

# Quickstart Guide

Author: Thomas Mikscha  
E-Mail: [thomas.mikscha@gmail.com](mailto:thomas.mikscha@gmail.com)  
Last modified: 1/26/2014 12:45:24 PM

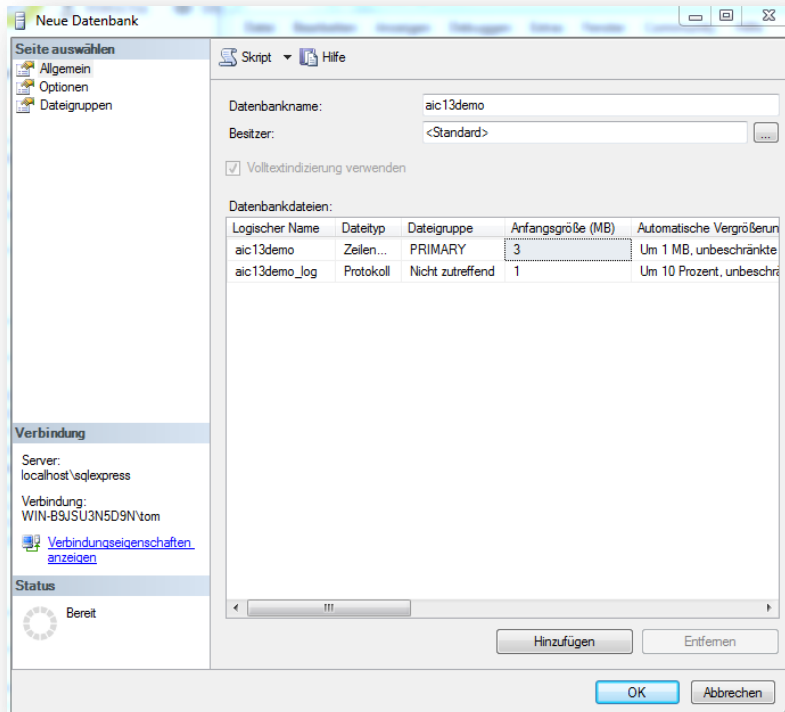
## Requirements

- Windows 7
- .NET Framework 4.0
- IIS 7.0 Webserver
- Microsoft SQL Server 2008 R2
- Microsoft SQL Server Management Studio (optional)
- LinqToTwitter
- Log4net
- JSON.net
- Twitter API key

