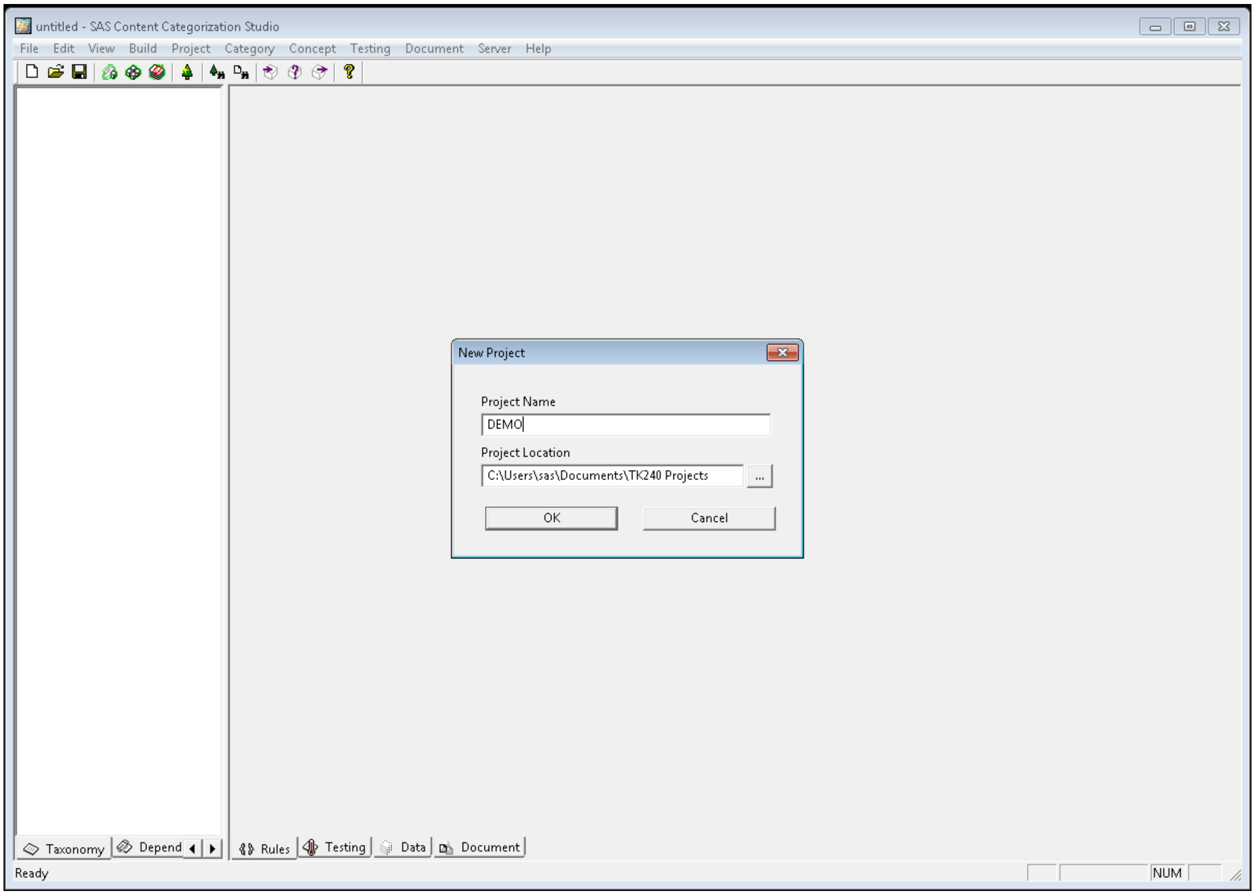
Here are steps to create a SAS Content Categorization project. There will be steps to create both a category rule as well as a concept rule. Each rule will be built or compiled, and then testing on the rule will be done on a set of documents that is part of the zip file.

Part 1: Creating a project

1. Create New Project and set the Location where the new Project Folder will be stored.



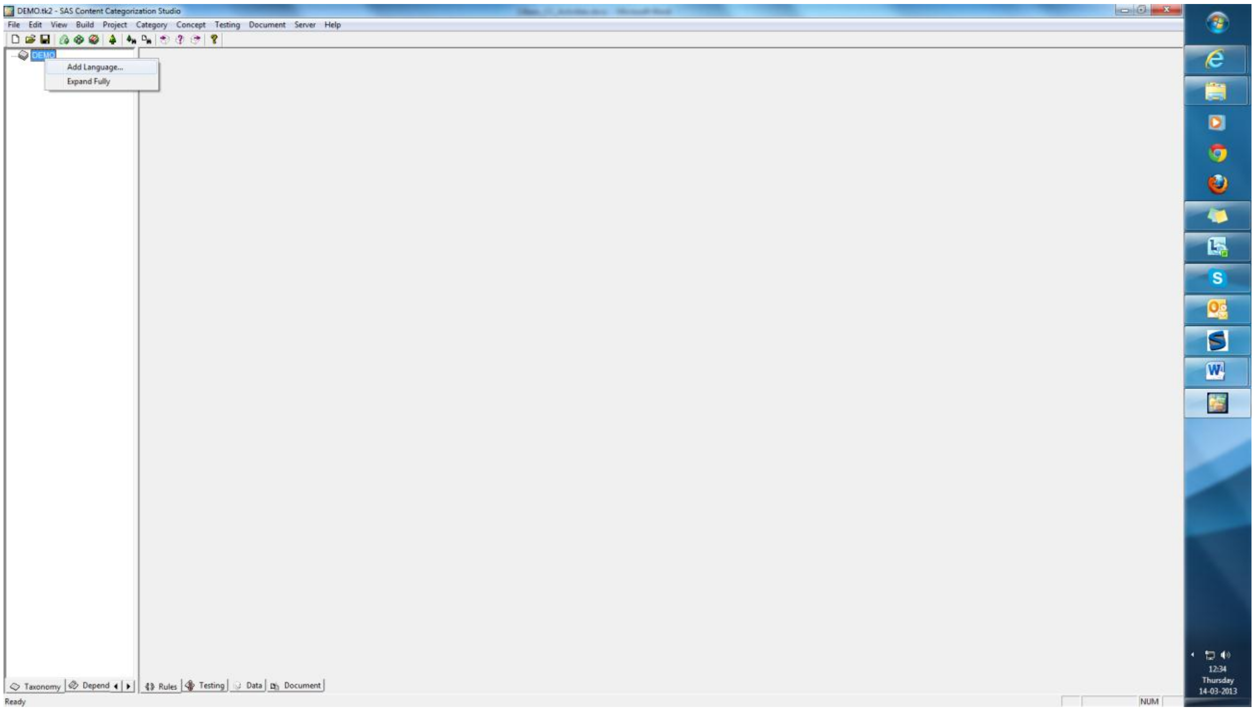
New Project

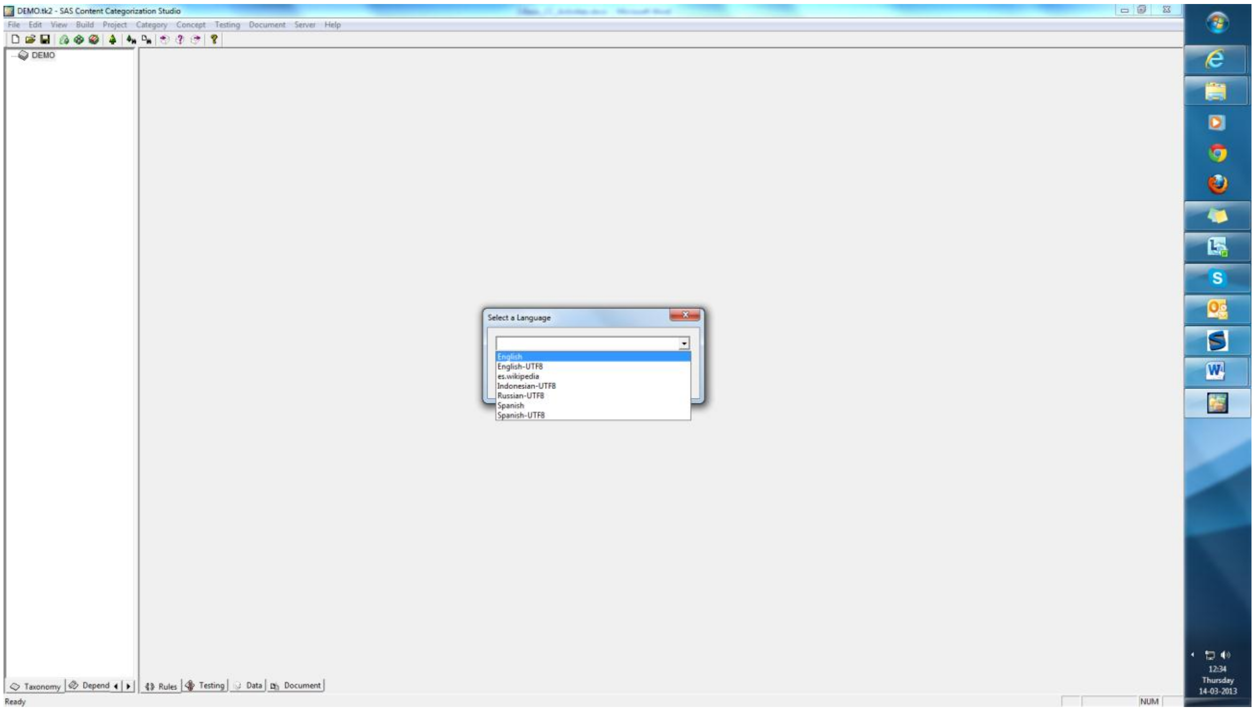
Project Name: DEMO

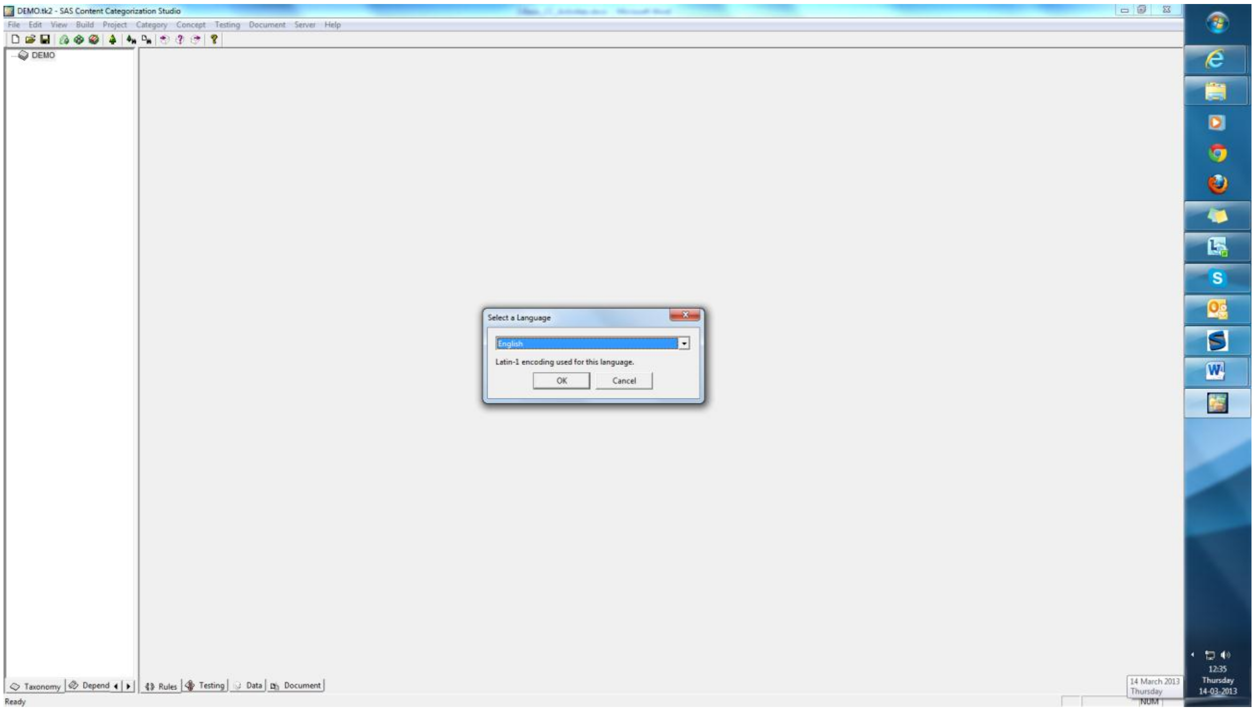
Project Location: C:\Users\sas\Documents\TK240 Projects

Now we will begin creating rules in the newly created *Demo* project.

2. Add language to the new project *Demo*


DEMO.tk2 - SAS Content Categorization Studio
File Edit View Build Project Category Concept Testing Document Server Help
DEMO
Add Language...
Expand Fully
Taxonomy
Depend
Rules
Testing
Data
Document
Ready
NUM
12:34
Thursday
14-03-2013

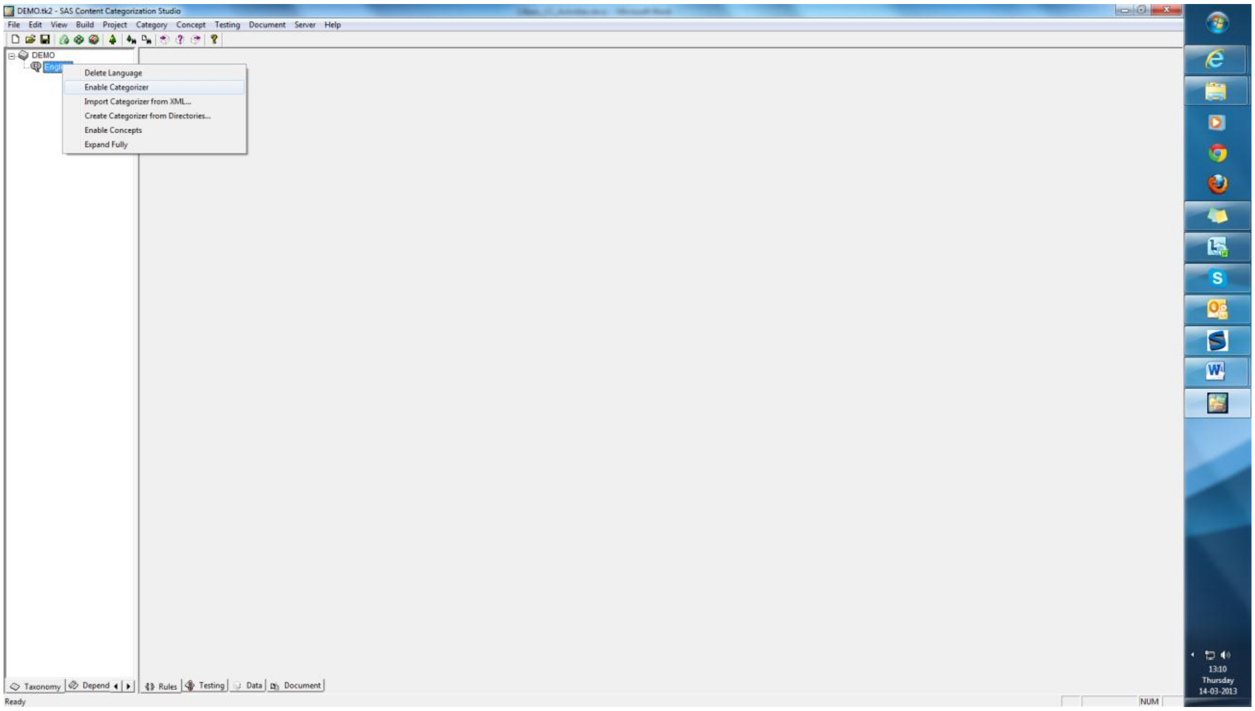

DEMO.tk2 - SAS Content Categorization Studio
File Edit View Build Project Category Concept Testing Document Server Help
DEMO
Select a Language
English
English-UTF8
es.wikipedia
Indonesian-UTF8
Russian-UTF8
Spanish
Spanish-UTF8
Taxonomy
Depend
Rules
Testing
Data
Document
Ready
NUM
12:34
Thursday
14-03-2013

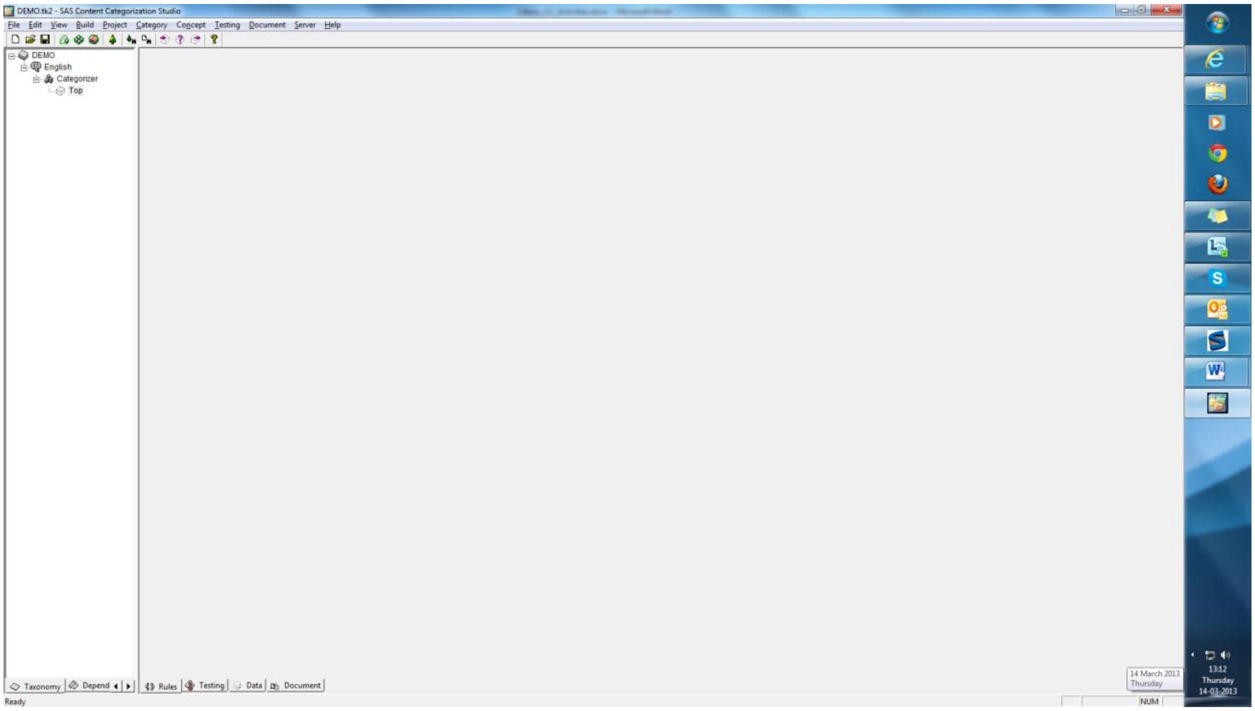

DEMO.tk2 - SAS Content Categorization Studio
File Edit View Build Project Category Concept Testing Document Server Help
DEMO
Select a Language
English
OK
Cancel
Taxonomy
Depend
Rules
Testing
Data
Document
Ready
NUM
12:35
Thursday
14-03-2013

Latin-1 encoding used for this language.

