Mails

- Copy the securiQ.Wall mail job **Anti-Spam Based on CORE (Body-Subject)**, then adjust it and finally enable it:
 - Assign a new name
 e.g. <company name or -acronym> – Anti-Spam Based on CORE (Body-Subject)

'BW - Anti-Spam Based on CORE (Body-Subject)'
securiQ.Wall Mail Job Advanced

Text analysis

Basics | Operations | Misc. | Comments

Basics

Job name	『BW - Anti-Spam Based on CORE (Body-Subject)』
Status	<input type="radio"/> Active <input checked="" type="radio"/> Not active
Priority	『100』
Runs on	<input type="radio"/> All mails <input checked="" type="radio"/> Selected mails
Attachment dependency	<input checked="" type="radio"/> All <input type="radio"/> Only with attachment
Positive selection rule dependency	<input checked="" type="radio"/> All true <input type="radio"/> At least one true
	『InetSender』
Negated selection rule dependency	<input checked="" type="radio"/> All false <input type="radio"/> At least one false
	『』
Selection rule summary	(InetSender)

The job is run on all incoming Internet mails with the **InetSender** selection rule.

- Under **Operations**, enter the **text analyzer** previously defined:
<company name or -acronym> CORE Analyzer

Basics	Operations	Misc.	Comments
Operations			
Mode	<input type="radio"/> .Xblock image analysis		
	<input checked="" type="radio"/> Text analysis		
Analyzer	BW - CORE Analyzer		
Conversion	Text Normalizer		
Analyse Elements	<input type="checkbox"/> Attachments		

- In the **Analyse Elements** field, select the same settings as for the teaching, validation and mail checking jobs:

Analyse Elements	<input type="checkbox"/> Attachments
	<input type="checkbox"/> Inline pictures
	<input checked="" type="checkbox"/> Text in subject item
	<input checked="" type="checkbox"/> Text in body item
	<input checked="" type="checkbox"/> Merge text items for analysis
	<input type="checkbox"/> Other text items

- In the **Categories** field, enter the categories to be blocked with threshold set to 1:

<u>Categories</u>	<u>Threshold for categories</u>
<input type="button" value="New"/> <input type="button" value="Edit"/> <input type="button" value="Remove"/>	
SPAM-DE SPAM-EN	1 1 <input type="button" value="▲"/> <input type="button" value="▼"/>

Click on **New** to add further categories. To change an existing category, select it and click on **Edit**. To delete a category, select it and click on **Remove**.

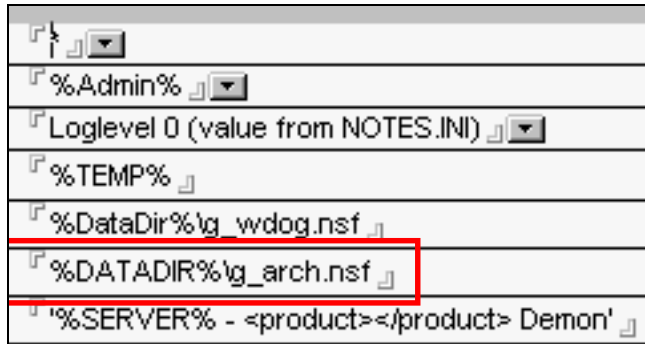


Only enter the mail categories the recipients are not supposed to receive, i.e. typically SPAM!. Remove any other category from this screen.

- Let the documents be deleted and moved to the quarantine. Define a meaningful category for the quarantine. Enabled notifications on both the **Alarm** and **System Errors** tab as required.