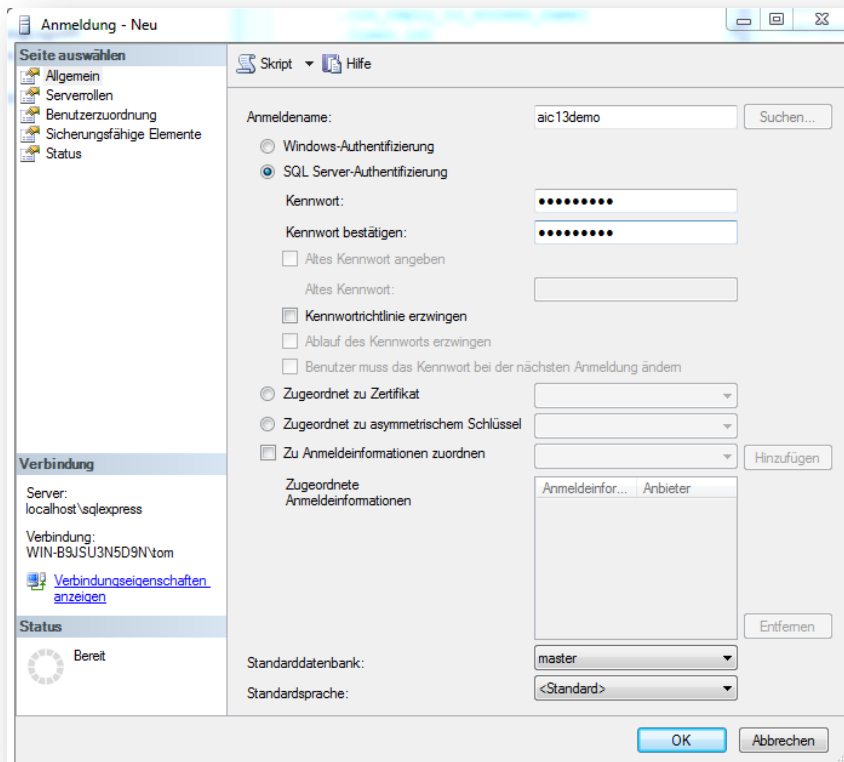
# Installation and Deployment

## Database

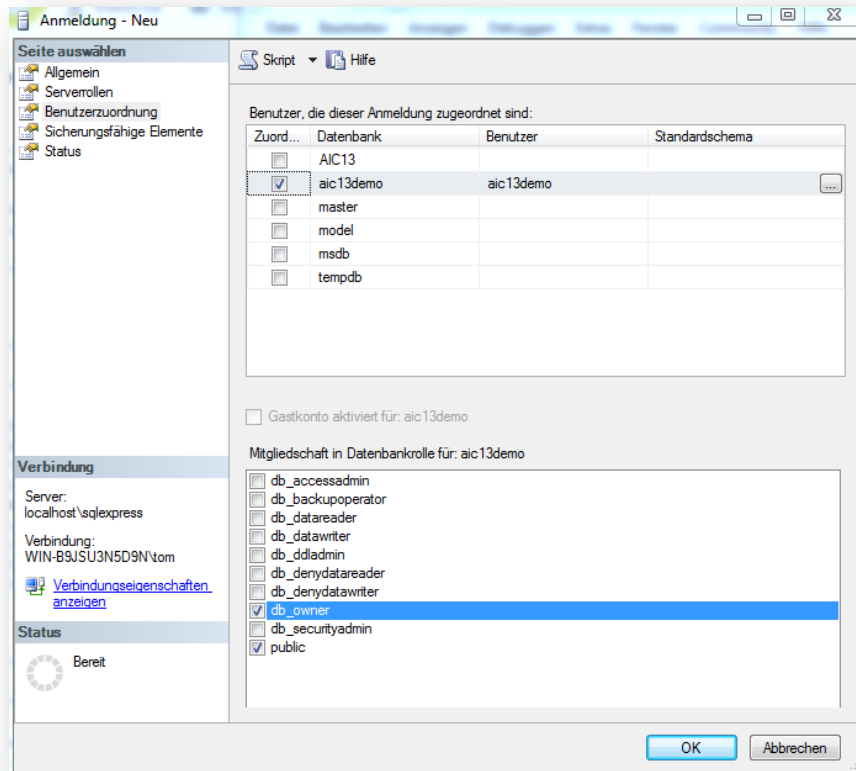
1. Create an empty database called aic13demo



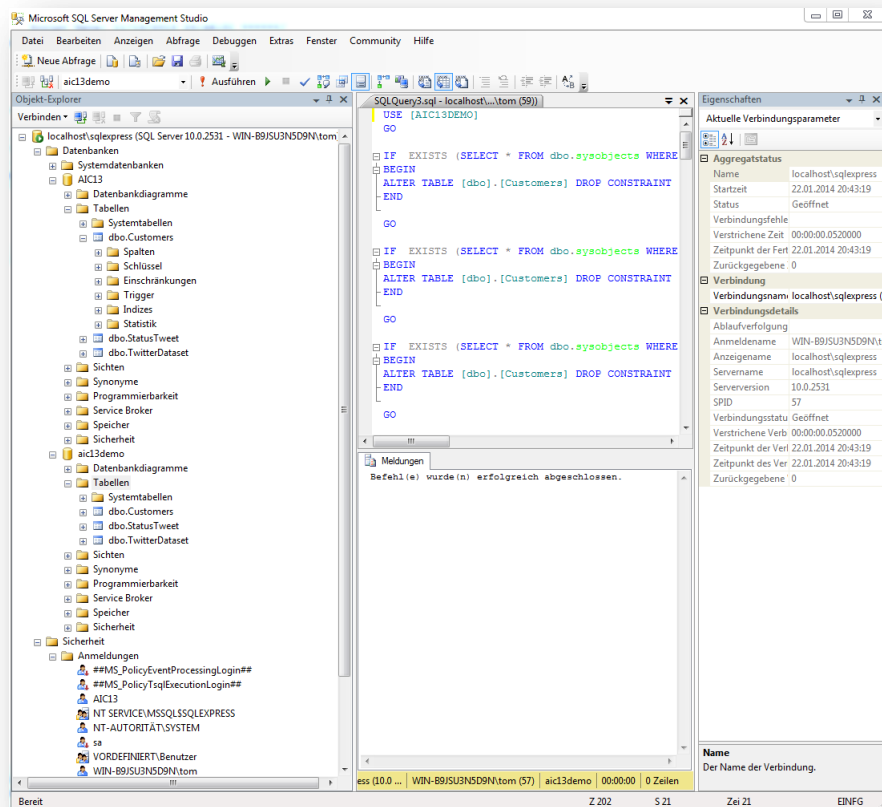
2. Create a new user login on the SQL server e.g. aic13demo



3. Add the user to the database and assign the roles "public" and "db\_owner"



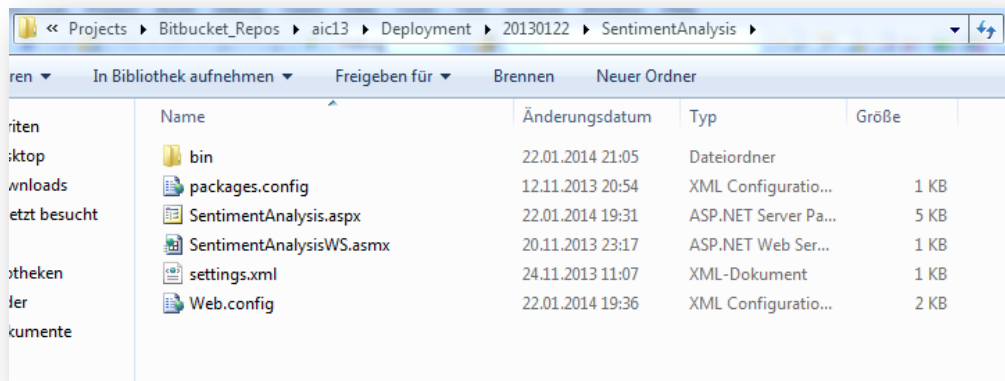
4. Open the Microsoft SQL Server Management Studio, connect to the database, load the CreateTables.sql script and execute it in order to create the table structure.