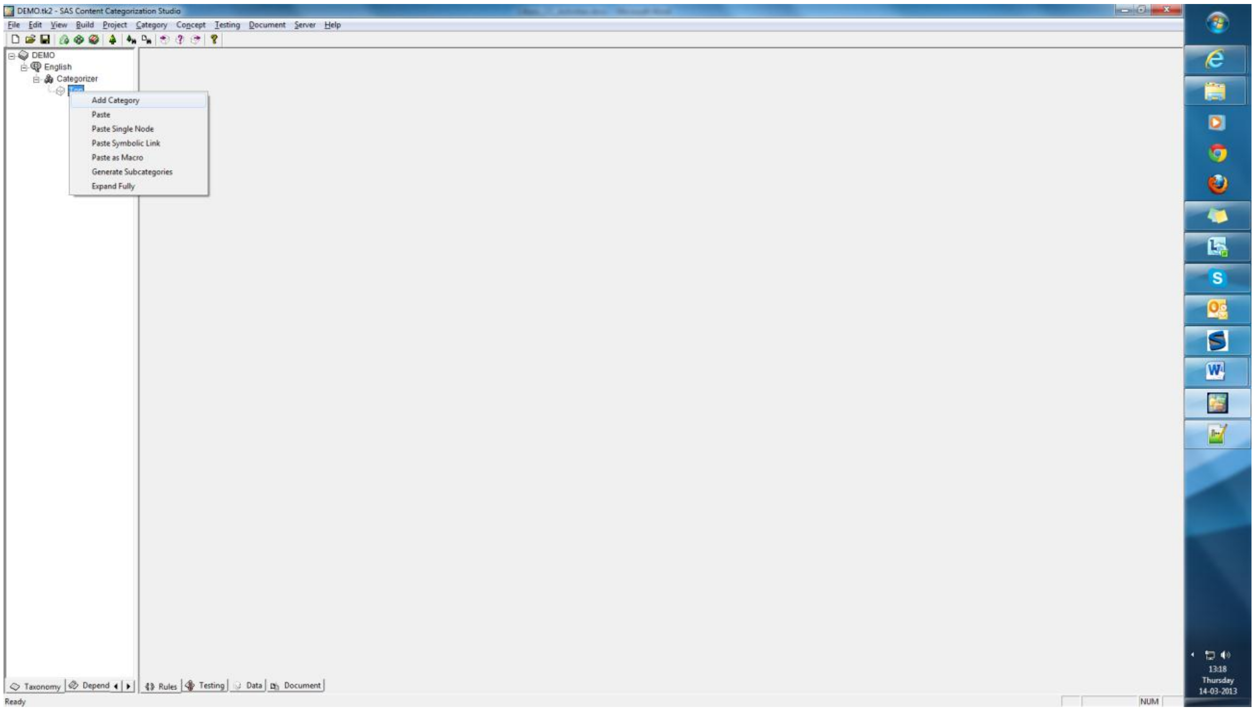
Part 2: Creating a Category

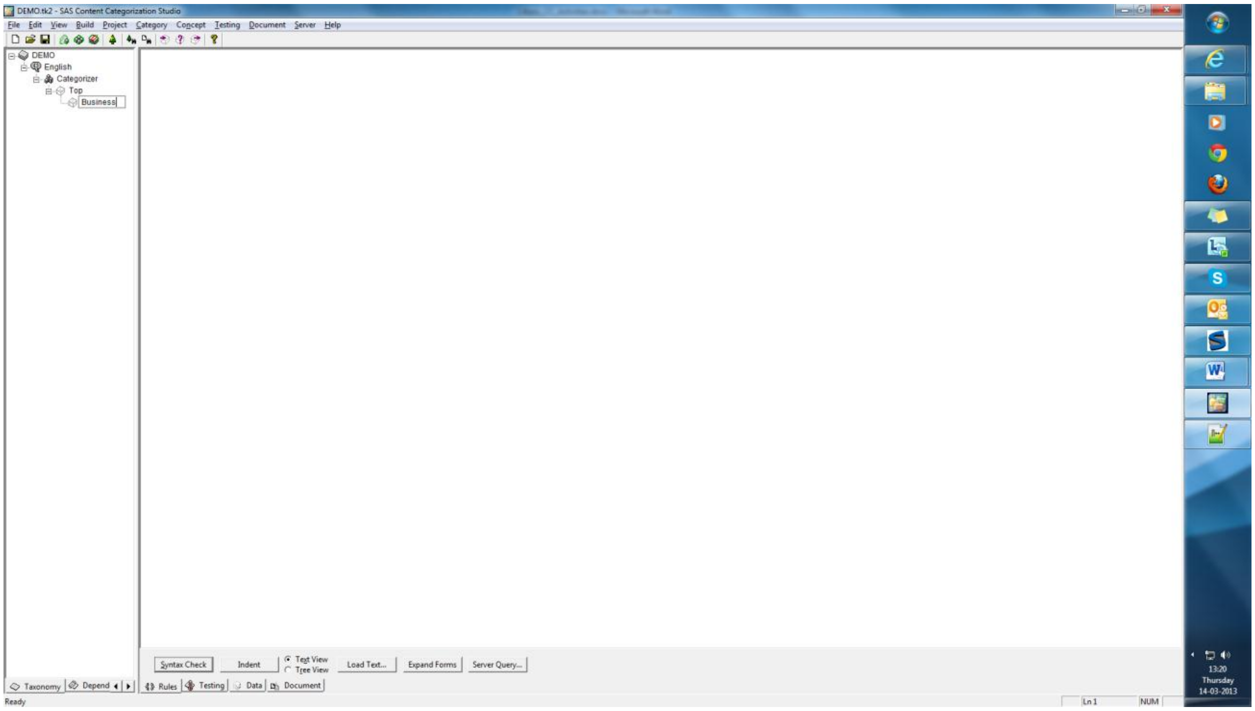
1. To create a category taxonomy, right click on English and choose Enable Categorizer.


DEMO.tk2 - SAS Content Categorization Studio
File Edit View Build Project Category Concept Testing Document Server Help
DEMO
English
Delete Language
Enable Categorizer
Import Categorizer from XML...
Create Categorizer from Directories...
Enable Concepts
Expand Fully
Taxonomy
Depend
Rules
Testing
Data
Document
Ready
NUM
13:10
Thursday
14-03-2013


DEMO.tk2 - SAS Content Categorization Studio
File Edit View Build Project Category Concept Testing Document Server Help
DEMO
English
Categorizer
Top
Taxonomy
Depend
Rules
Testing
Data
Document
14 March 2013
Thursday
Ready
NUM
13:12
Thursday
14-03-2013

2. Next, create a category by right mouse clicking on Top and select Add Category

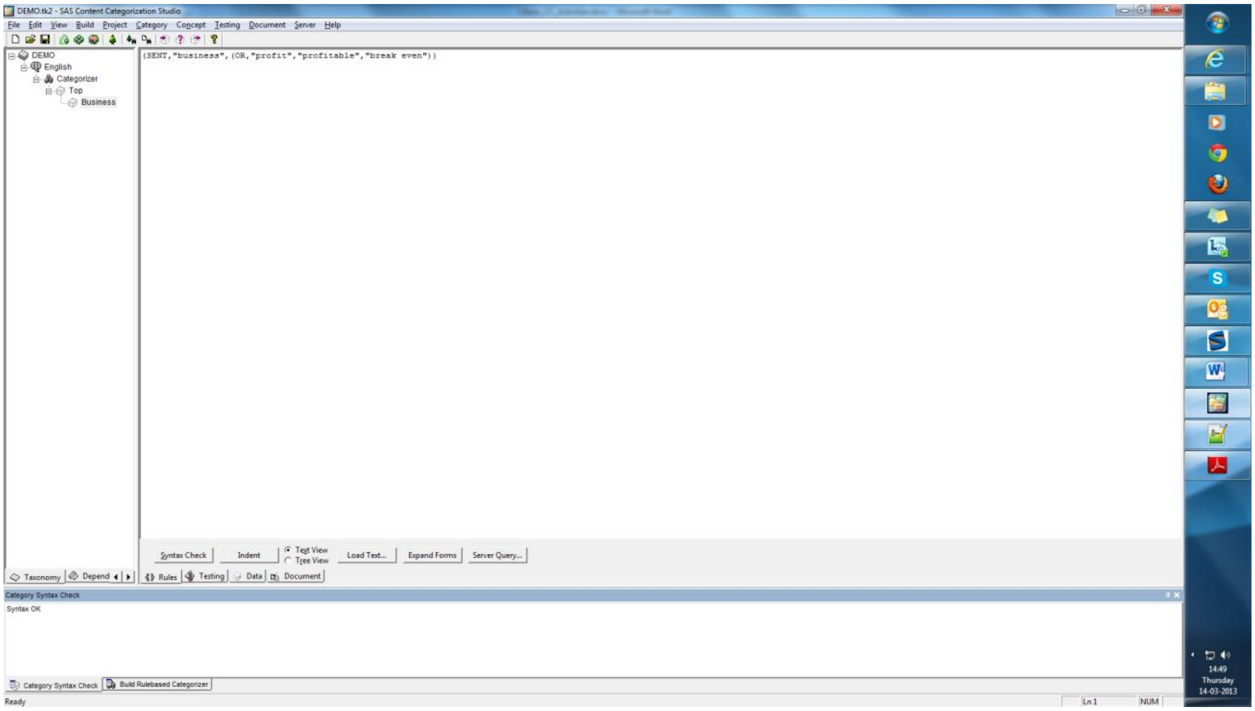




On the rule tab of the Business category write the rules using Text View or Tree View.

3. Within Text View, in the rule tab insert the following rule:

```
(SENT,business,(OR,profit,profitable,break even))
```

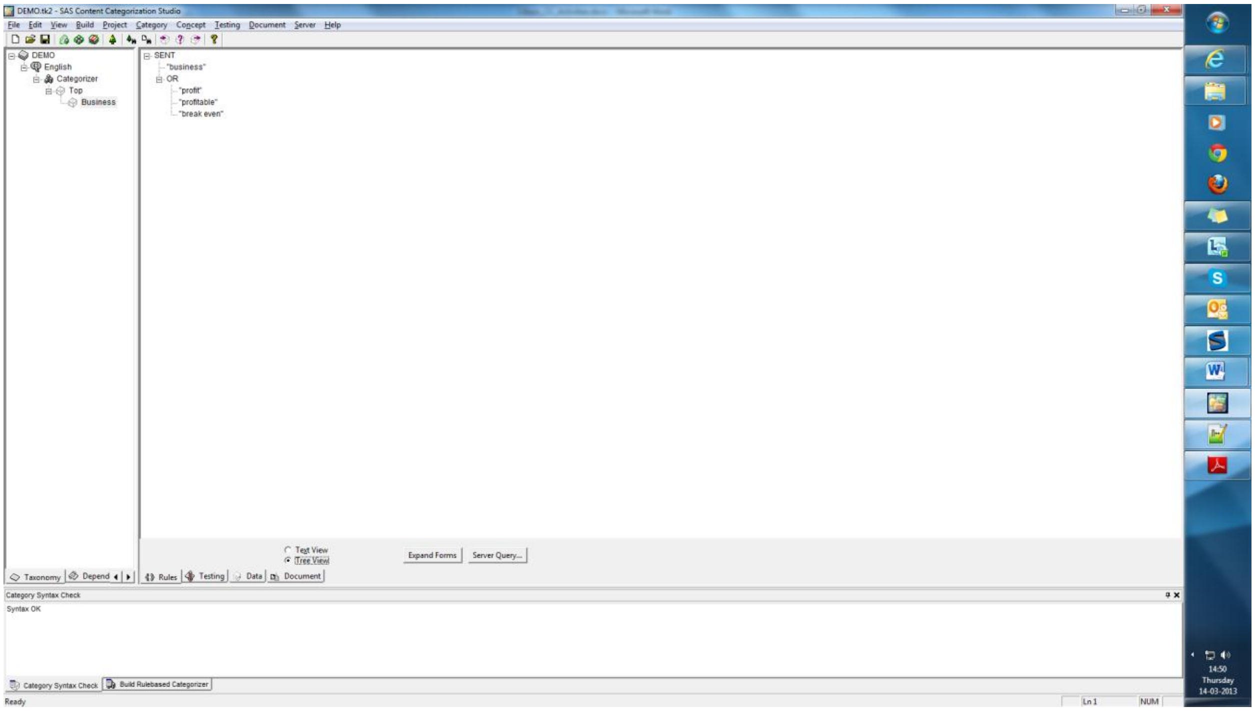


```
(SENT,"business",(OR,"profit","profitable","break even"))
```

Category Syntax Check

```
Syntax OK
```

Below is a screen shot of the tree view of the same rule.

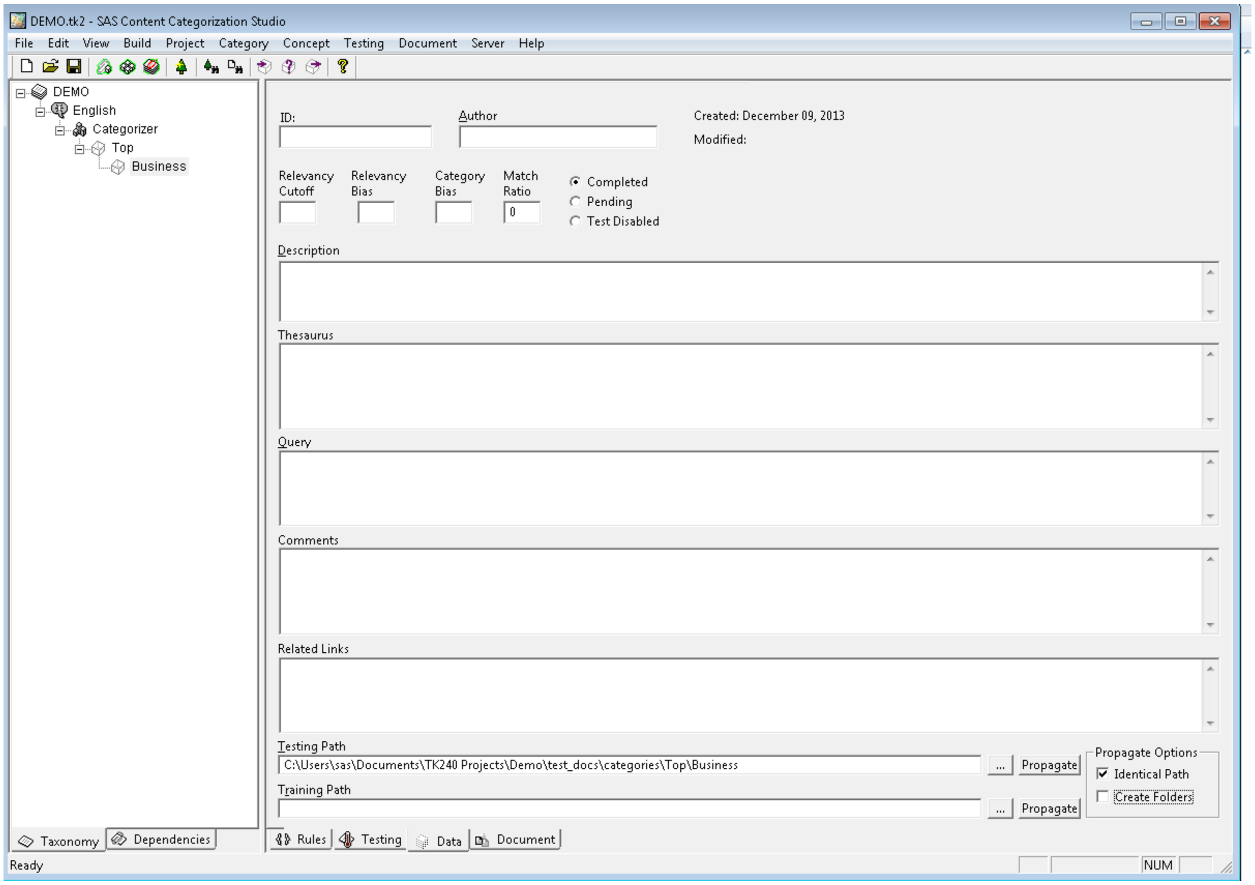

DEMO.tk2 - SAS Content Categorization Studio
File Edit View Build Project Category Concept Testing Document Server Help
Taxonomy
Depend
Rules
Testing
Data
Document
○ Text View
● Tree View
Expand Forms
Server Query...
Category Syntax Check
Build Rulebased Categorizer
Ready
Ln 1
NUM
14:50
Thursday
14-03-2013

```
DEMO
  English
    Categorizer
      Top
        Business
```

```
SENT
  "business"
  OR
    "profit"
    "profitable"
    "break even"
```

```
Syntax OK
```

4. On the data tab, set the path of the folder where the files to be tested are located and then click Propagate.



ID:

Author

Created: December 09, 2013

Modified:

Relevancy Cutoff

Relevancy Bias

Category Bias

Match Ratio: 0

● Completed
○ Pending
○ Test Disabled

Description

Thesaurus

Query

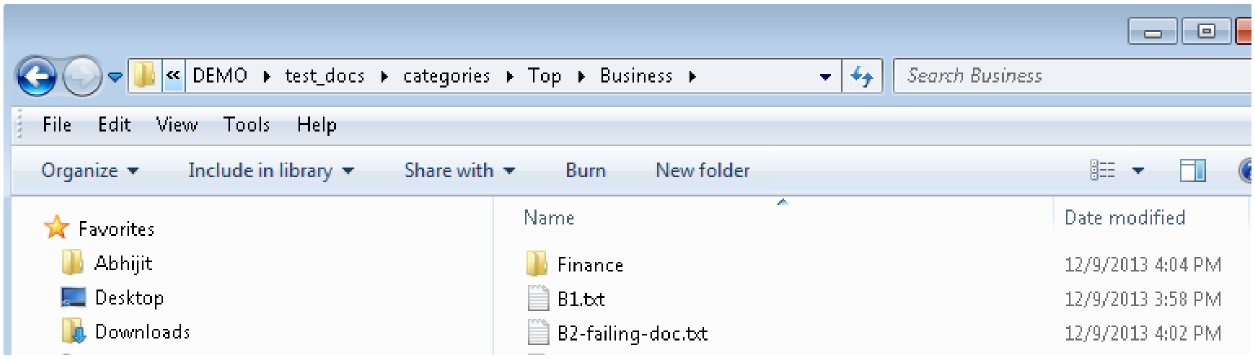
Comments

Related Links

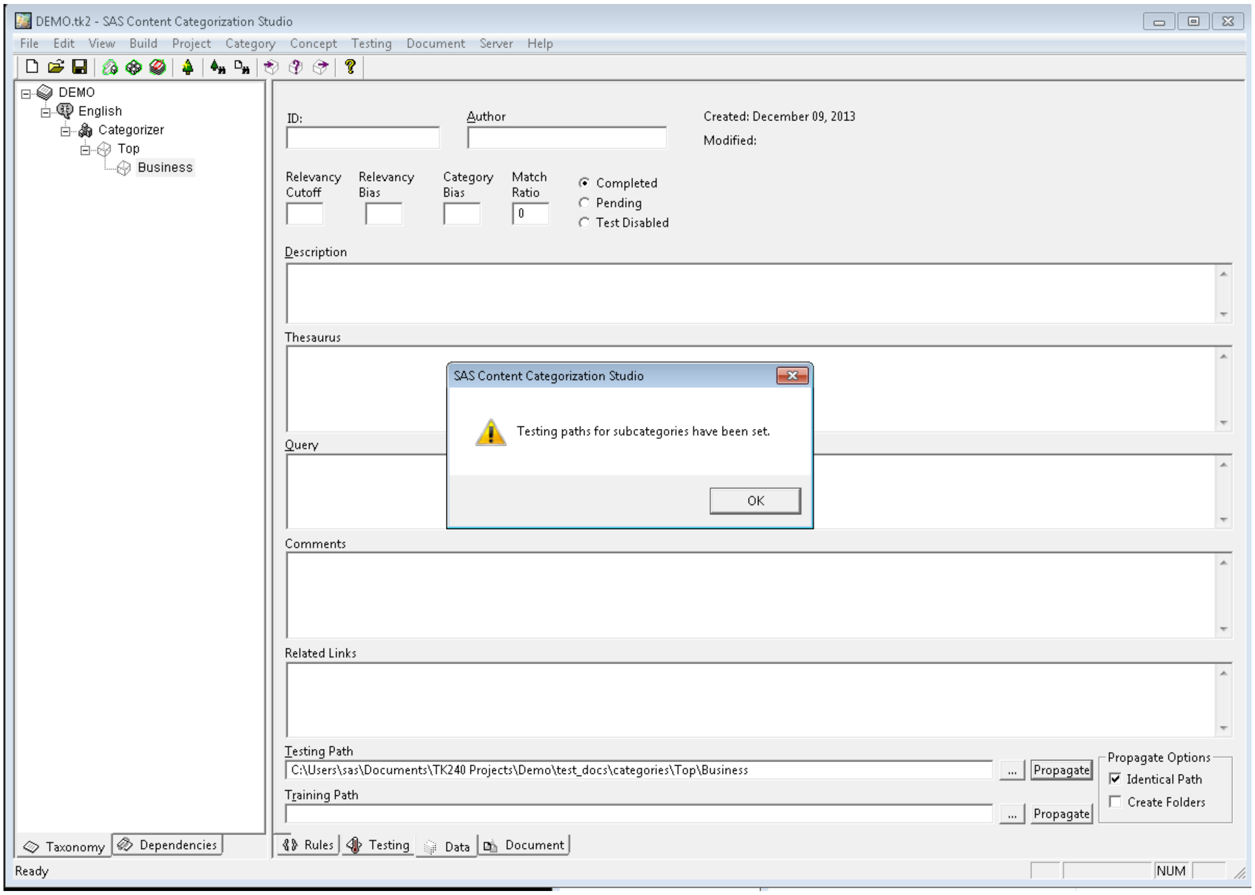
Testing Path: C:\Users\sas\Documents\TK240 Projects\Demo\test_docs\categories\Top\Business ... Propagate

Training Path ... Propagate

Propagate Options
☑ Identical Path
☐ Create Folders



Name	Date modified
Finance	12/9/2013 4:04 PM
B1.txt	12/9/2013 3:58 PM
B2-failing-doc.txt	12/9/2013 4:02 PM



DEMO.tk2 - SAS Content Categorization Studio
File Edit View Build Project Category Concept Testing Document Server Help
DEMO
English
Categorizer
Top
Business
Taxonomy
Dependencies
Rules
Testing
Data
Document
Propagate
Propagate
OK
Ready
NUM

SAS Content Categorization Studio

Testing paths for subcategories have been set.