Alarm	System Errors
Alarm	
Delete document	<input type="radio"/> No <input checked="" type="radio"/> Yes
Document in Quarantine?	<input type="radio"/> No <input checked="" type="radio"/> Yes
Category in Quarantine report	『SPAM with CORE』
Write analysis details to an e-mail field	<input checked="" type="radio"/> No <input type="radio"/> Yes
Notify administrator	<input type="radio"/> No <input checked="" type="radio"/> Yes
Administrator subject	『[SPAM: Denied Content]』
Administrator body	『This mail conta threshold.』
Add analysis details to notification message	<input type="radio"/> No <input checked="" type="radio"/> Yes
Notify recipient	<input type="radio"/> No <input checked="" type="radio"/> Yes
Recipient subject	『[Denied Content]』
Recipient body	『This mail conta threshold.』
Add analysis details to notification message	<input type="radio"/> No <input checked="" type="radio"/> Yes
Notify sender	<input type="radio"/> No <input checked="" type="radio"/> Yes
Sender subject	『[Denied Content]』
Sender body	『This mail conta threshold.』
Add analysis details to notification message	<input type="radio"/> No <input checked="" type="radio"/> Yes

- In the **Misc.** tab, enable your usual quarantine database (typically **g_arch.nsf**):



- After some time, check the result in the quarantine database (View **Originals – With Body**): If the result is unsatisfactory (e.g. business mails classified as SPAM), read-just the categorization.

2.2 Procedure in a Replicated Environment

In a replicated environment with two or more servers, you have two different options:

1. Run the teaching process on one server and then copy the classifier file **<company name or acronym>.param** to the `grptools` directory of the other servers. This classifier will then be used by the analysis jobs.
2. Run the teaching process on each server, i.e. replicate the training database with the categorized documents on the other servers. The teaching will then be executed on each server with the same reference set.

Then proceed as described in Sections 2.1.1 through 2.1.5.

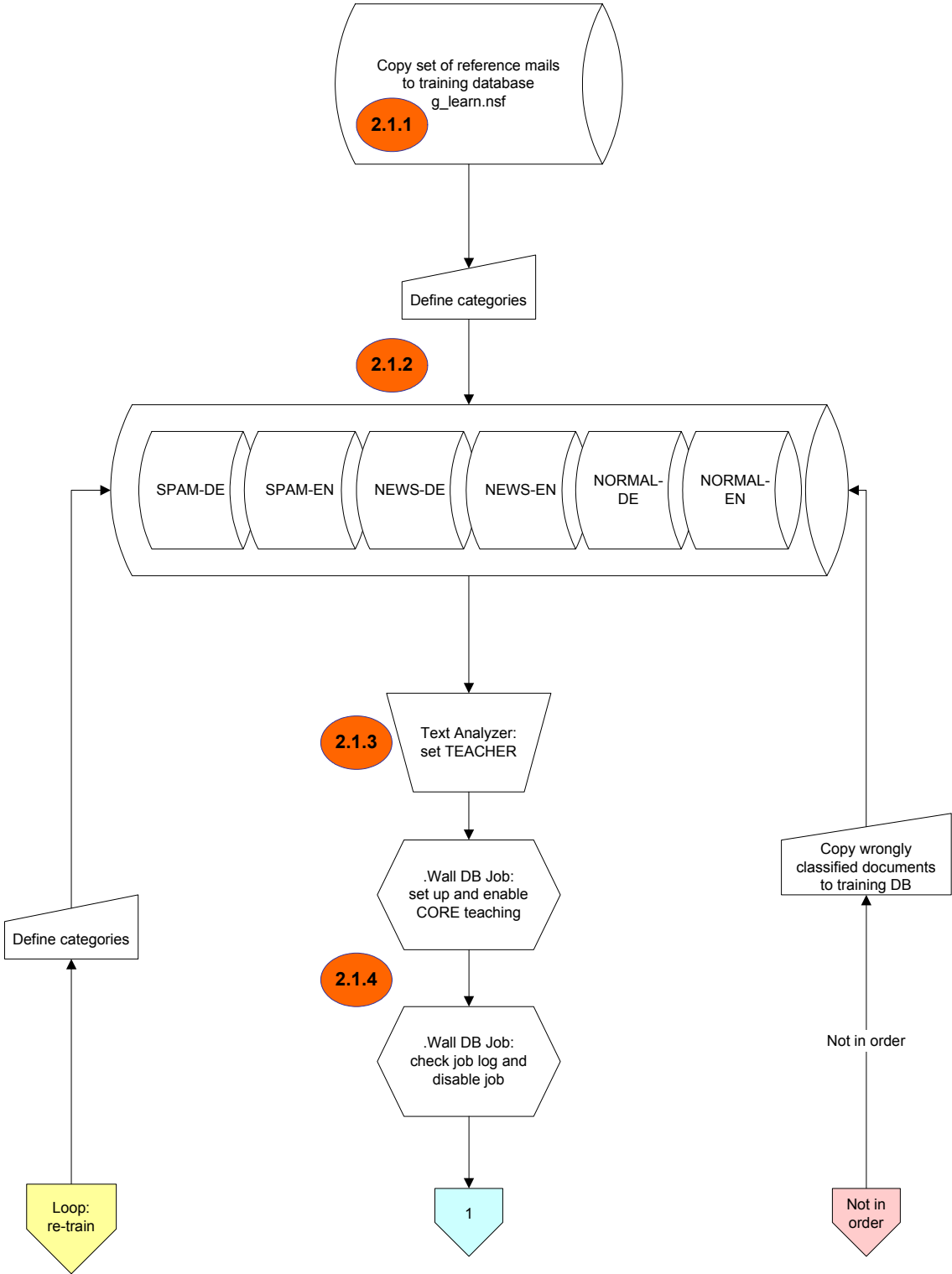
Further settings in a replicated environment include the following:

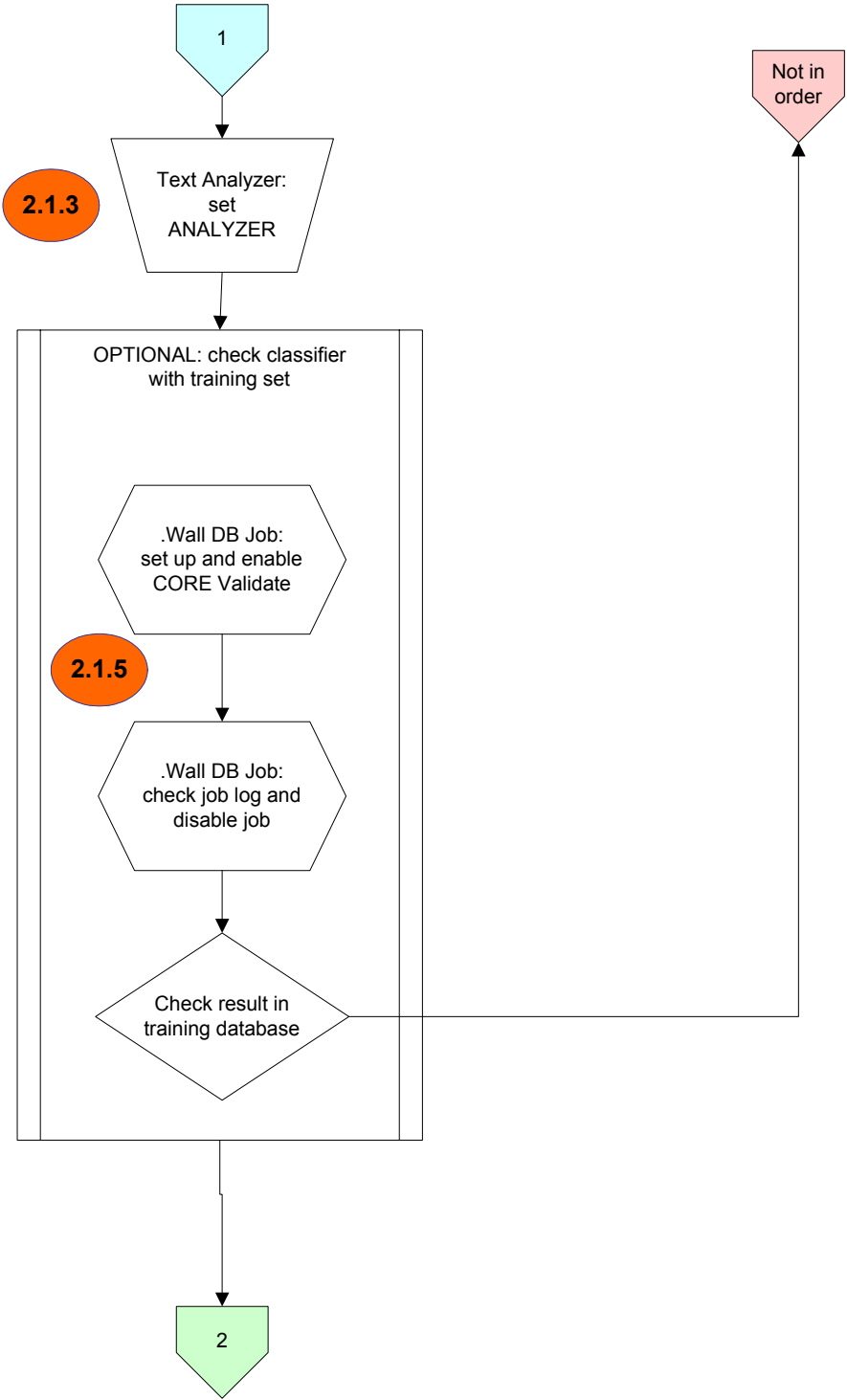
2.2.1 Additional Settings in a Replicated Environment