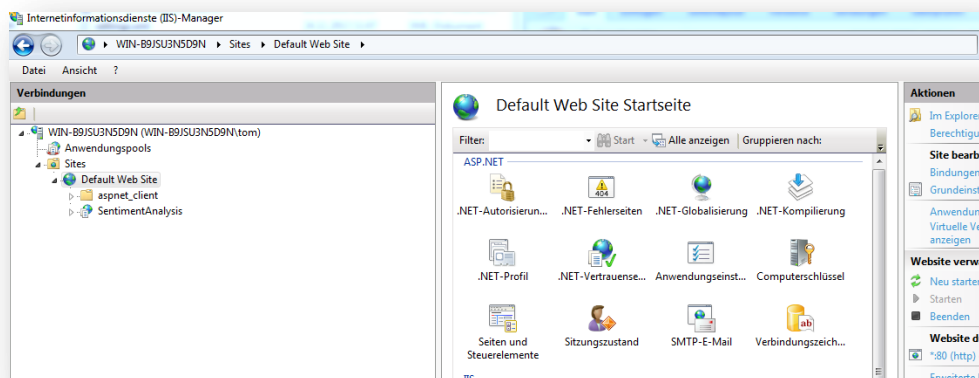
- Finally the Twitter dataset (corpus) from <http://cs.stanford.edu/people/alecmgo/trainingandtestdata.zip> must be imported into the table "TwitterDataset"
- (Optional) Database backup: The deployment folder contains a complete dump of the sentiment database already containing the training data (corpus), so the database can be restored using this dump without having to create it from scratch using the scripts. (File: aic13\_database\_dump.zip)

## Webserver

- Copy the folder "SentimentAnalysis" containing the webservice and webfrontend to your webserver



- Next, open the IIS Management Console, click on "Default website" and map the webapplication by creating a new "Web application"



- Specify an alias where the webapplication should be accessed later on

Anwendung hinzufügen

Sitename: Default Web Site  
 Pfad: /

Alias: aic13demo      Anwendungspool: DefaultAppPool      Auswählen...

Beispiel: Vertrieb

Physikalischer Pfad:  
 C:\inetpub\wwwroot\SentimentAnalysisDemo      ...

Pass-Through-Authentifizierung

Verbinden als...      Einstellungen testen...

OK      Abbrechen

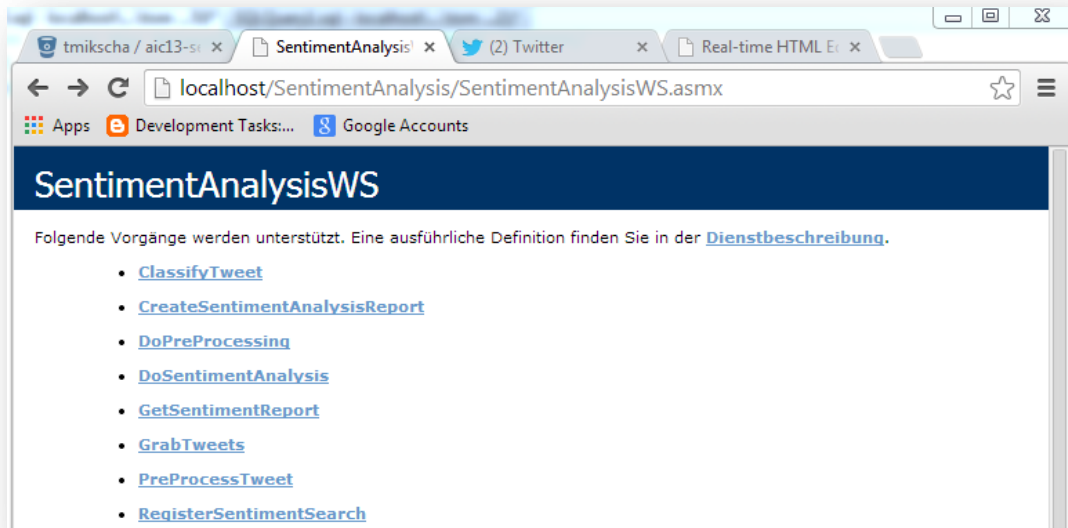
4. Ensure that the Framework is set to .NET 4.0
5. Open the settings.xml file and specify your credentials for accessing the Twitter API and your SQL server

```

1  <?xml version="1.0" encoding="utf-8"?>
2  <configuration>
3    <twitter>
4      <accesstoken>2168391810-hT7DphEiNHH0zKQ7yo8qGk1fLd33uVvK1pGbV18y</accesstoken>
5      <accesstokensecret>NJ8LyUexKQ5uaUQzC0ldZRIxG9Eyx9rG6Tz0Hyv6a2G9R</accesstokensecret>
6      <consumerkey>AIZbkCf8LvbLoaR4XR4RaQ</consumerkey>
7      <consumersecret>HPnzTzEte6FVzisYeb1hNUUSLewYsFGjwV4E9S4yA</consumersecret>
8      <accountname>aic13tmik</accountname>
9      <password>ms1595p8</password>
10    </twitter>
11    <database>
12      <dbserver>LOCALHOST\SQLEXPRESS</dbserver>
13      <dbname>aic13demo</dbname>
14      <user>aic13demo</user>
15      <password>aic13demo</password>
16    </database>
17  </configuration>