ID:

Author

Created: December 09, 2013

Modified:

Relevancy Cutoff

Relevancy Bias

Category Bias

Match Ratio: 0

● Completed
○ Pending
○ Test Disabled

Description

Thesaurus

Query

Comments

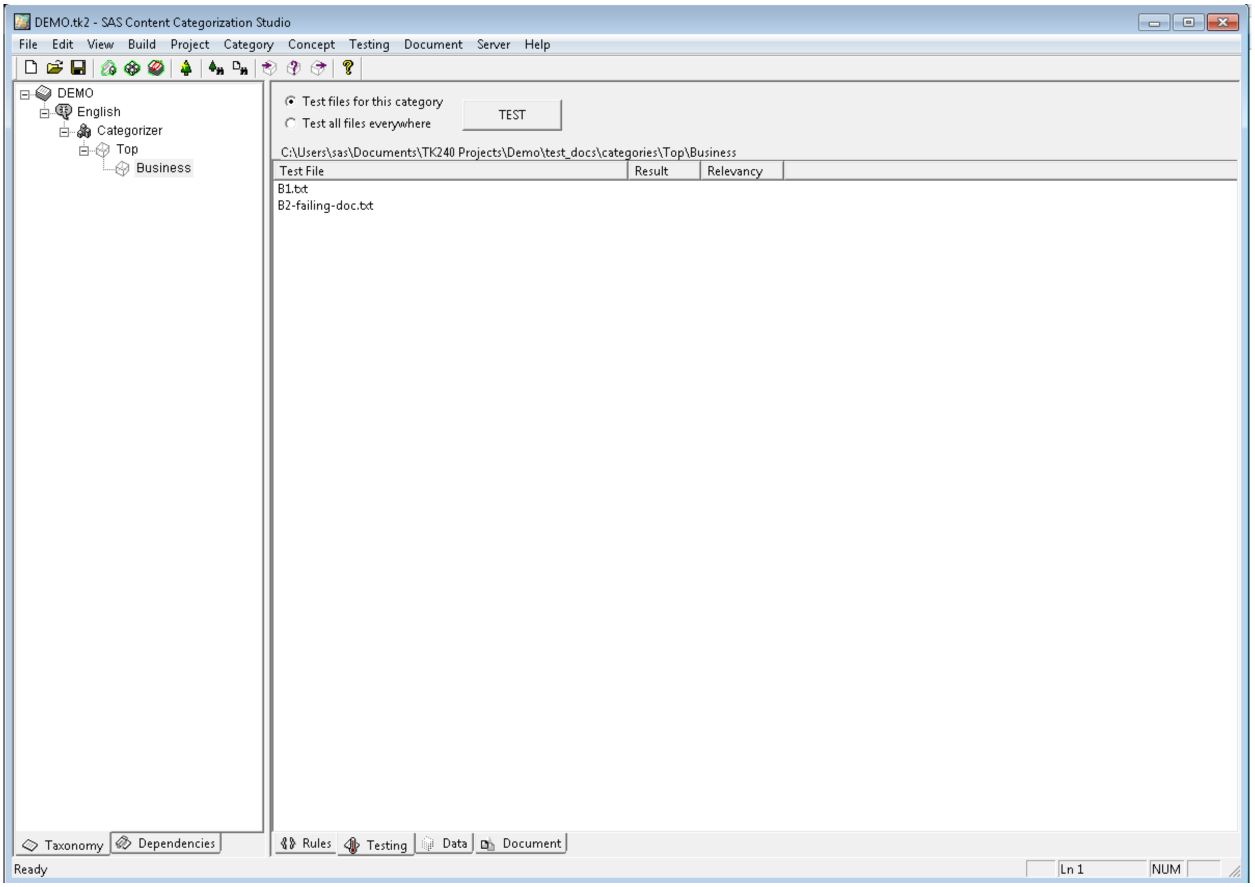
Related Links

Testing Path: C:\Users\sas\Documents\TK240 Projects\Demo\test_docs\categories\Top\Business

Training Path

Propagate Options
☑ Identical Path
☐ Create Folders

The files in the testing folder will now be visible in the testing tab.

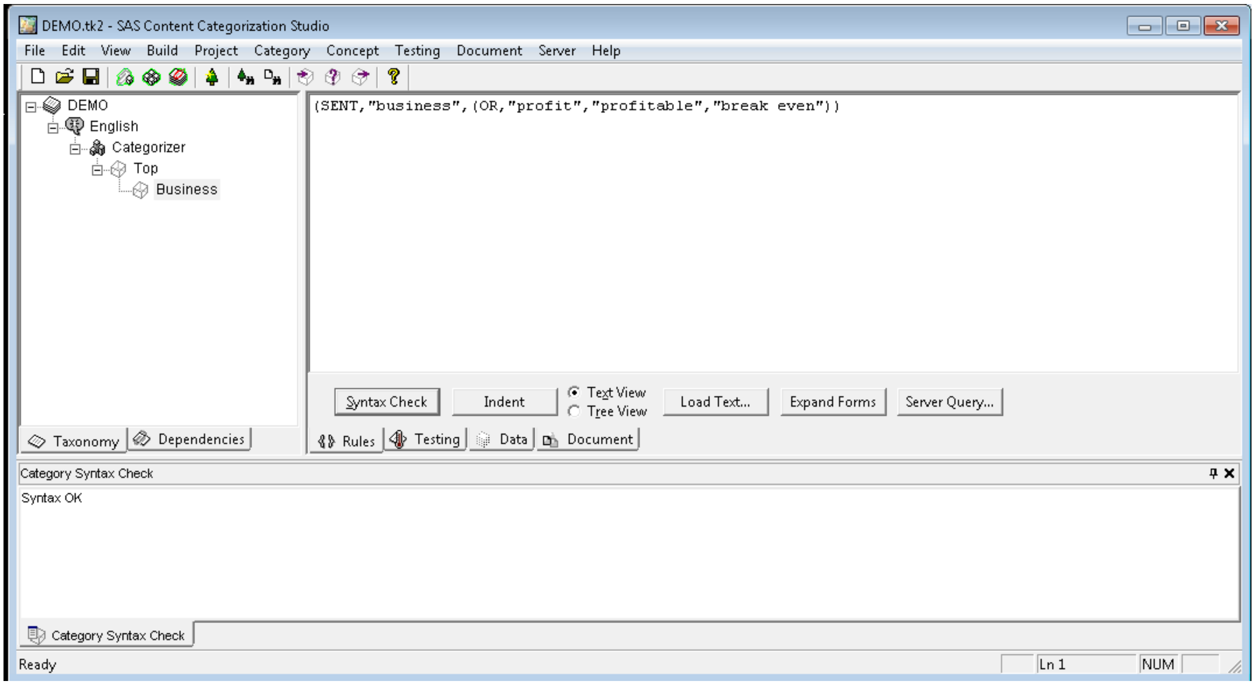

DEMO.tk2 - SAS Content Categorization Studio
File Edit View Build Project Category Concept Testing Document Server Help
DEMO
English
Categorizer
Top
Business
TEST
Taxonomy
Dependencies
Rules
Testing
Data
Document
Ready
Ln 1
NUM

● Test files for this category
○ Test all files everywhere

C:\Users\sas\Documents\TK240 Projects\Demo\test_docs\categories\Top\Business

Test File	Result	Relevancy
B1.txt		
B2-failing-doc.txt		

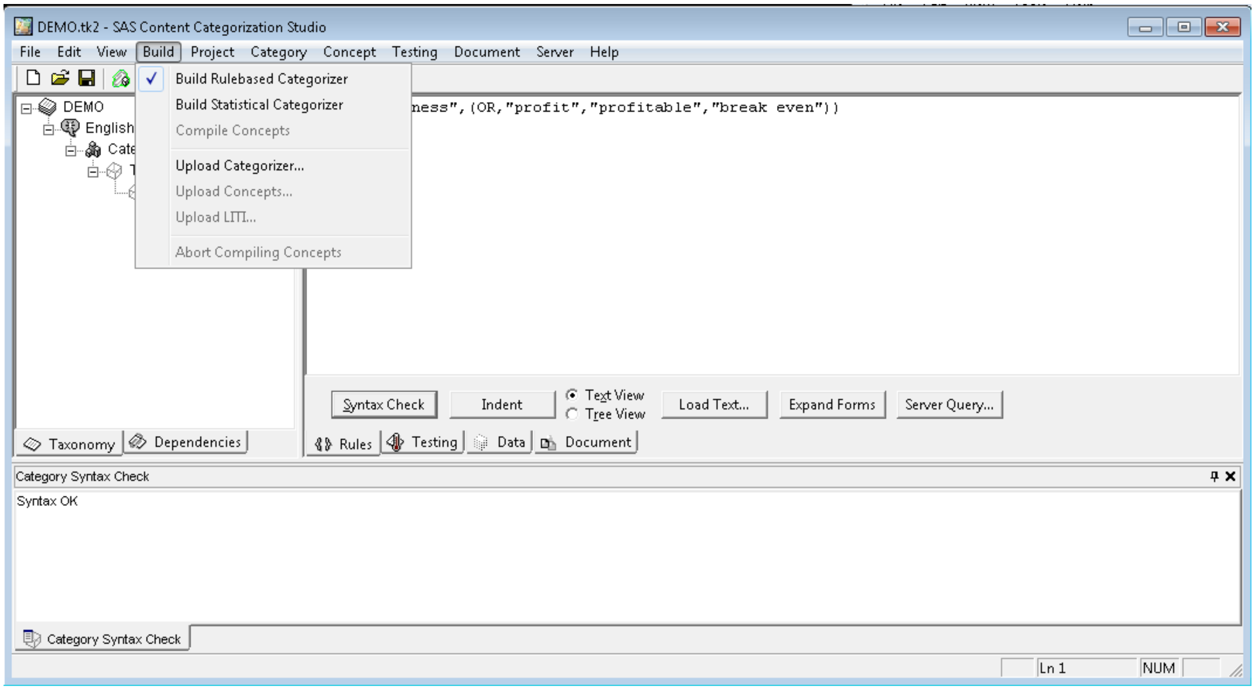
5. Now, build the category rules we created. To do that, check for the syntax of the rules using Syntax Check before building categories.



```
(SENT,"business",(OR,"profit","profitable","break even"))
```

```
Syntax OK
```

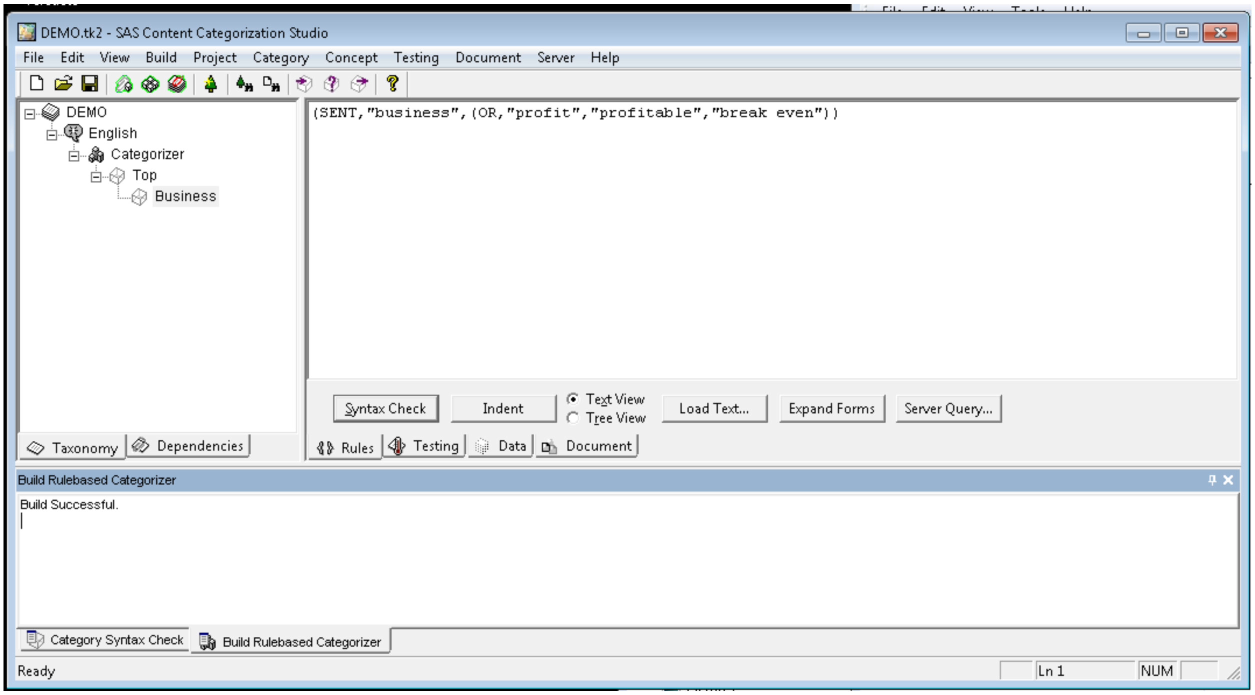
6. Above we created a boolean rule based category called Business, hence we will build the category using Build->Build Rulebased categorizer from the menu bar.



```
ness",(OR,"profit","profitable","break even"))
```

```
Syntax OK
```

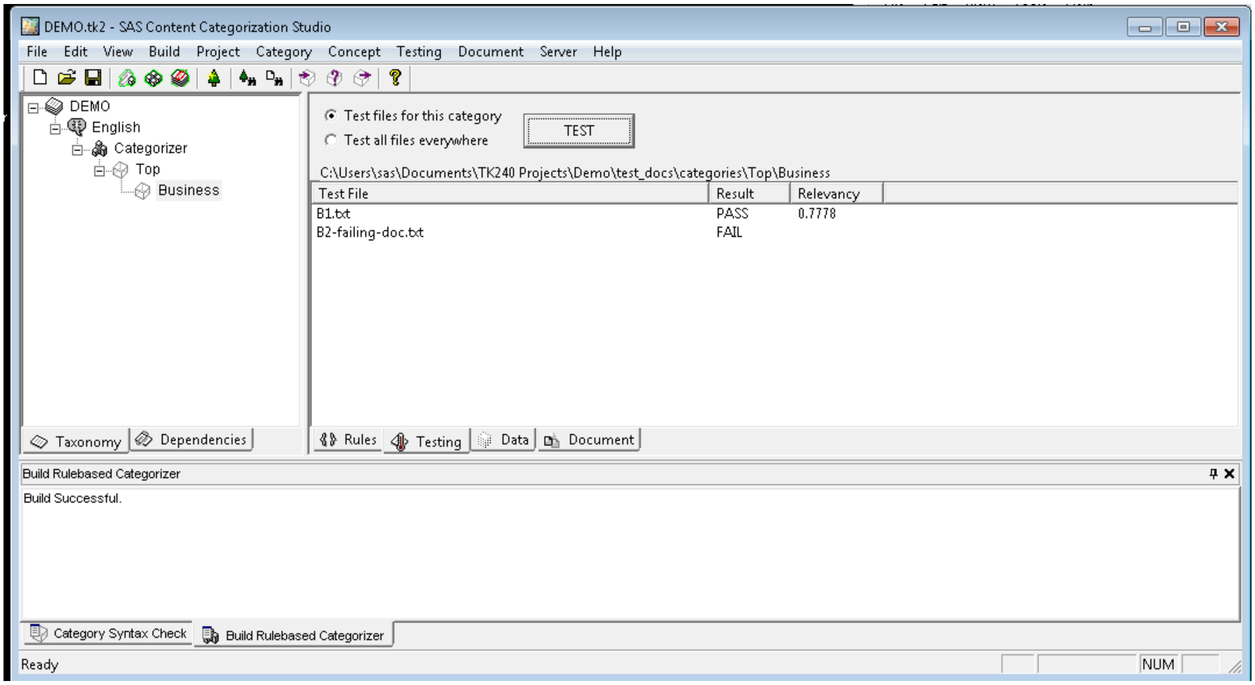
Build is successful.



```
(SENT,"business",(OR,"profit","profitable","break even"))
```

Build Successful.