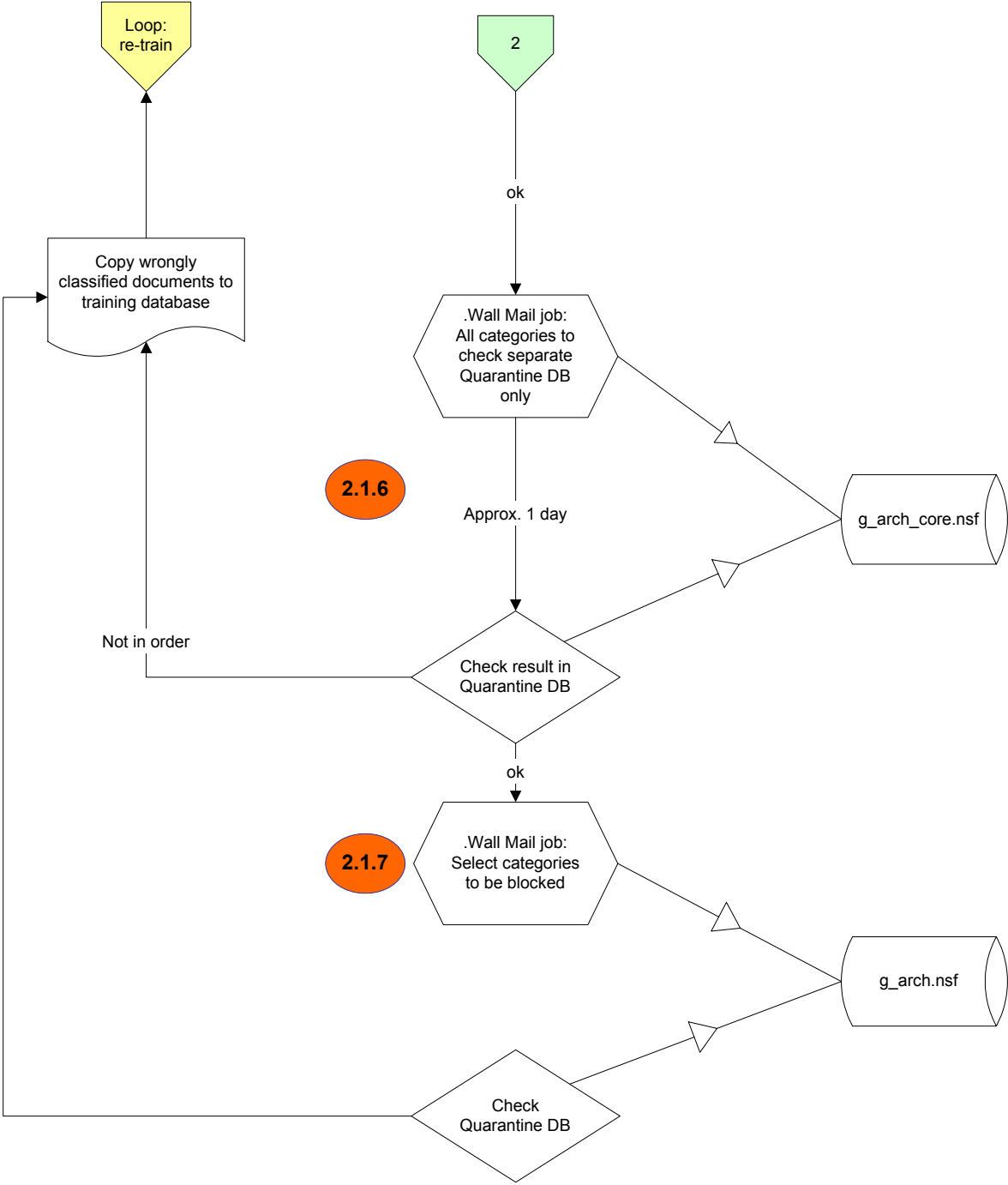
Repeat the following steps as often as required, i.e. until the classification result is satisfactory:

1. Copy the mails wrongly categorized to the training database and assign the correct category.
2. Disable the mail test job
→ Replicate the `grptools` directory with the second server.
3. Delete the job log of the database jobs.
4. In the quarantine database (`g_arch_core.nsf`), delete all documents for the test job.
5. Run the database teaching job
→ Replicate the `grptools` directory with the second server.
6. Check the job log to make sure the teaching job has been completed correctly (on both servers!).
7. After having run the teaching jobs on both servers, disable them and re-run the Wall mail test job.
→ Replicate the `grptools` directory with the second server.
8. After some time, check the result in the quarantine database (View **Originals – With Body**): If the result is unsatisfactory (e.g. business mails classified as SPAM), go back to Step 1.

2.3 Flow Chart







3 Company Profile - GROUP Technologies AG

GROUP Technologies AG is one of the world's leading manufacturers of e-mail security, organization and management software. The company's innovative and forward-thinking products have made GROUP one of the leaders in technologies and innovations in these areas. The optimally coordinated products are available for the Lotus Notes, Microsoft Exchange and SMTP platforms.

Using GROUP's iQ.Suite enables medium-sized and large-scale companies to optimize the cost and efficiency of their e-mail environment and raise work productivity.

The iQ.Suite is modular, scalable company-wide and offers customers the required degree of investment security. Through its fully server-based architecture, iQ.Suite can be administered centrally and economically. Our performance ranges from e-mail cryptography and virus protection to anti-spam and secure archiving of e-mails.

The products are available through direct sale and from OEM and trading partners. GROUP Technologies AG has been trading on the stock market since November 2000.

GROUP Technologies AG's customer base includes a wide variety of renowned companies, such as ABB, Deutsche Bank, Ernst & Young, Honda and Toshiba. Over two million users utilize GROUP Technologies AG products to protect their systems.

GROUP Technologies AG headquarters is in Karlsruhe, Germany. The company maintains offices internationally both in Europe and in Boston, USA.

www.group-technologies.com

© 2003 GROUP Technologies AG

The product descriptions are general and descriptive in nature. They can be interpreted neither as a promise of specific properties nor as a declaration of guarantee or warranty. The specifications and design of our products can be changed at any times without prior notice, especially to keep pace with technical developments.

The information contained in this documentation deals with issues as assessed by GROUP Technologies AG at the time of publication. As GROUP Technologies AG is bound to react to changing market requirements, this document by no means represents an obligation by GROUP Technologies AG and GROUP cannot guarantee the correctness of the information presented in this document after its publication.

This documentation is intended for information purposes only. GROUP Technologies AG hereby excludes any warranty, express or implied, for this document. GROUP Technologies AG is unable to guarantee, either explicitly or tacitly, the quality, execution, standardization or suitability for a specific purpose.

All product or company names in this document may be protected brand names of their respective owners.

Headquarters

GROUP Technologies AG

Ottostrasse 4

76227 Karlsruhe / Germany

Phone +49(0)721-4901-0

Fax +49(0)721-4901-199

info.de@group-technologies.com

www.group-technologies.com



North American Headquarters

GROUP Technologies

321 Fortune Blvd.

Milford, MA 01757/USA

Phone +1 508-473-3332

Phone 877-476-8755 (US and Canada)

Fax +1 508-473-9940

info.us@group-technologies.com

www.group-technologies.com