```

6. Then the webservises can be accessed using the URL  
<http://localhost/aic13demo/SentimentAnalysisWS.asmx>



7. The GUI for registering new keywords for sentiment analysis should be accessible using the URL <http://localhost/aic13demo/SentimentAnalysis.aspx>

localhost/aic13dem x aic13tmik (aic13tmi x tmikscha / aic13-se x Neuer Tab x

localhost/aic13demo/SentimentAnalysis.aspx

Apps Development Tasks... Google Accounts

Company ID:

Company Name:

E-Mail:

Keyword:

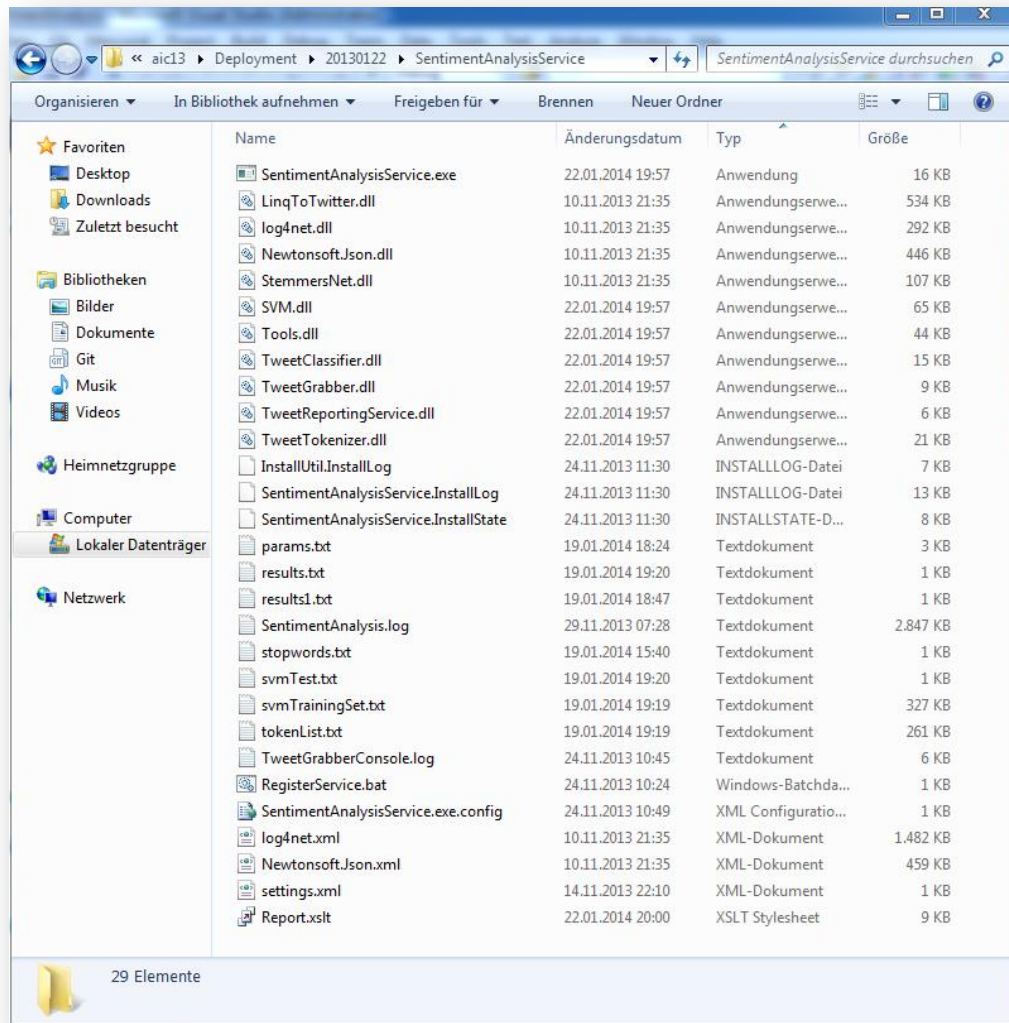
Algorithm: Simple Classifier (Bayes) ▼

Limit: 100 ▼

Register Reset Report

## Windows Service

1. Copy the folder "SentimentAnalysisService" to the desired target directory

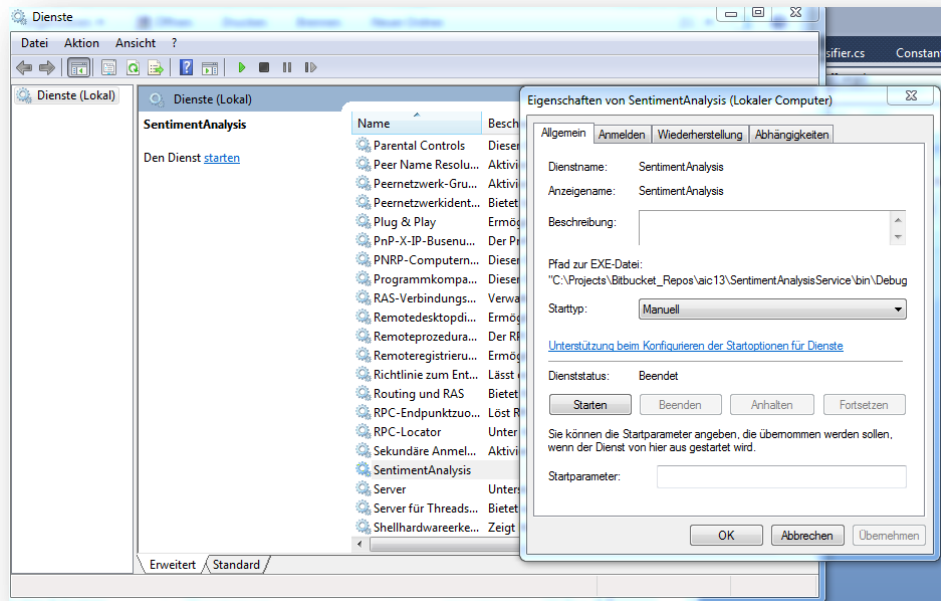


- Again, add your credentials to the settings.xml which can be copied from the web application