7. Now to test the files go to the Testing tab and click the Test button.


DEMO.tk2 - SAS Content Categorization Studio
File Edit View Build Project Category Concept Testing Document Server Help
DEMO
English
Categorizer
Top
Business
● Test files for this category
○ Test all files everywhere
TEST
Taxonomy
Dependencies
Rules
Testing
Data
Document
Build Rulebased Categorizer
Category Syntax Check
Build Rulebased Categorizer
Ready
NUM

C:\Users\sas\Documents\TK240 Projects\Demo\test_docs\categories\Top\Business

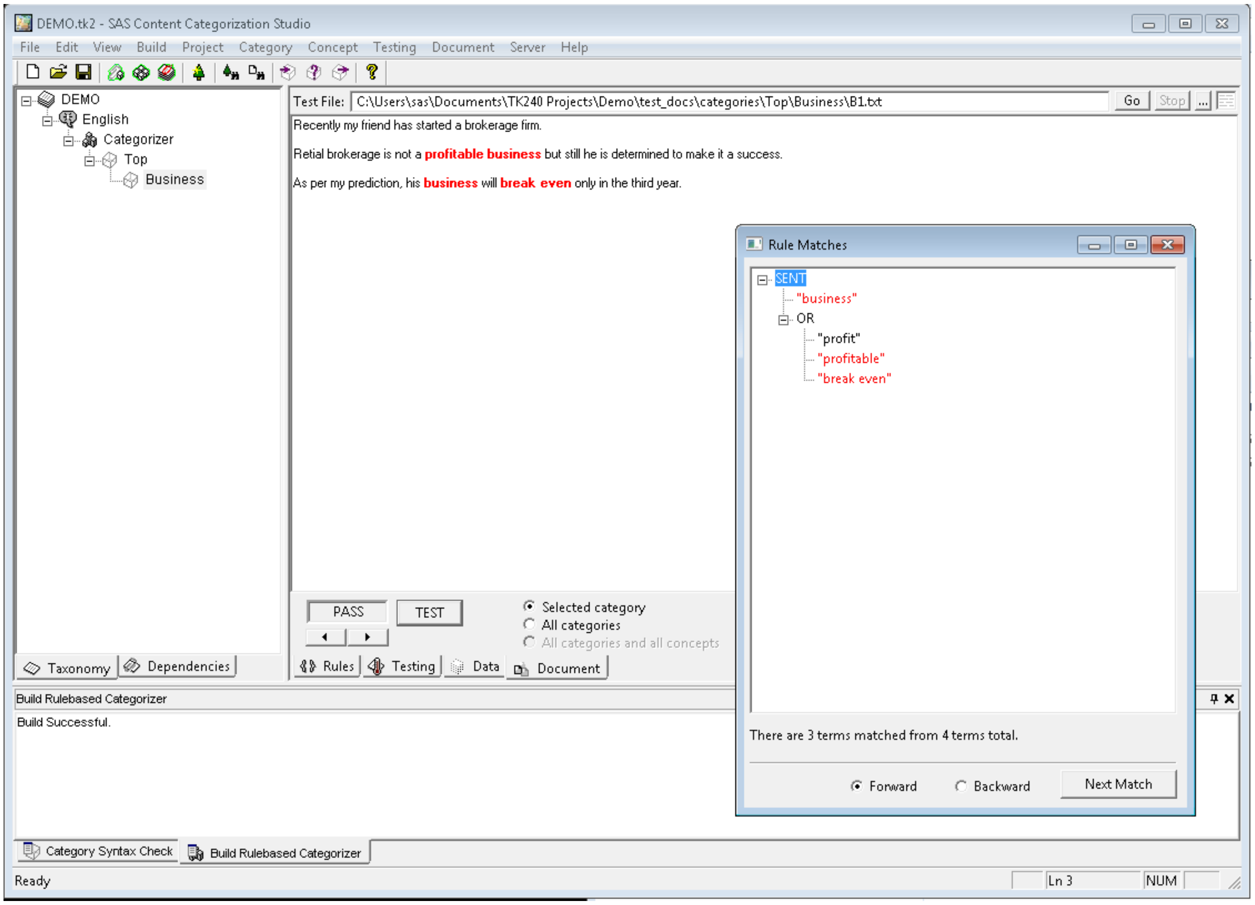
Test File	Result	Relevancy
B1.txt	PASS	0.7778
B2-failing-doc.txt	FAIL	

Build Successful.

We can double click any of the test documents shown above to see which words get matched or if the test documents did not match. Matched words will get highlighted in red.

Relevancy score shown above is useful to determine best match when the document matches more than one category.

8. Double click the B1.txt to open that file on Document tab and check which words get highlighted.



Test File: C:\Users\sas\Documents\TK240 Projects\Demo\test_docs\categories\Top\Business\B1.txt

Recently my friend has started a brokerage firm.

Retial brokerage is not a **profitable business** but still he is determined to make it a success.

As per my prediction, his **business** will **break even** only in the third year.

● Selected category
○ All categories
○ All categories and all concepts

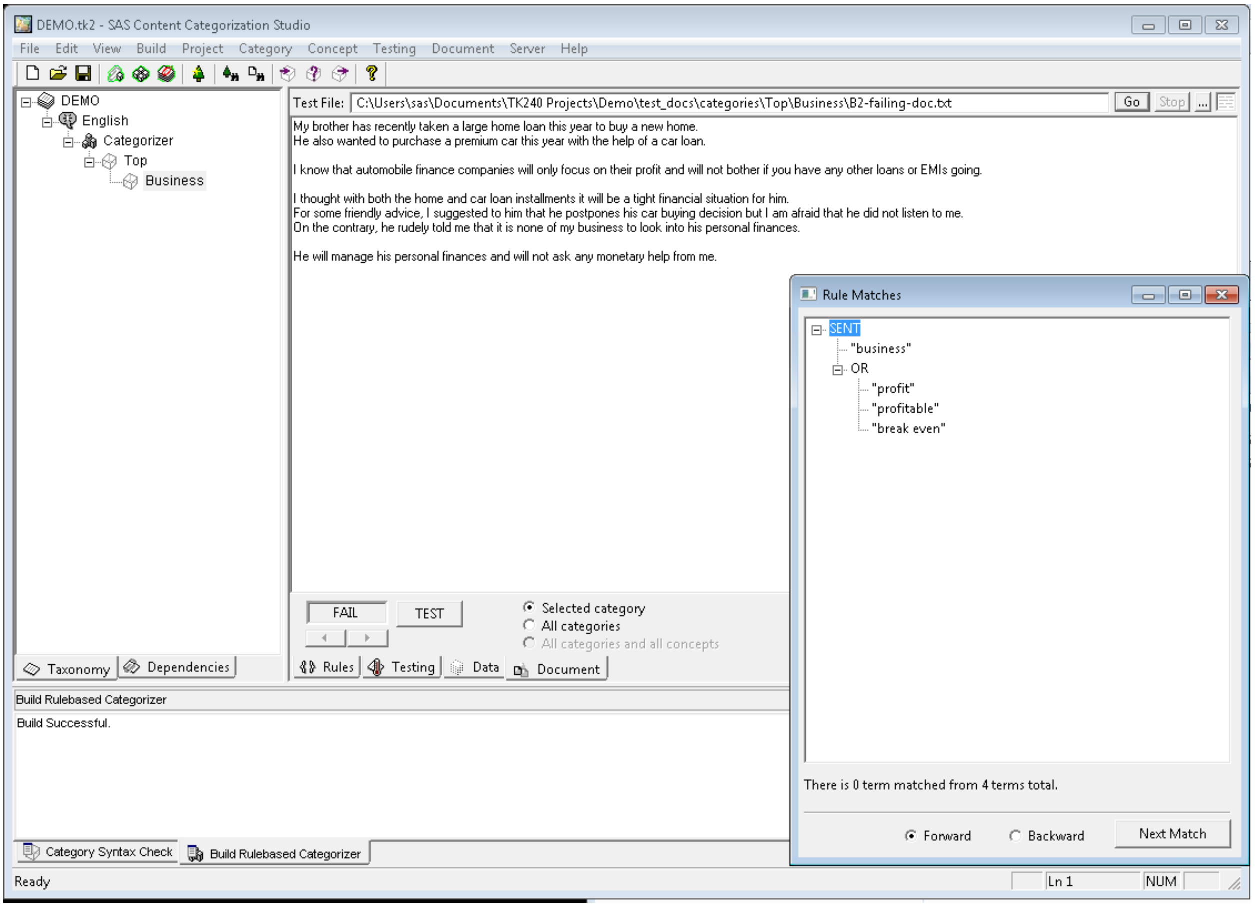
```
SENT
  "business"
  OR
    "profit"
    "profitable"
    "break even"
```

There are 3 terms matched from 4 terms total.

Build Rulebased Categorizer

Build Successful.

The file *B2-failing-doc.txt* fails because the terms profit and business do not match the rule defined.



Test File: C:\Users\sas\Documents\TK240 Projects\Demo\test_docs\categories\Top\Business\B2-failing-doc.txt

```
My brother has recently taken a large home loan this year to buy a new home.
He also wanted to purchase a premium car this year with the help of a car loan.

I know that automobile finance companies will only focus on their profit and will not bother if you have any other loans or EMIs going.

I thought with both the home and car loan installments it will be a tight financial situation for him.
For some friendly advice, I suggested to him that he postpones his car buying decision but I am afraid that he did not listen to me.
On the contrary, he rudely told me that it is none of my business to look into his personal finances.

He will manage his personal finances and will not ask any monetary help from me.
```

● Selected category
○ All categories
○ All categories and all concepts

```
SENT
  "business"
  OR
    "profit"
    "profitable"
    "break even"
```

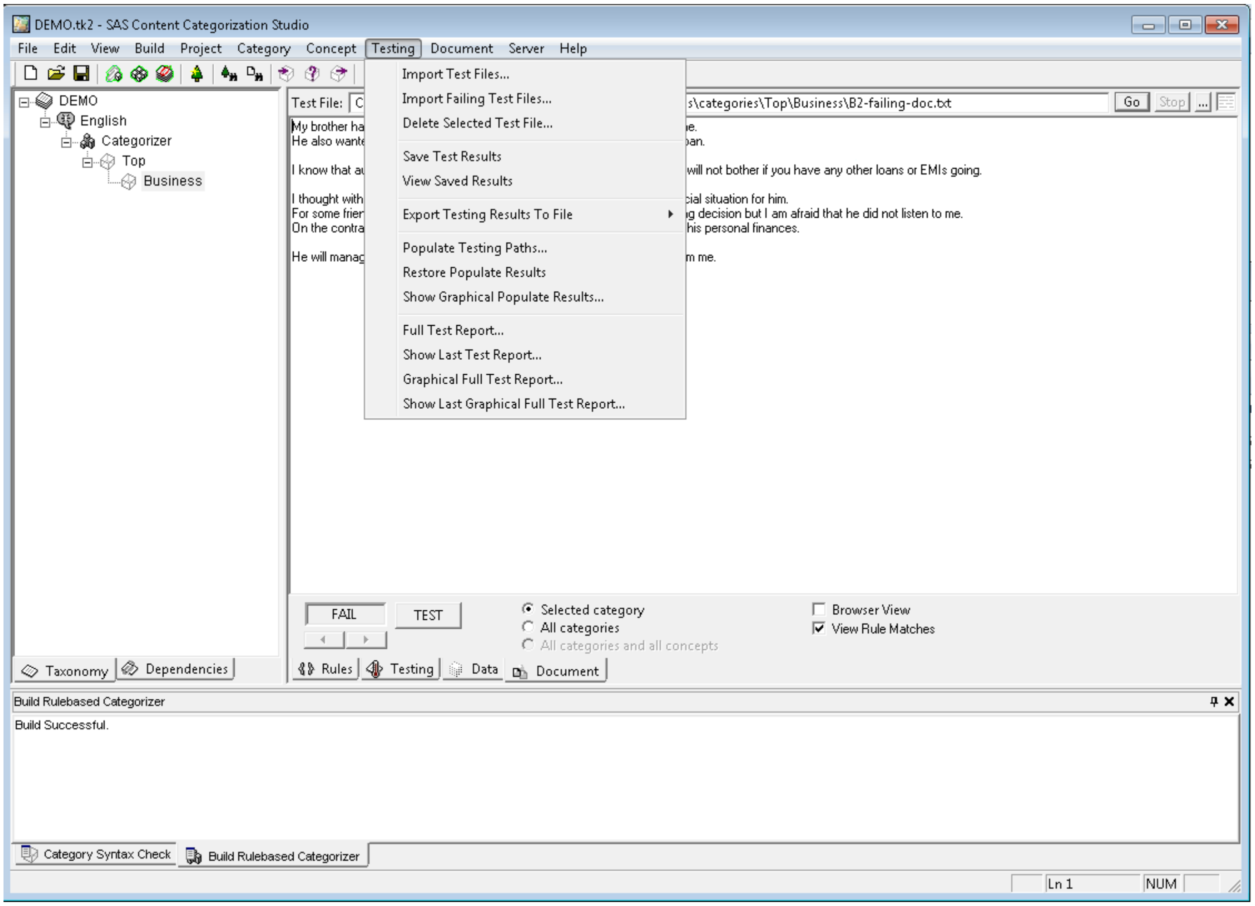
There is 0 term matched from 4 terms total.

● Forward ○ Backward

Build Rulebased Categorizer

```
Build Successful.
```

9. A good model should match all the relevant test documents (recall) while not matching irrelevant documents (precision). To test the failure of such irrelevant documents we can also check using Testing->Import Failing test Files….



Test File: C ... s\categories\Top\Business\B2-failing-doc.txt

My brother ha ... e.
He also wante ... oan.

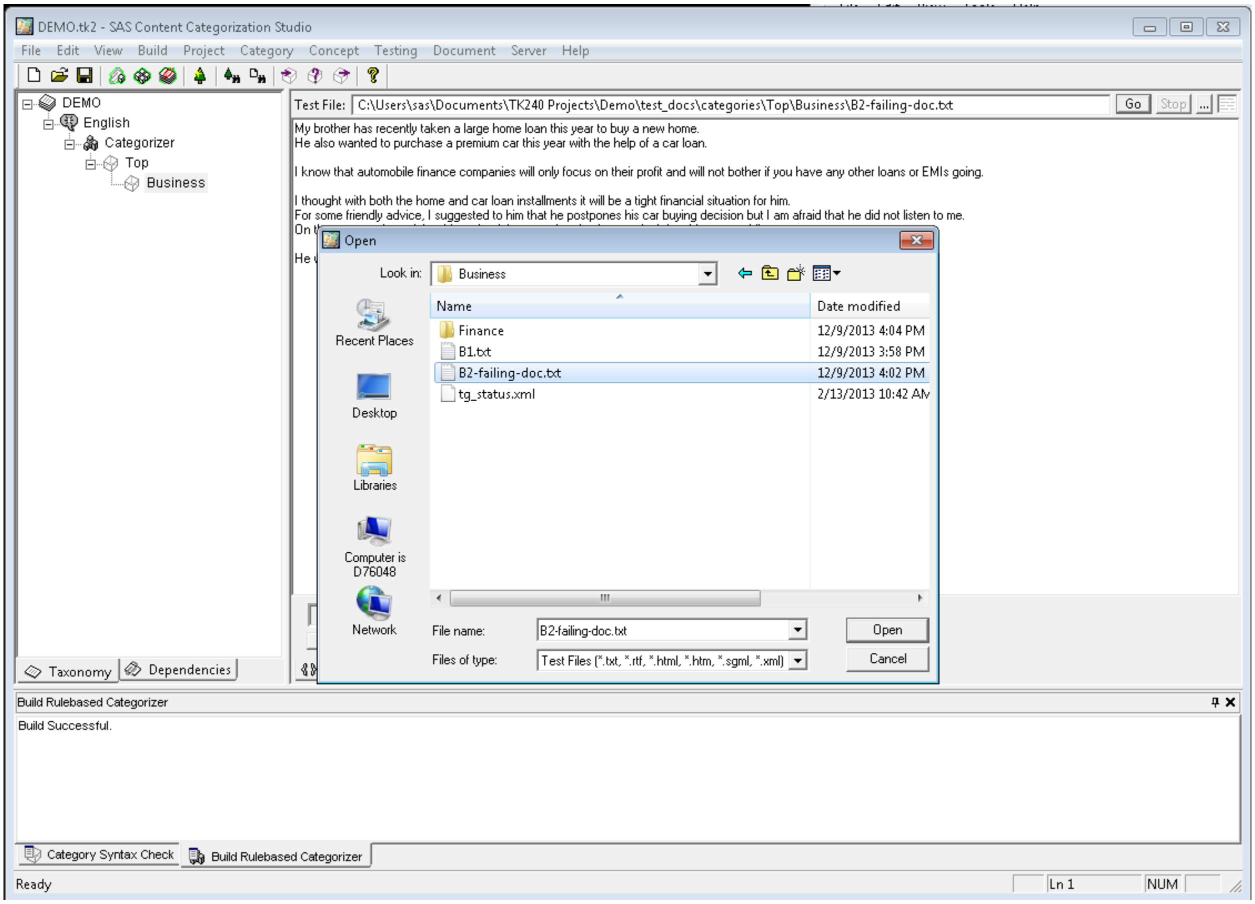
I know that a ... will not bother if you have any other loans or EMIs going.

I thought with ... cial situation for him.
For some frier ... g decision but I am afraid that he did not listen to me.
On the contra ... his personal finances.

He will manag ... m me.

Build Successful.

10. Open the Failing doc to be tested



Test File: C:\Users\sas\Documents\TK240 Projects\Demo\test_docs\categories\Top\Business\B2-failing-doc.txt

My brother has recently taken a large home loan this year to buy a new home.
He also wanted to purchase a premium car this year with the help of a car loan.

I know that automobile finance companies will only focus on their profit and will not bother if you have any other loans or EMIs going.

I thought with both the home and car loan installments it will be a tight financial situation for him.
For some friendly advice, I suggested to him that he postpones his car buying decision but I am afraid that he did not listen to me.
On t

He

Look in: Business

Name	Date modified
Finance	12/9/2013 4:04 PM
B1.txt	12/9/2013 3:58 PM
B2-failing-doc.txt	12/9/2013 4:02 PM
tg_status.xml	2/13/2013 10:42 AM

File name: B2-failing-doc.txt

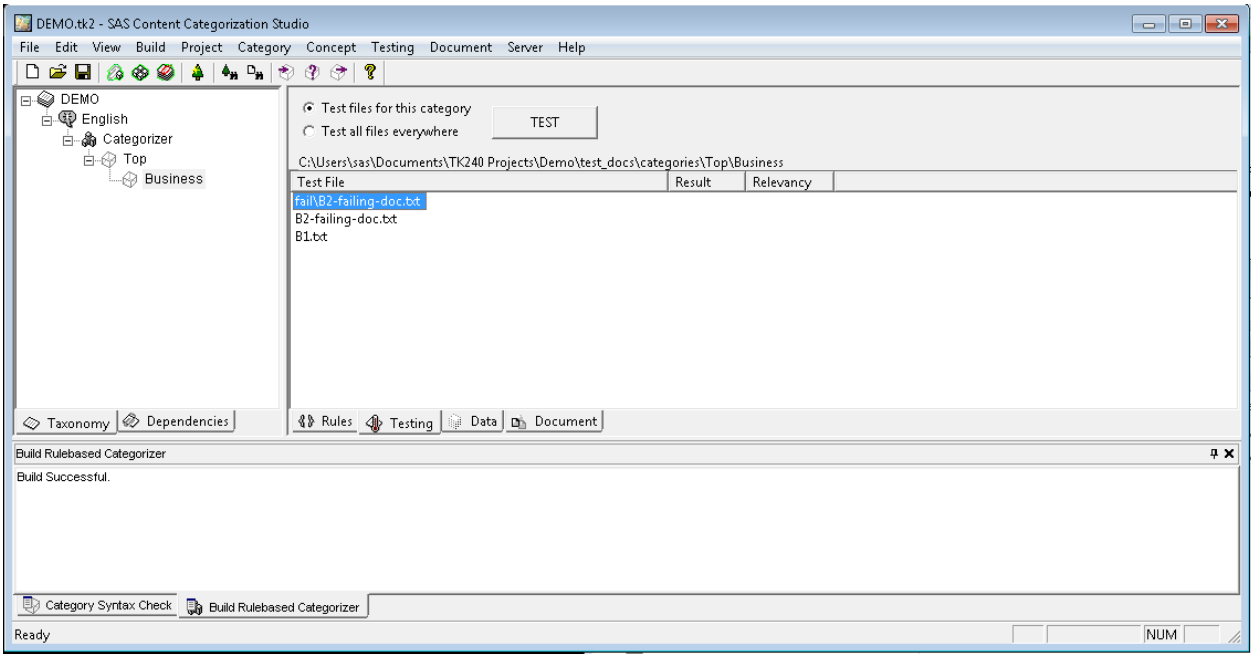
Files of type: Test Files (*.txt, *.rtf, *.html, *.htm, *.sgml, *.xml)

Build Rulebased Categorizer

Build Successful.

Ready

The imported file will be visible in the Testing tab.


DEMO.tk2 - SAS Content Categorization Studio
File Edit View Build Project Category Concept Testing Document Server Help
DEMO
English
Categorizer
Top
Business
TEST
Taxonomy
Dependencies
Rules
Testing
Data
Document
Category Syntax Check
Build Rulebased Categorizer
NUM

● Test files for this category
○ Test all files everywhere

C:\Users\sas\Documents\TK240 Projects\Demo\test_docs\categories\Top\Business

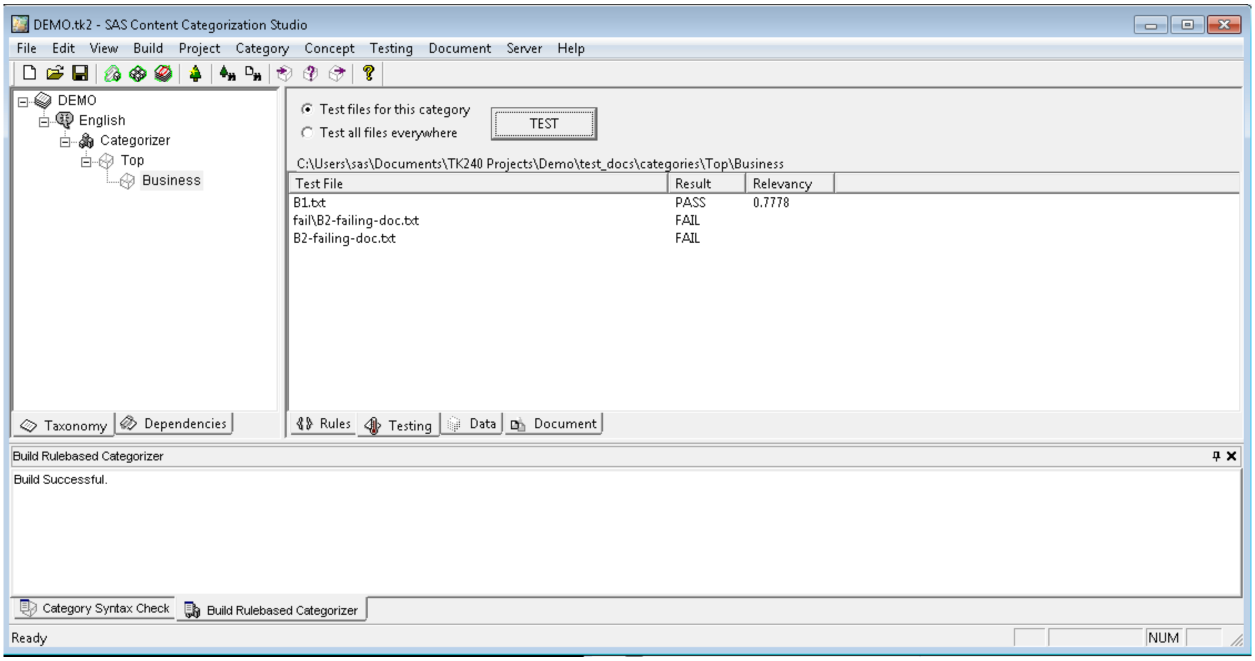
Test File	Result	Relevancy
fail\B2-failing-doc.txt		
B2-failing-doc.txt		
B1.txt		

Build Rulebased Categorizer

Build Successful.

Ready

In the Test the imported failing test file should FAIL, that proves our rule is precise (filtering irrelevant documents) and working correctly.


DEMO.tk2 - SAS Content Categorization Studio
File Edit View Build Project Category Concept Testing Document Server Help
DEMO
English
Categorizer
Top
Business
TEST
Taxonomy
Dependencies
Rules
Testing
Data
Document
Category Syntax Check
Build Rulebased Categorizer
Ready
NUM

● Test files for this category
○ Test all files everywhere

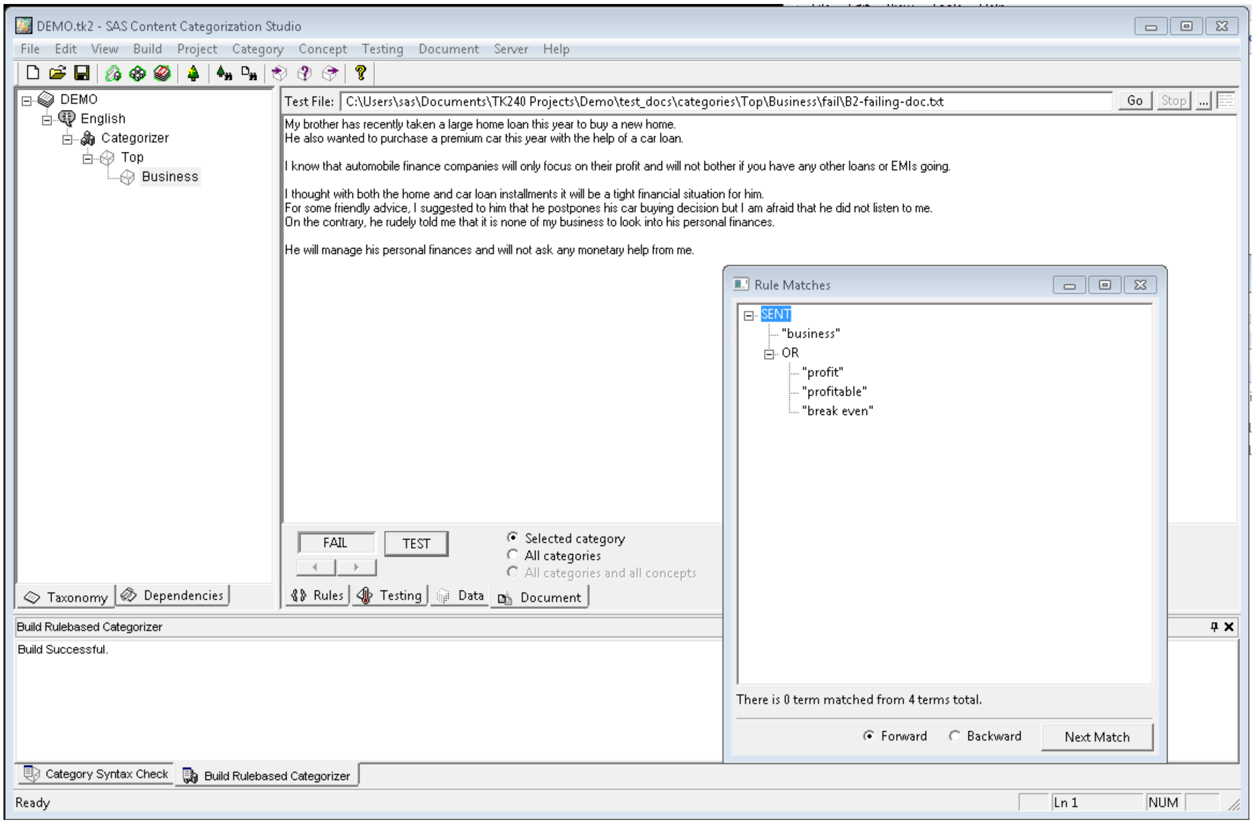
C:\Users\sas\Documents\TK240 Projects\Demo\test_docs\categories\Top\Business

Test File	Result	Relevancy
B1.txt	PASS	0.7778
fail\B2-failing-doc.txt	FAIL	
B2-failing-doc.txt	FAIL	

Build Rulebased Categorizer

Build Successful.

11. Double clicking the Failing document opens it in the Document Tab.



Test File: C:\Users\sas\Documents\TK240 Projects\Demo\test_docs\categories\Top\Business\fail\B2-failing-doc.txt

My brother has recently taken a large home loan this year to buy a new home.
He also wanted to purchase a premium car this year with the help of a car loan.

I know that automobile finance companies will only focus on their profit and will not bother if you have any other loans or EMIs going.

I thought with both the home and car loan installments it will be a tight financial situation for him.
For some friendly advice, I suggested to him that he postpones his car buying decision but I am afraid that he did not listen to me.
On the contrary, he rudely told me that it is none of my business to look into his personal finances.

He will manage his personal finances and will not ask any monetary help from me.

● Selected category
○ All categories
○ All categories and all concepts

```
SENT
  "business"
  OR
    "profit"
    "profitable"
    "break even"
```

There is 0 term matched from 4 terms total.

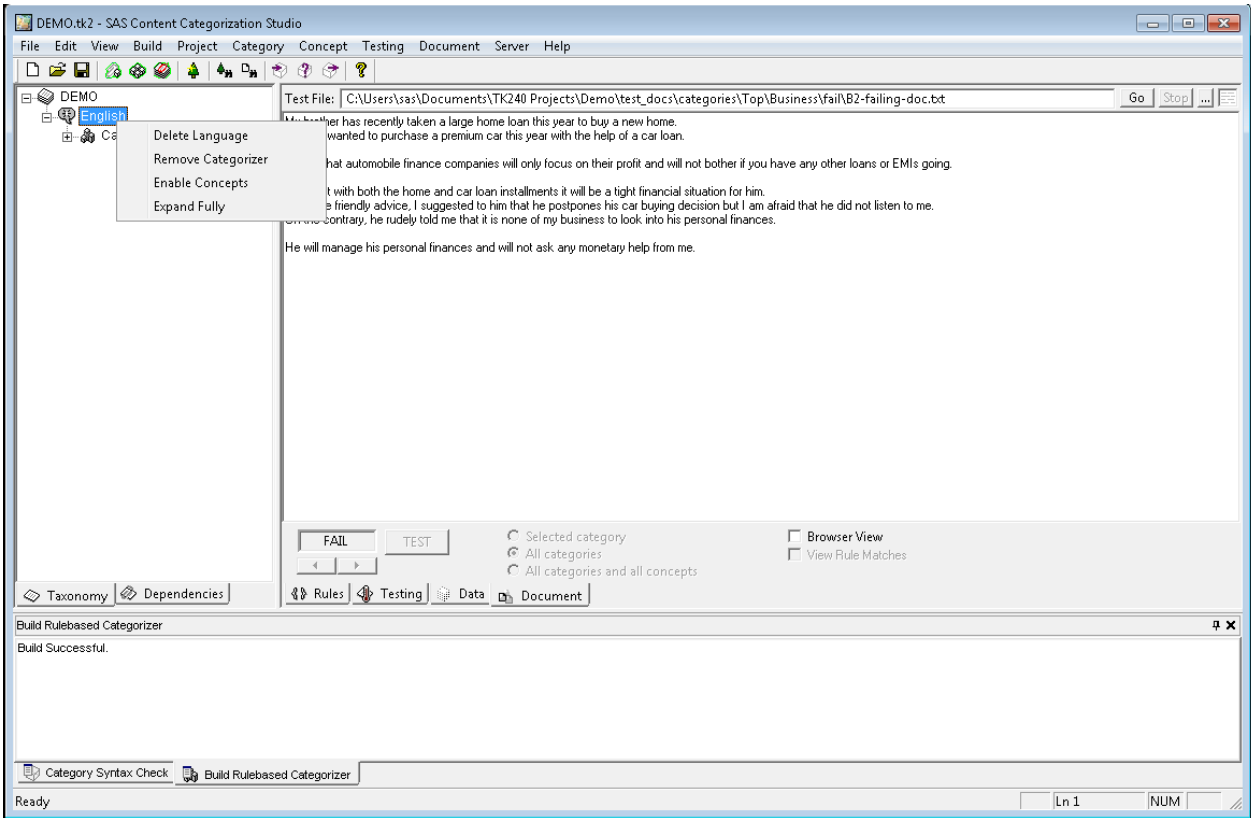
● Forward ○ Backward

Build Rulebased Categorizer

Build Successful.

Part 3: Creating LITI Definition Concepts

1. Right mouse click on English and then click Enable Concepts to create Concept Taxonomy



Test File: C:\Users\sas\Documents\TK240 Projects\Demo\test_docs\categories\Top\Business\fail\B2-failing-doc.txt

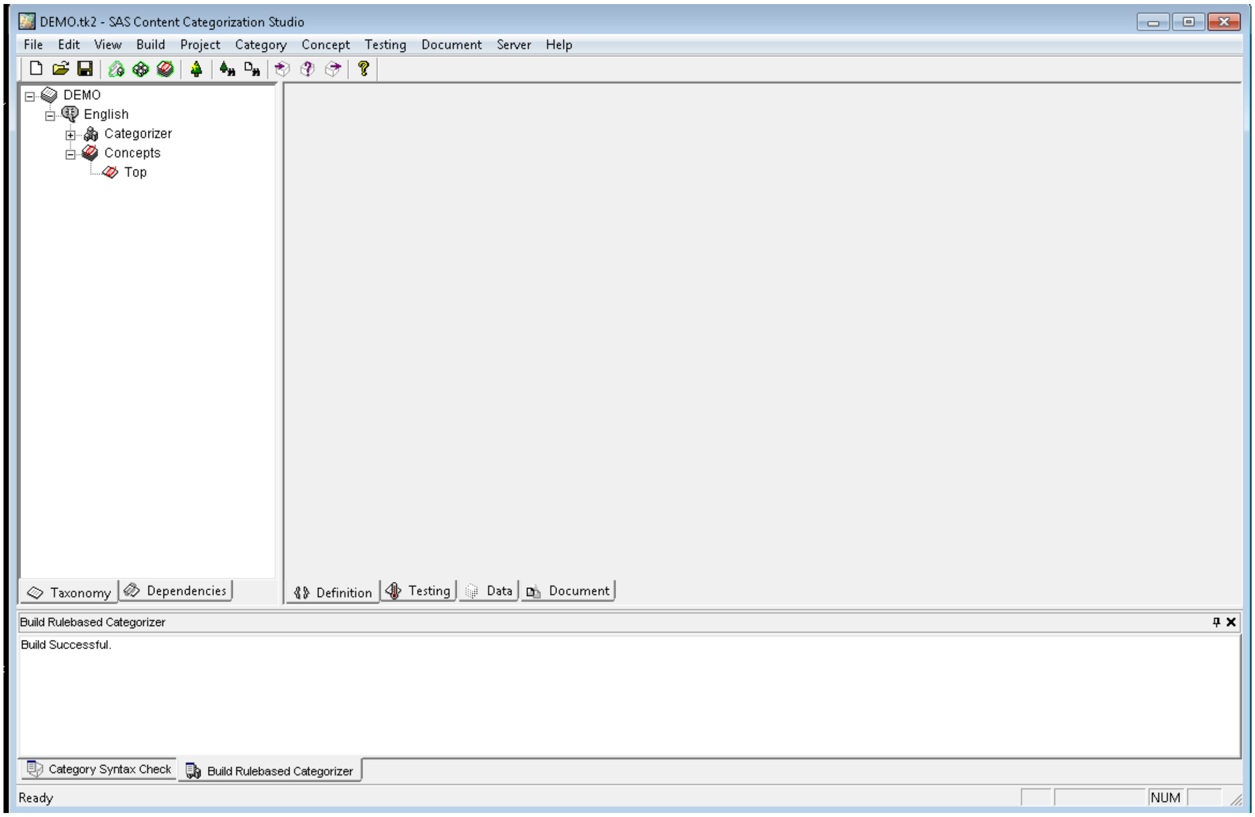
My brother has recently taken a large home loan this year to buy a new home.
wanted to purchase a premium car this year with the help of a car loan.

hat automobile finance companies will only focus on their profit and will not bother if you have any other loans or EMIs going.

t with both the home and car loan installments it will be a tight financial situation for him.
e friendly advice, I suggested to him that he postpones his car buying decision but I am afraid that he did not listen to me.
On the contrary, he rudely told me that it is none of my business to look into his personal finances.

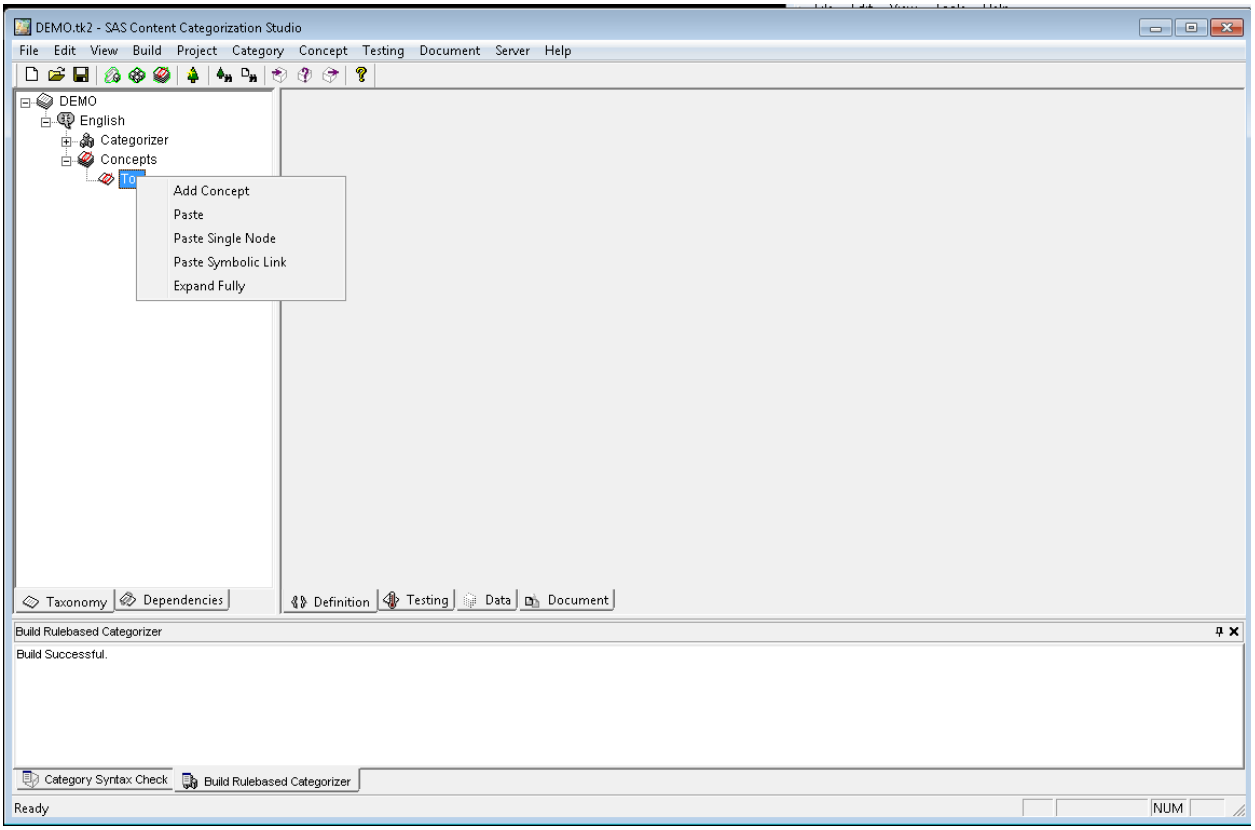
He will manage his personal finances and will not ask any monetary help from me.

Build Successful.