- In order to run the sentiment analysis process, the service must be registered. This can be either done by executing the RegisterService.bat file located in the service directory or by running the following commands from an administrator console

```
%windir%/Microsoft.NET/Framework/v4.0.30319/InstallUtil /u
SentimentAnalysisService.exe
%windir%/Microsoft.NET/Framework/v4.0.30319/InstallUtil
SentimentAnalysisService.exe
```

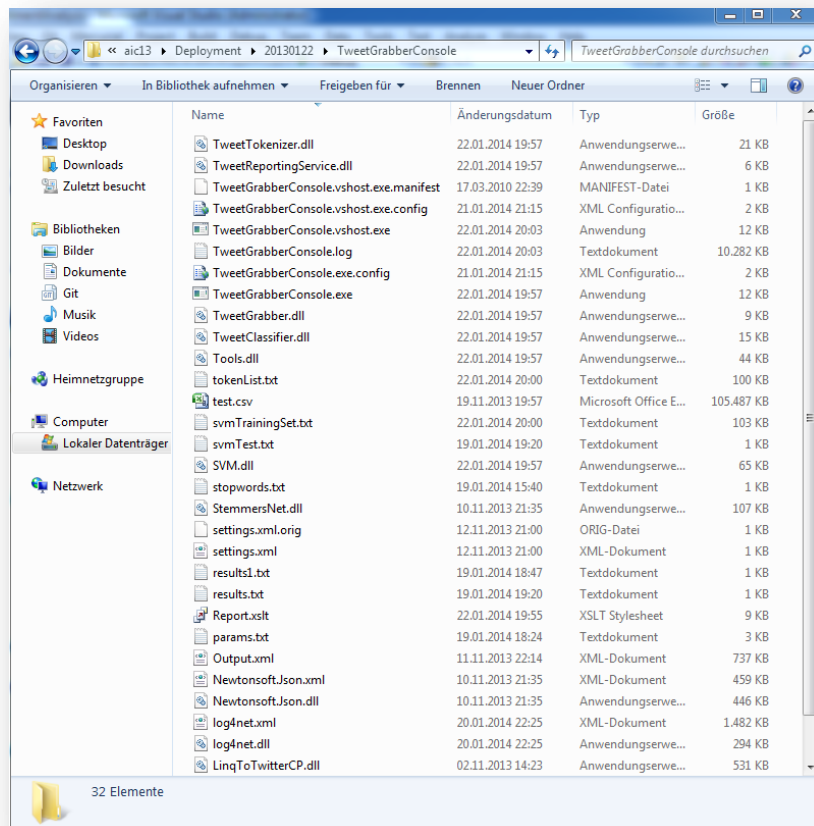
- Then you should see a new entry in your service console





## TweetGrabberConsole

1. For testing and debugging or simply running the collection of tweets manually, the TweetGrabberConsole can be used. The console offers better capabilities to monitor the whole sentiment analysis process compared to the windows service
2. Copy the folder TweetGrabberConsole to the desired target folder