DEMO.tk2 - SAS Content Categorization Studio
File Edit View Build Project Category Concept Testing Document Server Help
DEMO
English
Categorizer
Concepts
Top
Taxonomy
Dependencies
Definition
Testing
Data
Document
Build Rulebased Categorizer
Category Syntax Check
Build Rulebased Categorizer
Ready
NUM

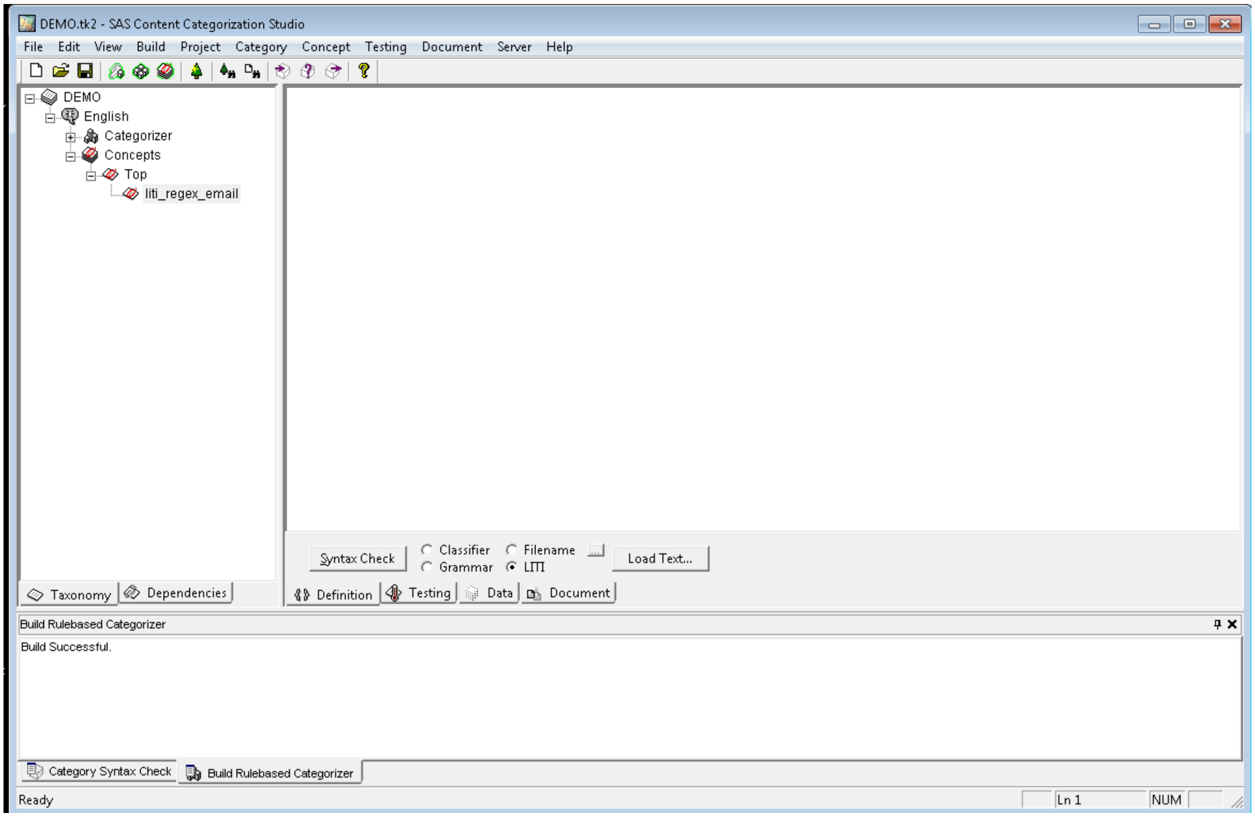
Build Successful.

2. To add new Concept do Right Click on Top and select Add Concept



Build Successful.

3. Create a concept definition named liti_regex_email. In the below screenshot the LITI concept uses Regular Expressions to capture the email address ending with .com. To create the LITI concept, select the LITI radio button on the Definition tab.



○ Classifier ○ Filename
○ Grammar ● LITI

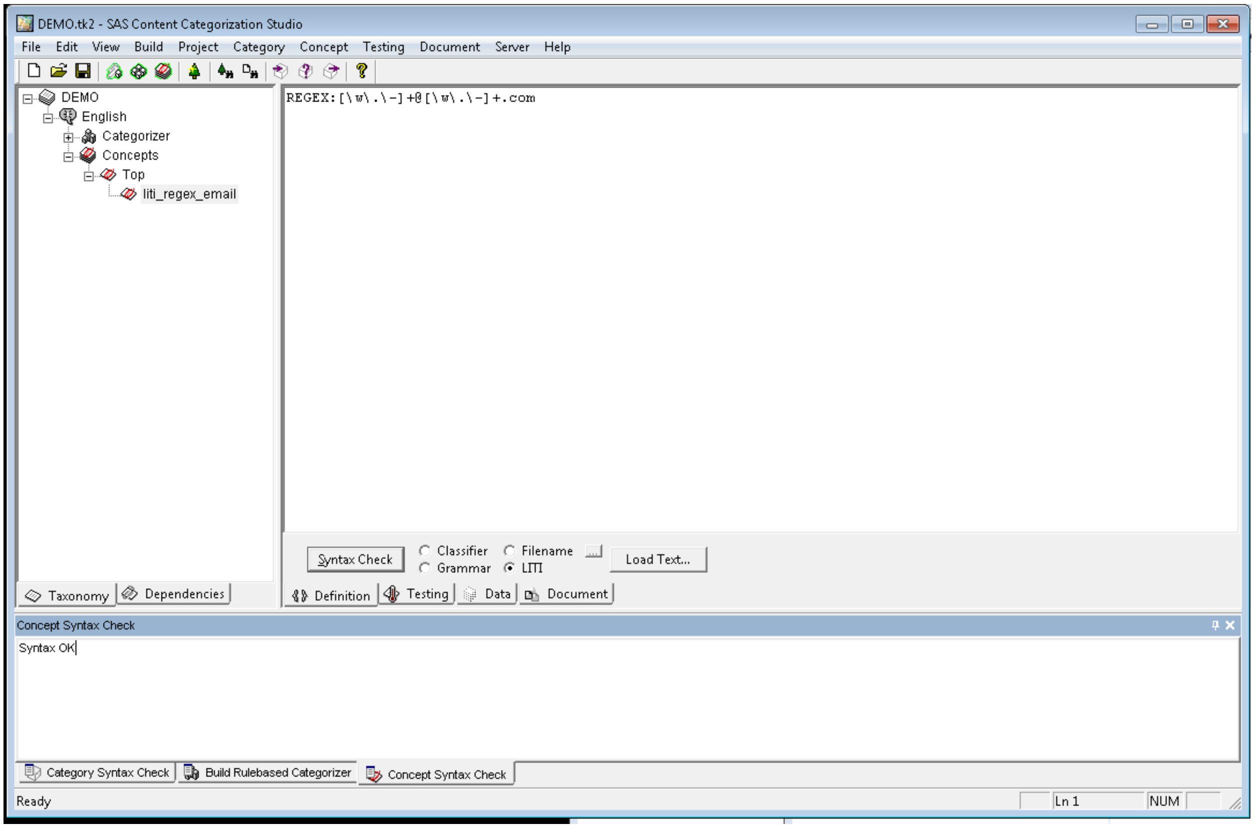
Build Rulebased Categorizer

```
Build Successful.
```

Write the definition

```
REGEX:[\w\.\-]+@[\w\.\-]+.com
```

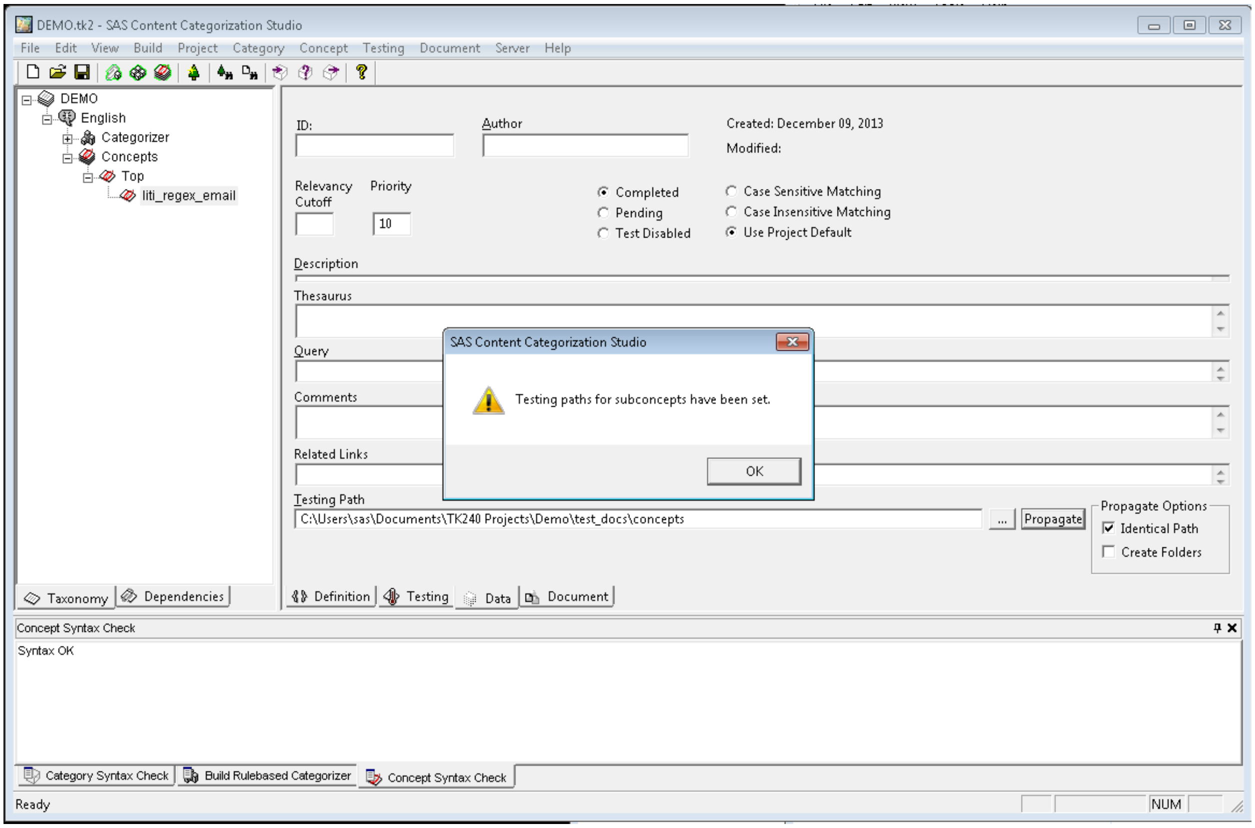
on the Definition tab and then check for the syntax by clicking Syntax Check button.



```
REGEX:[\w\.\-]+@[\w\.\-]+.com
```

```
Syntax OK
```

4. On the Data tab set the path of the folder where the files to be tested are located and then click Propagate.


DEMO.tk2 - SAS Content Categorization Studio
File Edit View Build Project Category Concept Testing Document Server Help
DEMO
English
Categorizer
Concepts
Top
liti_regex_email
Taxonomy
Dependencies
Definition
Testing
Data
Document
Category Syntax Check
Build Rulebased Categorizer
Concept Syntax Check
NUM

ID:

Author

Created: December 09, 2013

Modified:

Relevancy Cutoff

Priority 10

● Completed
○ Pending
○ Test Disabled

○ Case Sensitive Matching
○ Case Insensitive Matching
● Use Project Default

Description

Thesaurus

Query

Comments

Related Links

Testing Path C:\Users\sas\Documents\TK240 Projects\Demo\test_docs\concepts ... Propagate

Propagate Options
☑ Identical Path
□ Create Folders

SAS Content Categorization Studio

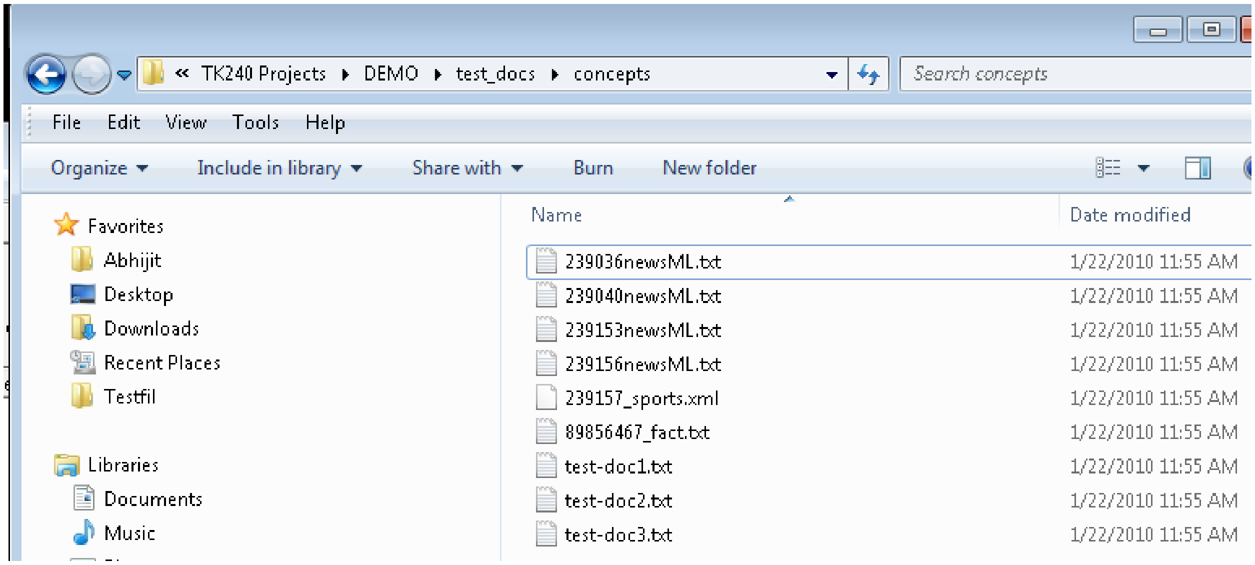
Testing paths for subconcepts have been set.

OK

Concept Syntax Check

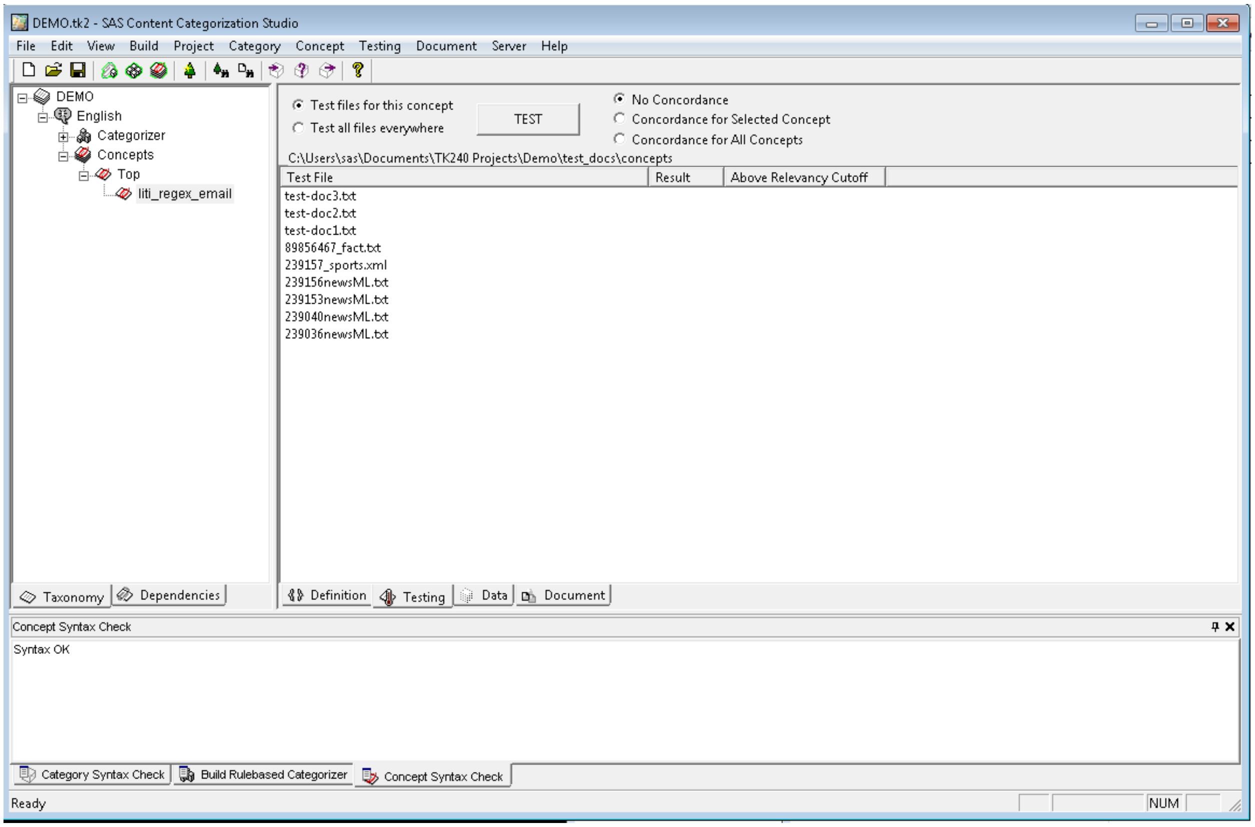
Syntax OK

Ready


« TK240 Projects ▸ DEMO ▸ test_docs ▸ concepts
Search concepts
File Edit View Tools Help
Organize ▾ Include in library ▾ Share with ▾ Burn New folder
Favorites
Abhijit
Desktop
Downloads
Recent Places
Testfil
Libraries
Documents
Music

Name	Date modified
239036newsML.txt	1/22/2010 11:55 AM
239040newsML.txt	1/22/2010 11:55 AM
239153newsML.txt	1/22/2010 11:55 AM
239156newsML.txt	1/22/2010 11:55 AM
239157_sports.xml	1/22/2010 11:55 AM
89856467_fact.txt	1/22/2010 11:55 AM
test-doc1.txt	1/22/2010 11:55 AM
test-doc2.txt	1/22/2010 11:55 AM
test-doc3.txt	1/22/2010 11:55 AM

Files in the Testing Folder will now be visible in the Testing Tab.



● Test files for this concept
○ Test all files everywhere

● No Concordance
○ Concordance for Selected Concept
○ Concordance for All Concepts

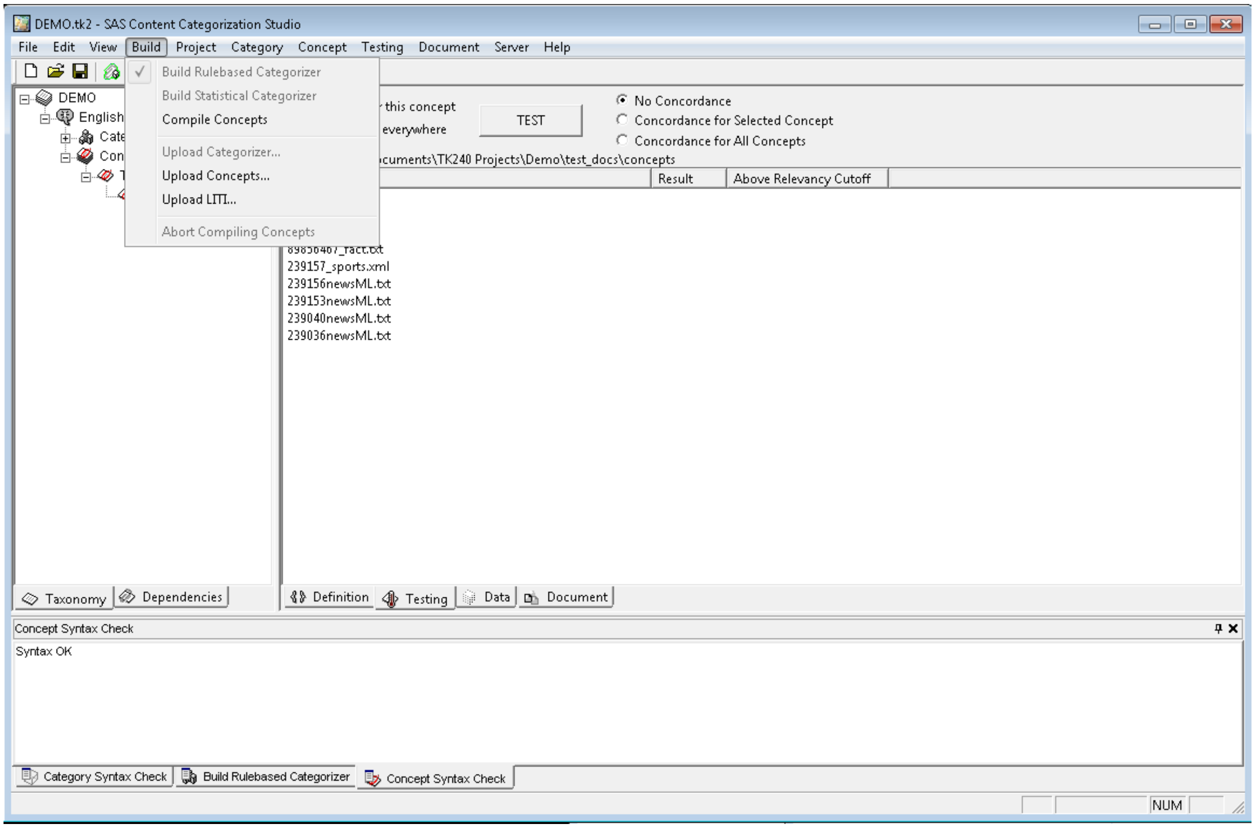
C:\Users\sas\Documents\TK240 Projects\Demo\test_docs\concepts

Test File	Result	Above Relevancy Cutoff
test-doc3.txt		
test-doc2.txt		
test-doc1.txt		
89856467_fact.txt		
239157_sports.xml		
239156newsML.txt		
239153newsML.txt		
239040newsML.txt		
239036newsML.txt		

Concept Syntax Check

```
Syntax OK
```

5. Compile the concepts by clicking Build->Compile Concepts on the menu bar



● No Concordance

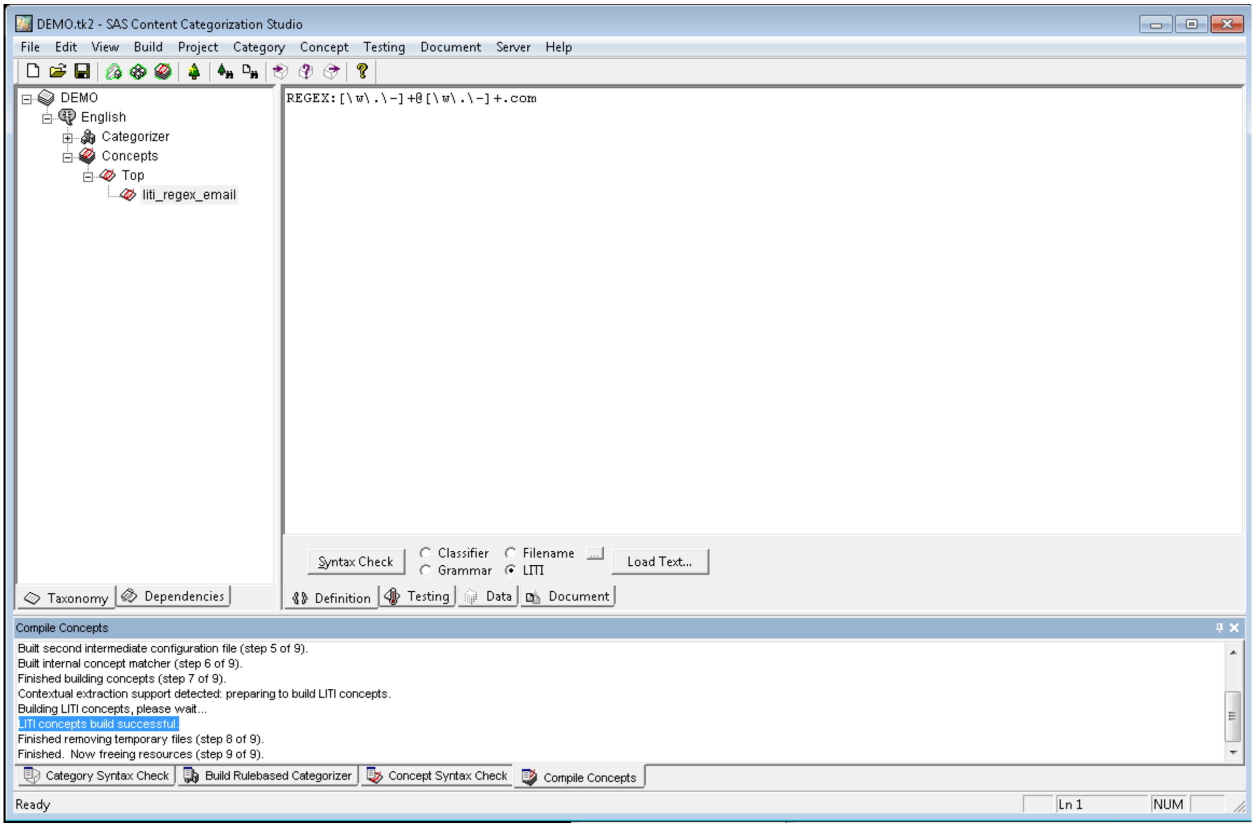
cuments\TK240 Projects\Demo\test_docs\concepts

```
89856467_fact.txt
239157_sports.xml
239156newsML.txt
239153newsML.txt
239040newsML.txt
239036newsML.txt
```

Concept Syntax Check

```
Syntax OK
```

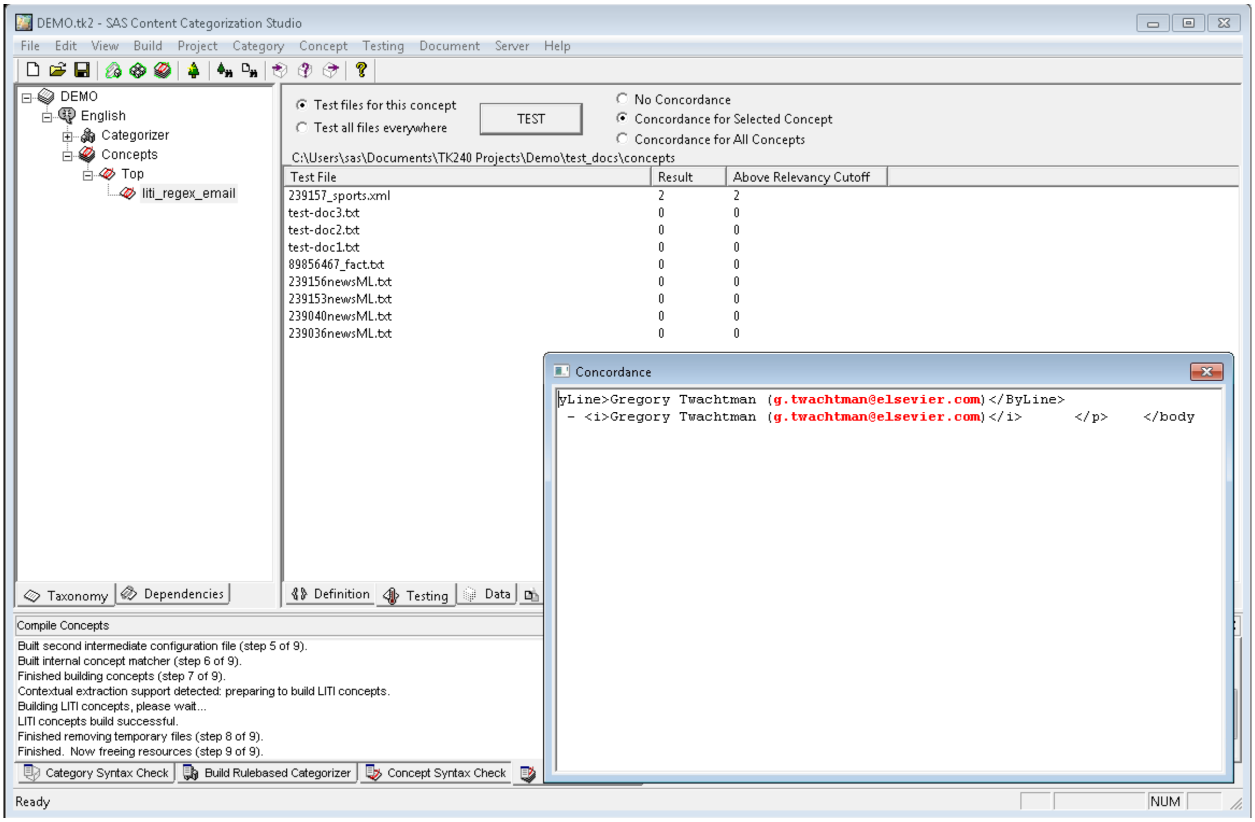
LITI concept build is successful.



```
REGEX:[\w\.\-]+@[\w\.\-]+.com
```

```
Built second intermediate configuration file (step 5 of 9).
Built internal concept matcher (step 6 of 9).
Finished building concepts (step 7 of 9).
Contextual extraction support detected: preparing to build LITI concepts.
Building LITI concepts, please wait...
LITI concepts build successful.
Finished removing temporary files (step 8 of 9).
Finished.  Now freeing resources (step 9 of 9).
```

6. Now test the files for this concept by clicking the Test button on the Testing tab. With Concordance selected, we can control the number of characters, words or sentences displayed before and after each match.


DEMO.tk2 - SAS Content Categorization Studio
File Edit View Build Project Category Concept Testing Document Server Help
DEMO
English
Categorizer
Concepts
Top
liti_regex_email
TEST
Taxonomy
Dependencies
Definition
Testing
Data
Category Syntax Check
Build Rulebased Categorizer
Concept Syntax Check
Ready
NUM

● Test files for this concept
○ Test all files everywhere

○ No Concordance
● Concordance for Selected Concept
○ Concordance for All Concepts

C:\Users\sas\Documents\TK240 Projects\Demo\test_docs\concepts

Test File	Result	Above Relevancy Cutoff
239157_sports.xml	2	2
test-doc3.txt	0	0
test-doc2.txt	0	0
test-doc1.txt	0	0
89856467_fact.txt	0	0
239156newsML.txt	0	0
239153newsML.txt	0	0
239040newsML.txt	0	0
239036newsML.txt	0	0

Concordance

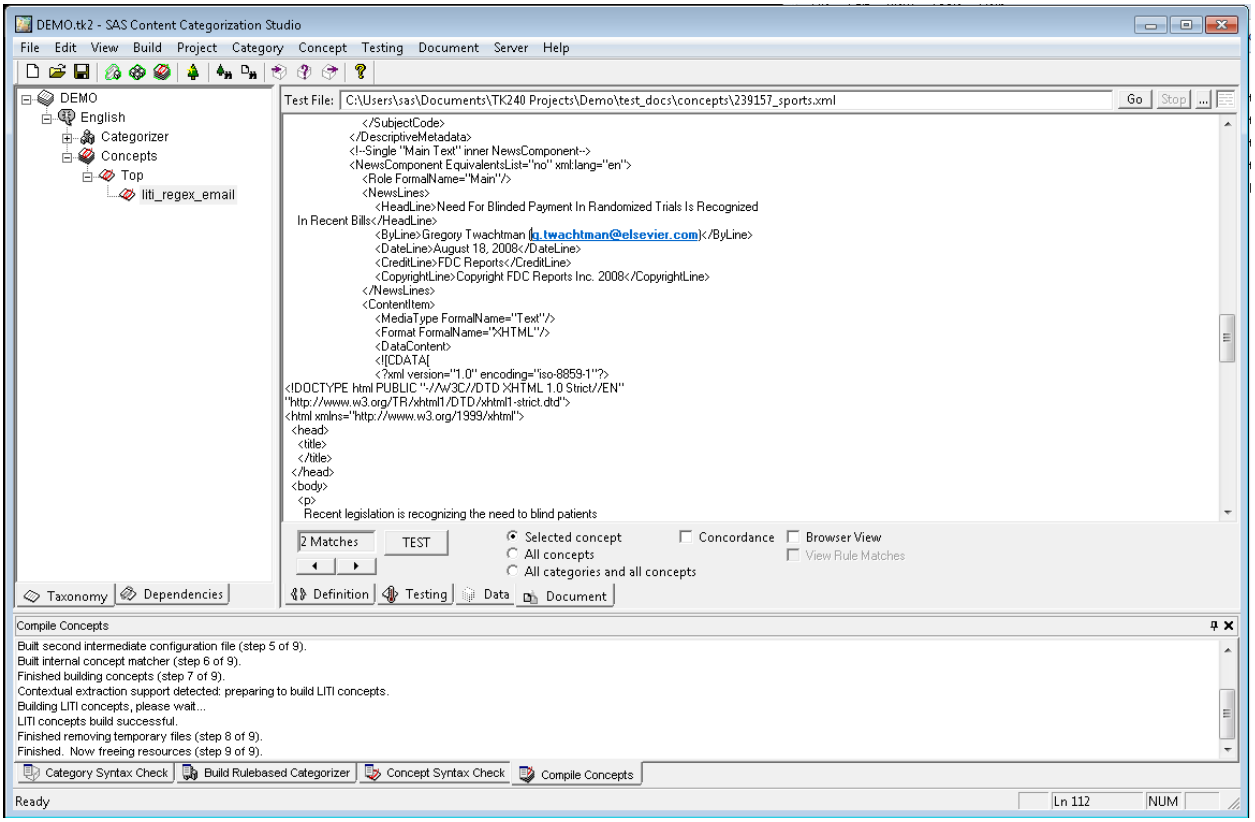
```
yLine>Gregory Twachtman (g.twachtman@elsevier.com)</ByLine>
 - <i>Gregory Twachtman (g.twachtman@elsevier.com)</i>      </p>     </body
```

Compile Concepts

```
Built second intermediate configuration file (step 5 of 9).
Built internal concept matcher (step 6 of 9).
Finished building concepts (step 7 of 9).
Contextual extraction support detected: preparing to build LITI concepts.
Building LITI concepts, please wait...
LITI concepts build successful.
Finished removing temporary files (step 8 of 9).
Finished.  Now freeing resources (step 9 of 9).
```

7. We can double click any of the test documents shown above to see which strings get matched. Matched strings will get highlighted in red as shown below.

Double click the 239157_sports.xml to open the file on Document tab and check which words get highlighted.



Test File: C:\Users\sas\Documents\TK240 Projects\Demo\test_docs\concepts\239157_sports.xml

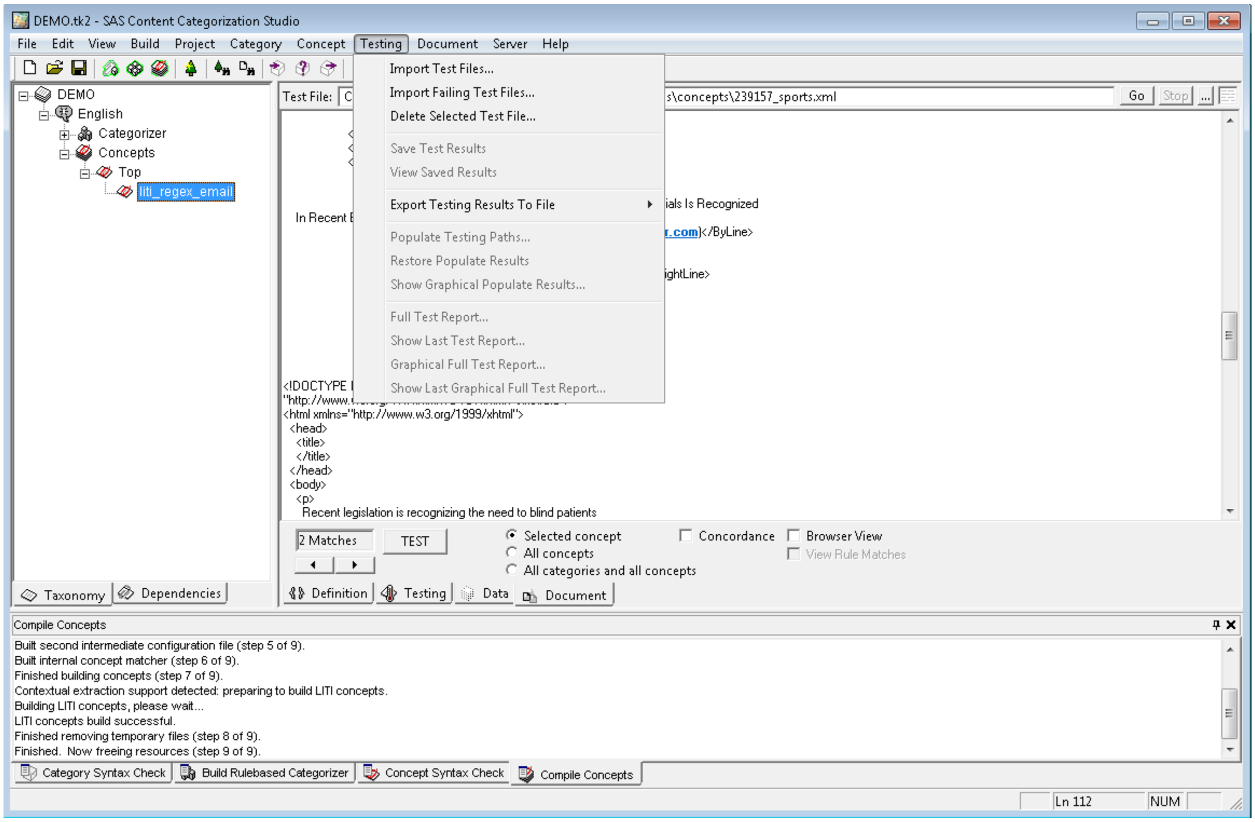
```
                </SubjectCode>
            </DescriptiveMetadata>
            <!--Single "Main Text" inner NewsComponent-->
            <NewsComponent EquivalentsList="no" xml:lang="en">
                <Role FormalName="Main"/>
                <NewsLines>
                    <HeadLine>Need For Blinded Payment In Randomized Trials Is Recognized
   In Recent Bills</HeadLine>
                    <ByLine>Gregory Twachtman (g.twachtman@elsevier.com)</ByLine>
                    <DateLine>August 18, 2008</DateLine>
                    <CreditLine>FDC Reports</CreditLine>
                    <CopyrightLine>Copyright FDC Reports Inc. 2008</CopyrightLine>
                </NewsLines>
                <ContentItem>
                    <MediaType FormalName="Text"/>
                    <Format FormalName="XHTML"/>
                    <DataContent>
                    <![CDATA[
                    <?xml version="1.0" encoding="iso-8859-1"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
 <head>
  <title>
  </title>
 </head>
 <body>
  <p>
   Recent legislation is recognizing the need to blind patients
```