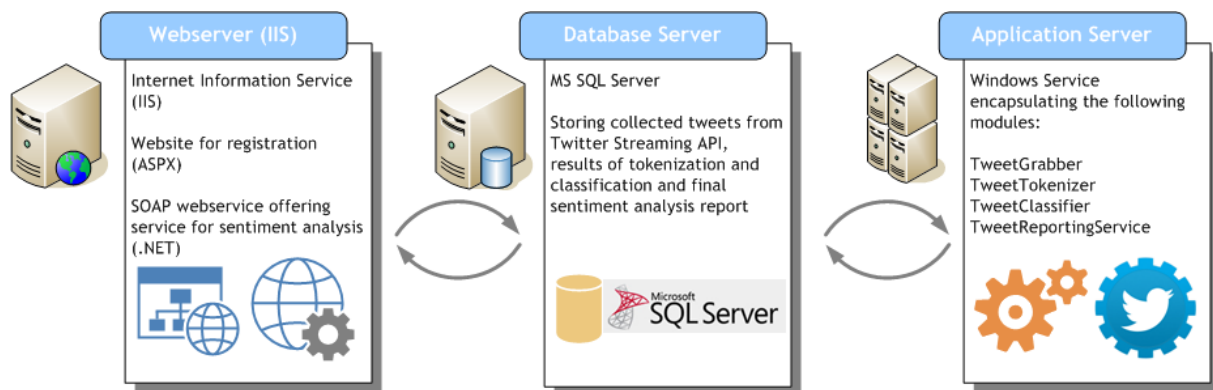
3. Modify the settings.xml by adding your credentials for the SQL server and the Twitter API

# Usage

## Overview of Architecture

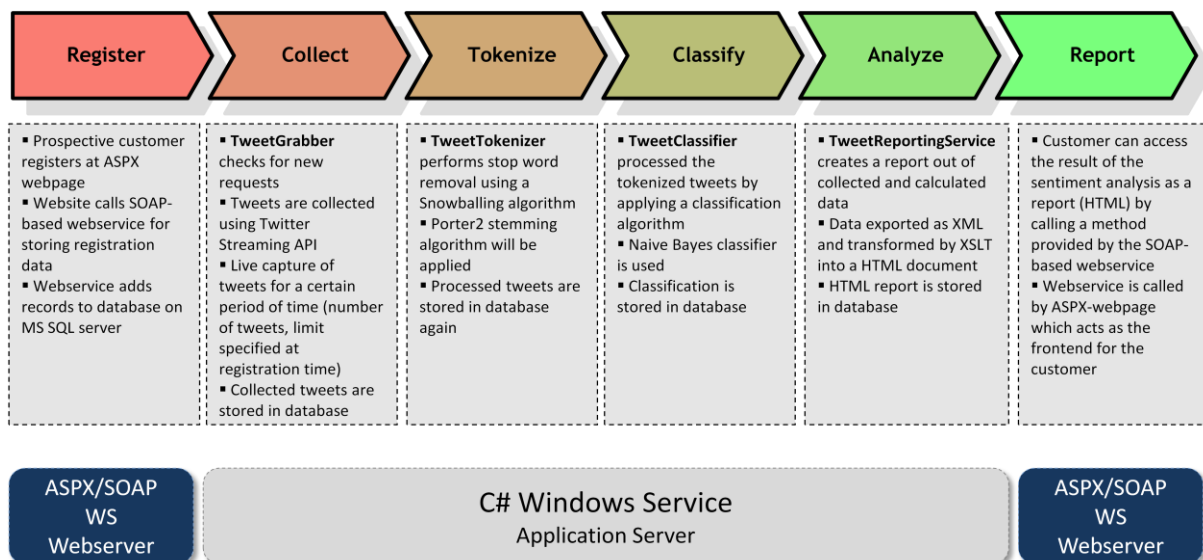
The sentiment analysis process is basically split into three modules:

- A common database server where all collected tweets and jobs (e.g. keywords to analyze) are stored
- A webapplication that is used for registering new keywords and accessing the reports after the classification (=user frontend for interaction)
- TweetGrabberConsole/Windows Service for running the classification process



Generally speaking whole classification process consists of the following steps/phases

- Register a new search term or keyword for analysis (Webapplication)
- Collect, tokenize, classify and analyze the tweets (TweetGrabberConsole or Windows Service)
- Reporting using XSLT (Webapplication)



## Usage of Components

1. Register a new search term that should be analyzed in the Twitter universe using the webapplication / - frontend e.g. Justin Bieber

localhost/SentimentAnaly x aic13tmik (aic13tmik) auf x tmikscha / aic13-sentimer x

localhost/SentimentAnalysis/SentimentAnalysis.aspx

Apps Development Tasks:... Google Accounts

Company ID:

Company Name:

E-Mail:

Keyword:

Algorithm:

Limit:

2. After pressing the Register button, the search term will be added to the database for further processing by either the TweetGrabberConsole or the SentimentAnalysis Windows service

Microsoft SQL Server Management Studio

SQLQuery4.sql - localhost\...tom (56) CreateTables.sql - localhost\...tom...)

```

/***** Skript für SelectTopNRows-Befehl aus SSMS *****/
SELECT TOP 1000 [id]
,[company_name]
,[email_address]
,[keyword]
,[status]
,[stage]
,[classifier]
,[htmlreport]
,[limit]
FROM [AIC13].[dbo].[Customers]

```

	id	company_name	email_address	keyword	status	stage	classifier	htmlreport
1	136	justin bieber	asd	Justin Bieber	2	3	2	NULL
2	137	Lady Gaga	weird@office.com	Lady Gaga	2	4	3	0xEFBBBF3C68746D6C207
3	138	Justin Bieber	Justin Bieber	Justin Bieber	2	4	3	0xEFBBBF3C68746D6C207
4	139	Justin Bieber	test@test.com	Justin Bieber	1	1	3	NULL

Eigenschaften

Aktuelle Verbindungsparameter

Aggregatstatus

Name localhost\sqlxpres

Startzeit 22.01.2014 21:47:32

Status Geöffnet

Verbindungsfehler

Verstrichene Ze 00:00:00.0920092

Zeitpunkt der F 22.01.2014 21:47:32

Zurückgeber 4

Verbindung

Verbindungsna localhost\sqlxpres

Verbindungsdetails

Ablaufverfolgung

Anmeldename WIN-B9JSU3N5D9N

Anzeigenname localhost\sqlxpres

Servername localhost\sqlxpres

Serverversion 10.0.2531

3. Running the TweetGrabberConsole will start the classification process. The tool will query the database looking for open tasks and running the
  - a. Collection of tweets
  - b. Tokenization (stop word removal, stemming etc.)
  - c. Classification using the selected algorithm (Simple Bayesian classifier, Naïve Bayes or SVM)
  - d. Evaluation (Reporting/XSLT)

```
C:\Projects\Bitbucket_Repos\aic13\TweetGrabberConsole\bin\Debug\TweetGrabberConsole.exe

#33 - ?Spread Your Wings? "Swaggy_xz" tweeted at 22.01.2014 21:03:42: RT @JBSour
ce: People seem to forget that Justin Bieber is a human being just like the rest
of us.
#33 - Follow Jasmine, ily "VILLEGASPORRA" tweeted at 22.01.2014 21:03:42: Foto:
lovemisterbieber: Photo - Last chance to see Justin Bieber's tour here http://t.
co/x0qUG0sHWD http://t.co/Zm8vQD2xiy
#35 - Lujain "curlyromeox" tweeted at 22.01.2014 21:03:42: RT @birdcagewifi: "T
he Wanted split"
"Justin Bieber retiring"

we're slowly taking over twitter

are you scared bc i am
#36 - clara "thoughtdrew" tweeted at 22.01.2014 21:03:43: RT @selfiejustins: "Yo
u're too obsessed with Justin Bieber" http://t.co/LfeaeA3gw3
#37 - Djulayn "DjulsCorputty" tweeted at 22.01.2014 21:03:43: Justin Bieber - Al
l that matters ??
#38 - FOLLOW ME JUSTIN ? "kidrauhlninjax" tweeted at 22.01.2014 21:03:43: RT @se
lfiejustins: "You're too obsessed with Justin Bieber" http://t.co/LfeaeA3gw3
#39 - Sam Everson "samevo2" tweeted at 22.01.2014 21:03:43: My psych teacher sai
d justin bieber is a horrible person because he spent 75 grand at a strip club..
. That's not a reason to hate lol
#40 - Believe "Sonia_MadiDamon" tweeted at 22.01.2014 21:03:43: "@jxstinspizza:
dylan obrien, justin bieber & tyler posey aka my life" Same
```

```
C:\Projects\Bitbucket_Repos\aic13\TweetGrabberConsole\bin\Debug\TweetGrabberConsole.exe

Nipple .
#97 - paul walker "mymjnd" tweeted at 22.01.2014 21:04:07: RT @selfiejustins: "Y
ou're too obsessed with Justin Bieber" http://t.co/LfeaeA3gw3
#98 - Follow Jasmine, ily "VILLEGASPORRA" tweeted at 22.01.2014 21:04:09: Foto:
lovemisterbieber: Photo - Last chance to see Justin Bieber's tour here http://t.
co/x0qUG0sHWD http://t.co/0cg00KNcFp
#98 - Cristiano Ronaldo. ♥ "Cnovas_PaulaCR7" tweeted at 22.01.2014 21:04:09: RT
@InJustinWorld: Every Belieber wants to hear these 5 words, 8 syllables, 25 lett
ers, one dream come true: "You are meeting Justin Bieber.
#100 - 8 Jaque Beatriz ? "JBeatriz_s2" tweeted at 22.01.2014 21:04:10: Hey, you'
re the best idol knew? I love you very much @justinbieber ♥ quarenta e quatro'
#ArtistaDoSéculo Justin Bieber
#101 - Alyssia Rodriguez "4_14_13_baby" tweeted at 22.01.2014 21:04:10: RT @robh
oma3: Justin Bieber got 100,000 retweets for tweeting "Live life full". That's j
ust 3 random words. I'm going to try now.

Nipple .

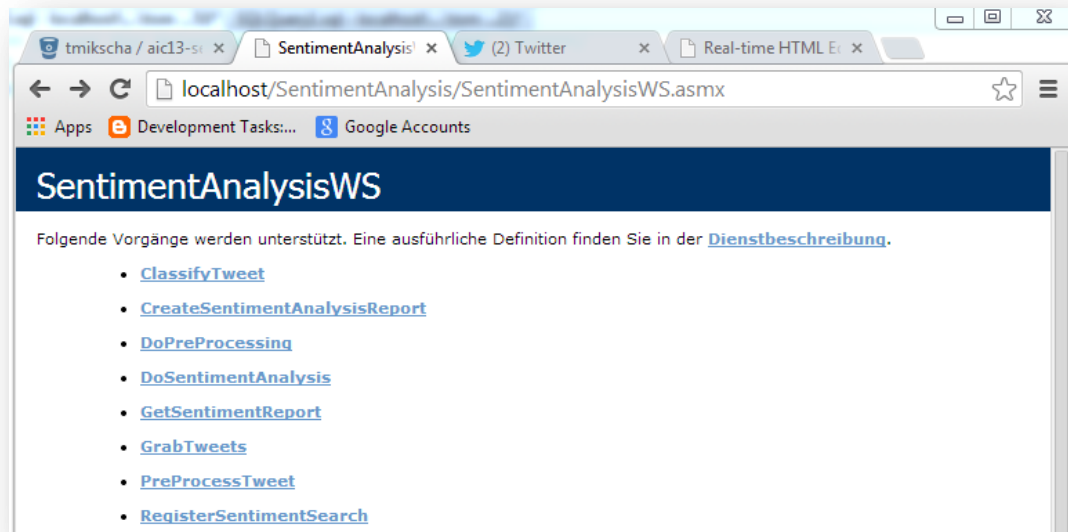
Step 2: Tokenization finished.
Step 3: Classification finished.
Step 4: Report generation finished.
```