Compile Concepts

```
Built second intermediate configuration file (step 5 of 9).
Built internal concept matcher (step 6 of 9).
Finished building concepts (step 7 of 9).
Contextual extraction support detected: preparing to build LITI concepts.
Building LITI concepts, please wait...
LITI concepts build successful.
Finished removing temporary files (step 8 of 9).
Finished.  Now freeing resources (step 9 of 9).
```

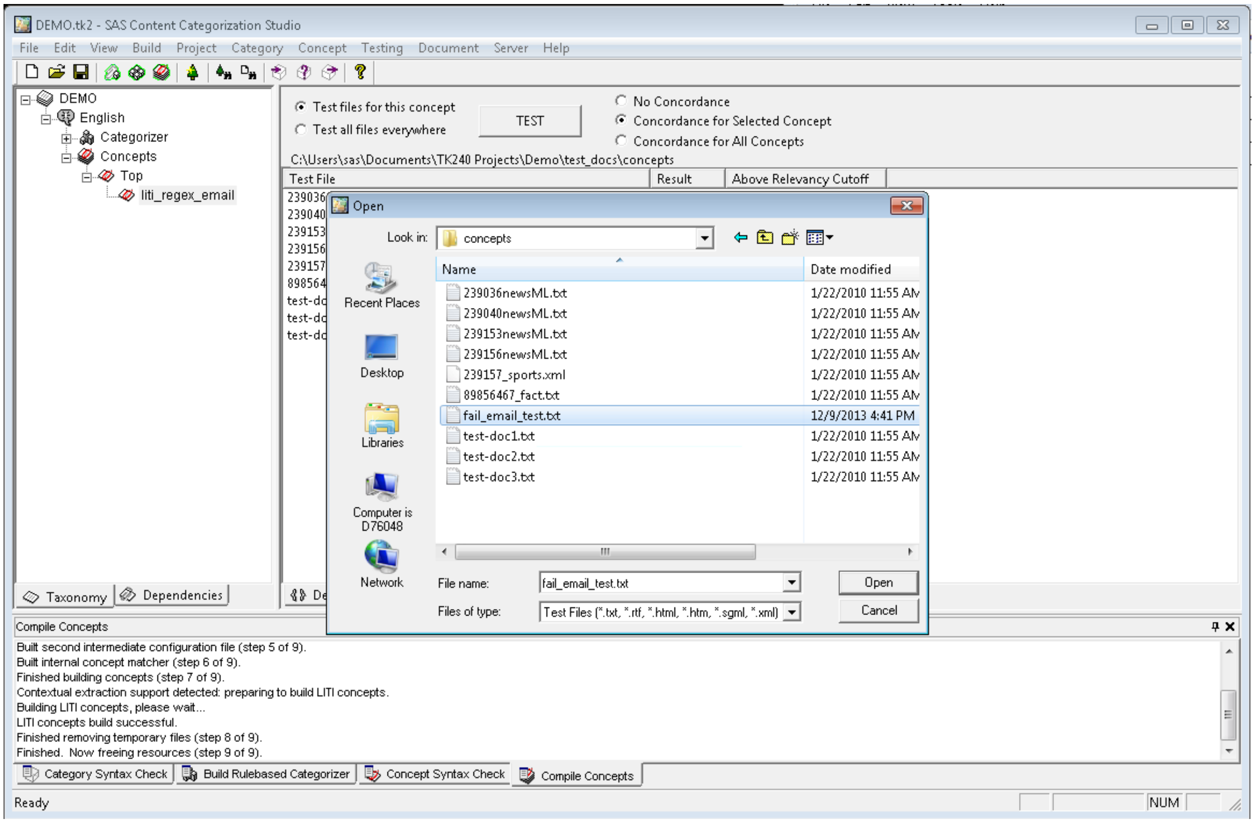
A good model should match all the relevant test documents (recall) while not matching irrelevant test documents (precision).

8. To test the failure of such irrelevant documents which do not satisfy the LITI definition of the concept, we can check by clicking Testing->Import Failing test Files… on the menu bar.


DEMO.tk2 - SAS Content Categorization Studio
File Edit View Build Project Category Concept Testing Document Server Help
Import Test Files...
Import Failing Test Files...
Delete Selected Test File...
Save Test Results
View Saved Results
Export Testing Results To File
Populate Testing Paths...
Restore Populate Results
Show Graphical Populate Results...
Full Test Report...
Show Last Test Report...
Graphical Full Test Report...
Show Last Graphical Full Test Report...
DEMO
English
Categorizer
Concepts
Top
liti_regex_email
Test File:
s\concepts\239157_sports.xml
Go
Stop
ials Is Recognized
In Recent
.com</ByLine>
ightLine>
<!DOCTYPE
"http://www.
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<title>
</title>
</head>
<body>
<p>
Recent legislation is recognizing the need to blind patients
2 Matches
TEST
Selected concept
All concepts
All categories and all concepts
Concordance
Browser View
View Rule Matches
Taxonomy
Dependencies
Definition
Testing
Data
Document
Compile Concepts
Category Syntax Check
Build Rulebased Categorizer
Concept Syntax Check
Compile Concepts
Ln 112
NUM

```
Built second intermediate configuration file (step 5 of 9).
Built internal concept matcher (step 6 of 9).
Finished building concepts (step 7 of 9).
Contextual extraction support detected: preparing to build LITI concepts.
Building LITI concepts, please wait...
LITI concepts build successful.
Finished removing temporary files (step 8 of 9).
Finished.  Now freeing resources (step 9 of 9).
```

Open the failing document *fail_email_test.txt* to be tested as shown below



● Test files for this concept
○ Test all files everywhere

○ No Concordance
● Concordance for Selected Concept
○ Concordance for All Concepts

C:\Users\sas\Documents\TK240 Projects\Demo\test_docs\concepts

Look in: concepts

Name	Date modified
239036newsML.txt	1/22/2010 11:55 AM
239040newsML.txt	1/22/2010 11:55 AM
239153newsML.txt	1/22/2010 11:55 AM
239156newsML.txt	1/22/2010 11:55 AM
239157_sports.xml	1/22/2010 11:55 AM
89856467_fact.txt	1/22/2010 11:55 AM
fail_email_test.txt	12/9/2013 4:41 PM
test-doc1.txt	1/22/2010 11:55 AM
test-doc2.txt	1/22/2010 11:55 AM
test-doc3.txt	1/22/2010 11:55 AM

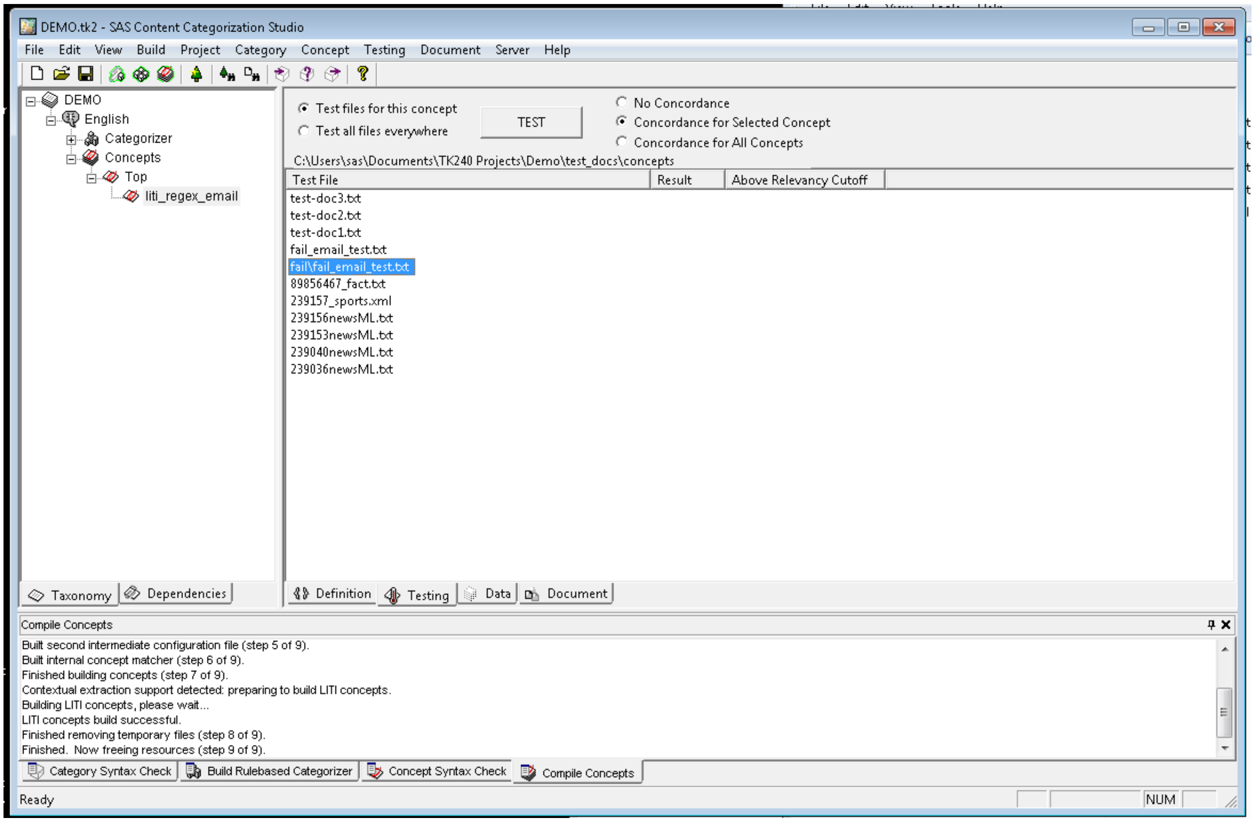
File name: fail_email_test.txt

Files of type: Test Files (*.txt, *.rtf, *.html, *.htm, *.sgml, *.xml)

Compile Concepts

```
Built second intermediate configuration file (step 5 of 9).
Built internal concept matcher (step 6 of 9).
Finished building concepts (step 7 of 9).
Contextual extraction support detected: preparing to build LITI concepts.
Building LITI concepts, please wait...
LITI concepts build successful.
Finished removing temporary files (step 8 of 9).
Finished.  Now freeing resources (step 9 of 9).
```

The imported file will be visible in the Testing tab.



● Test files for this concept
○ Test all files everywhere

○ No Concordance
● Concordance for Selected Concept
○ Concordance for All Concepts

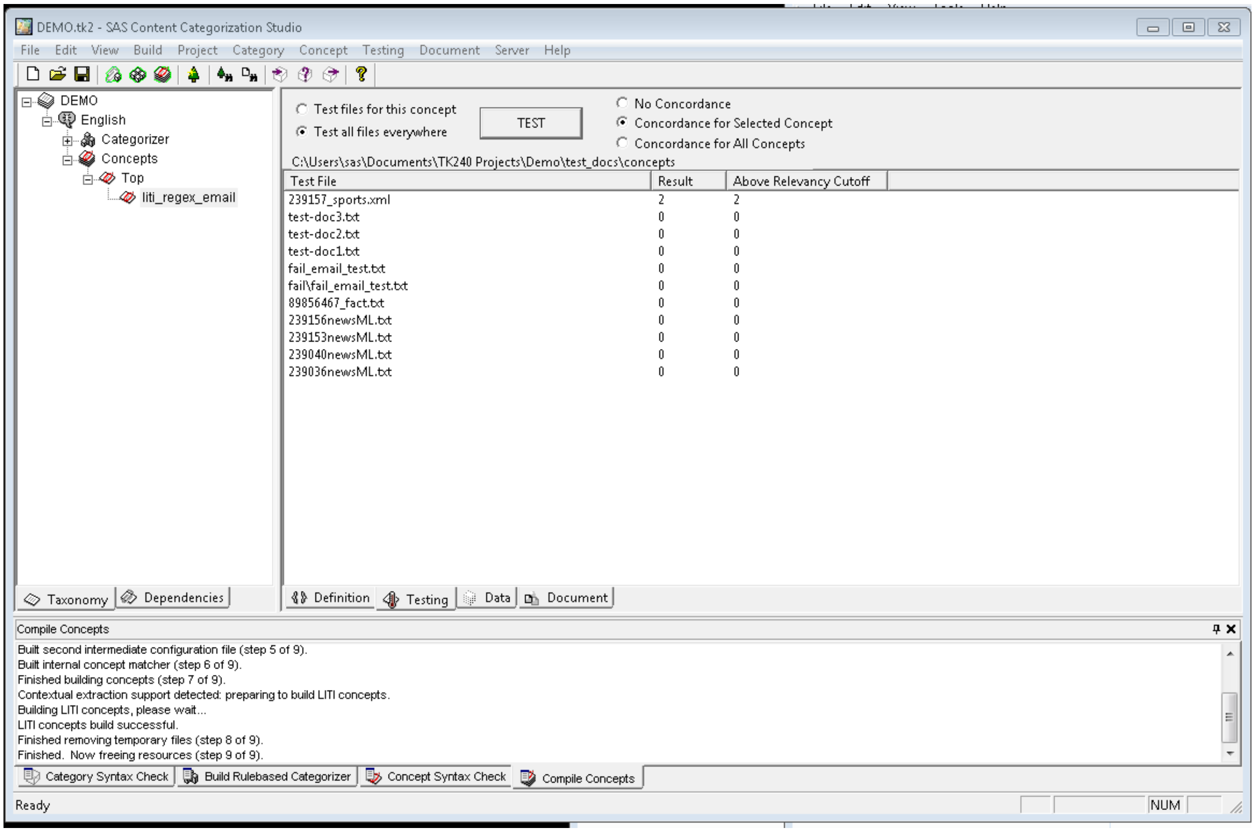
C:\Users\sas\Documents\TK240 Projects\Demo\test_docs\concepts

Test File	Result	Above Relevancy Cutoff
test-doc3.txt		
test-doc2.txt		
test-doc1.txt		
fail_email_test.txt		
fail\fail_email_test.txt		
89856467_fact.txt		
239157_sports.xml		
239156newsML.txt		
239153newsML.txt		
239040newsML.txt		
239036newsML.txt		

Compile Concepts

```
Built second intermediate configuration file (step 5 of 9).
Built internal concept matcher (step 6 of 9).
Finished building concepts (step 7 of 9).
Contextual extraction support detected: preparing to build LITI concepts.
Building LITI concepts, please wait...
LITI concepts build successful.
Finished removing temporary files (step 8 of 9).
Finished.  Now freeing resources (step 9 of 9).
```

In the test, the imported file fail_email_test.txt should not match any strings. You can verify this by noticing that the value for the Result column is 0. This shows that our concept definition is precise (filtering irrelevant documents) and working as expected.



○ Test files for this concept
● Test all files everywhere

○ No Concordance
● Concordance for Selected Concept
○ Concordance for All Concepts

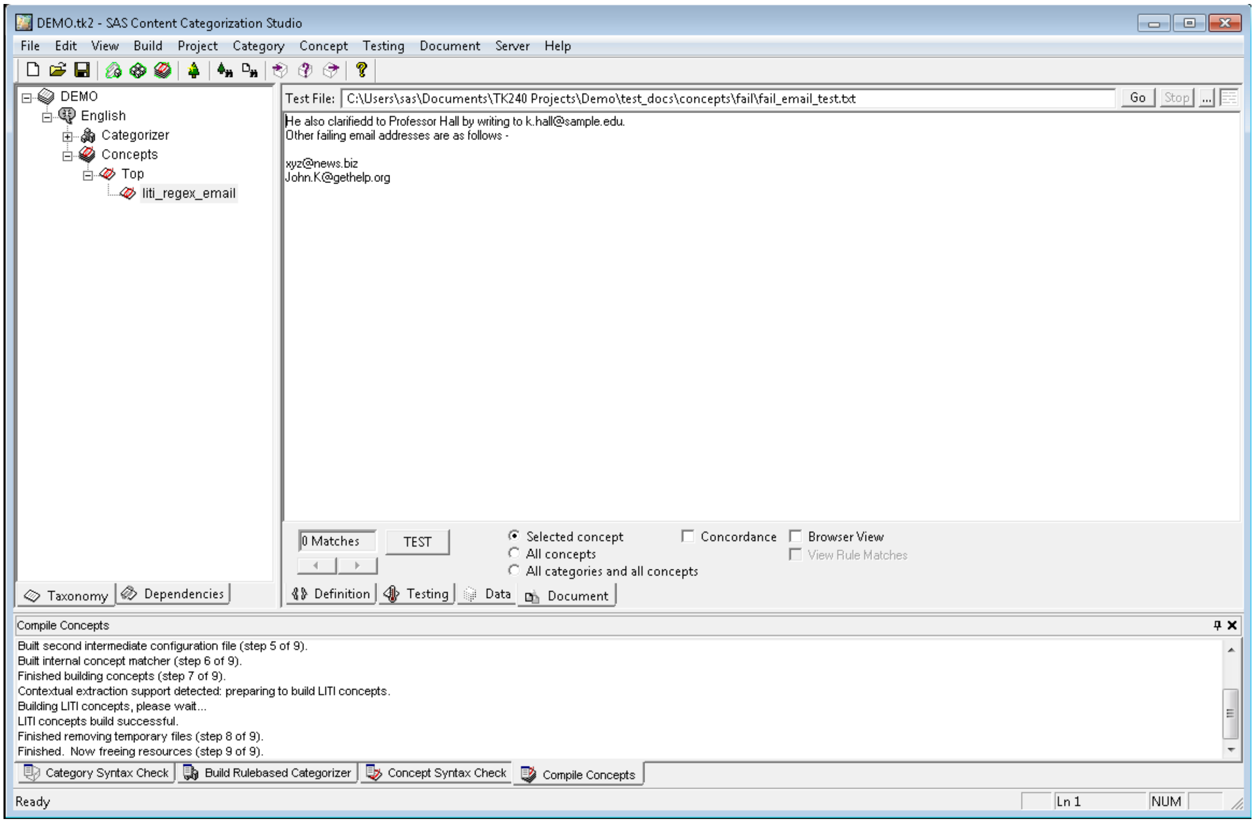
C:\Users\sas\Documents\TK240 Projects\Demo\test_docs\concepts

Test File	Result	Above Relevancy Cutoff
239157_sports.xml	2	2
test-doc3.txt	0	0
test-doc2.txt	0	0
test-doc1.txt	0	0
fail_email_test.txt	0	0
fail\fail_email_test.txt	0	0
89856467_fact.txt	0	0
239156newsML.txt	0	0
239153newsML.txt	0	0
239040newsML.txt	0	0
239036newsML.txt	0	0

Compile Concepts

```
Built second intermediate configuration file (step 5 of 9).
Built internal concept matcher (step 6 of 9).
Finished building concepts (step 7 of 9).
Contextual extraction support detected: preparing to build LITI concepts.
Building LITI concepts, please wait...
LITI concepts build successful.
Finished removing temporary files (step 8 of 9).
Finished.  Now freeing resources (step 9 of 9).
```

No strings get highlighted for the file fail_email_test.txt as it does not satisfy our REGEX LITI Definition



Test File: C:\Users\sas\Documents\TK240 Projects\Demo\test_docs\concepts\fail\fail_email_test.txt

```
He also clarifiedd to Professor Hall by writing to k.hall@sample.edu.
Other failing email addresses are as follows -

xyz@news.biz
John.K@gethelp.org
```

0 Matches

● Selected concept
○ All concepts
○ All categories and all concepts

□ Concordance
□ Browser View
□ View Rule Matches

Compile Concepts

```
Built second intermediate configuration file (step 5 of 9).
Built internal concept matcher (step 6 of 9).
Finished building concepts (step 7 of 9).
Contextual extraction support detected: preparing to build LITI concepts.
Building LITI concepts, please wait...
LITI concepts build successful.
Finished removing temporary files (step 8 of 9).
Finished.  Now freeing resources (step 9 of 9).
```