#### 4. Accessing the final report using the webfrontend again

Company		Email	Keyword	Sentiment		Sentiment (Weighted)	
				Positive	Negative	Positive	Negative
Justin Bieber		test@test.com	Justin Bieber	65.34653465346535%	34.65346534653465%	66.14440807712259%	33.855591922877444%

User name	Followers	Weighting	Weighting %	Classification	Avg Followers	Text (Orig.)	Text (Processed)
	13041	7.8526050329988735	785.26%	1.0000	1660.7227722772277	1.0000	RT weheartjustin justin bieber is like a book he judg by the cover but nobodi read his stori i've never seen a statement more t
Lujain	10371	6.244871313337268	624.49%	0.0000	1660.7227722772277	0.0000	RT birdcagewifi the want split justin bieber retir we'r slowly take over twitter are you scare bc i am
?	10198	6.1406998026625645	614.07%	1.0000	1660.7227722772277	1.0000	ariana is my queen life she is everyth to me she isn't famous like justin bieber or 1D but she is grow bigger and bigger
maria	9823	5.914894504957283	591.49%	0.0000	1660.7227722772277	0.0000	you'r all that matter to me artistasdo século justin bieber
?follow me justin?	8061	4.853910679472733	485.39%	1.0000	1660.7227722772277	1.0000	RT bieberloft i'm justin bieber and i'll make u question your sexual http://t.co/8aagrkcy0uu
News.Stream	6971	4.197569947475048	419.76%	1.0000	1660.7227722772277	1.0000	justin bieber smoke a cuban cigar as singer camp say his ego is out of control http://t.co/vhnvov7pav
	5555	3.344929143340905	334.49%	0.0000	1660.7227722772277	0.0000	RT alexisbrodi I just met justin drew bieber I cant contain MY feel right now cmg I love you with all MY heart justin http://t

5. Another approach is to access all components using the webservice methods available under <http://localhost/SentimentAnalysis/SentimentAnalysisWS.asmx>



- Helper-Method: ClassifyTweet
  - Takes a string containing a preprocessed text and an integer for the desired classifier e.g. 1=NaïveBayes, 2=SVM and 3=SimpleClassifier
- Helper-Method: PreProcessTweet
  - Takes a string containing a tweet and the language of the tweet ("EN", "DE", etc.) and applied the pre-processing (stemming, stopword-removal, etc.)
- Helper-Method: GetSentimentReport
  - Returns the previously created HTML report containing the sentiment analysis for the specified customer/job id
- Step 1: RegisterSentimentSearch
  - Allows the registration of a new search term for sentiment analysis
  - Adds the "keyword search" as a job to the database for further processing
  - Returns the ID of the search job
- Step 2: GrabTweets
  - Starts the collection of tweets for the specified ID
  - Uses the Twitter streaming API
  - Adds the tweets to the database and returns the number of collected tweets
- Step 3: DoPreProcessing
  - Applies the stemming, stopword-removal, etc. on all collected tweets for the specified customer/job id
  - Returns an HTML message containing all processed tweets
- Step 4: DoSentimentAnalysis
  - Starts the classification for the specified customer/job id
  - Return a HTML string containing all tweets together with their classification
- Step 5: CreateSentimentAnalysisReport
  - Creates a HTML report of the collected tweets containing the results of the classification
  - XML data will be transformed to HTML using XSLT
  - HTML file is stored in database and returned to the client