## **READING:**

Retriving nEws Around the worlD for dIscovery and kNowledge mininG

# News Classifier User Guide

## 1. Requirements

- .NET Framework 1.1
- MySQL Server 5
- MySQL ODBC Driver 3.1

## 2. Setup

### Database Setup

The file "notion-backup.sql" is the current backup of the database. The figure below shows the tables in the database.



In addition the following users with the following passwords should be created. The passwords can be changed, but will require editing the news classifier source (notiondatabase.cs).

USERNAME	PASSWORD
localprogram	kant2006
webuser	kant2006

In addition a DSN source should be created using the MySQL ODBC Driver. The source should be named "notion" and it should use the "localprogram" user and point to the "notion" database.

Administrator	<u>?</u> ×		
User DSN   System DSN   File DSN   Drivers   Tracing   Connection Pooling   About			
Driver Microsoft dBase Driver (*.dbf) Microsoft Excel Driver (*.ds) MySQL ODBC 3.51 Driver Microsoft Access Driver (*.mdb) MySQL ODBC 3.51 Driver	A <u>d</u> d <u>R</u> emove <u>C</u> onfigure		
An ODBC User data source stores information about how to connect to the indicated data provider. A User data source is only visible to you, and can only be used on the current machine.			
	V       File DSN       Drivers       Tracing       Connection         Driver       Microsoft dBase Driver (*.dbf)         Microsoft Excel Driver (*.dsf)       Microsoft Excel Driver (*.dsf)         MySQL ODBC 3.51 Driver       Microsoft Access Driver (*.mdb)         MySQL ODBC 3.51 Driver         Microsoft Access Driver (*.mdb)         MySQL ODBC 3.51 Driver         Jser data source stores information about how t         d data provider.       A User data source is only visily be used on the current machine.         OK       Cancel       Apply		

Connector/ODBC 3.51.12 - Configure Data Source Name		
Connector	VODBC	MySQL
Login Connect Opt	ions Advanced	Connector/ODBC Configuration
Data Source Name	Notion	This dialog is used to edit a Data Source Name (DSN).
Description	Notion	
Server		
User	localprogram	
Password		
Database	notion	
<u>T</u> est	Diagnostics >> Ok	<u>C</u> ancel <u>H</u> elp

#### **Program Setup**

Copy the News Classifier application & source folder to someplace on your computer  $(C:\ is ok)$ . The binary (NewsClassifier.exe) is in the "bin" directory. The data directory contains the training data for the category classifier and the rules and dictionaries for the named entity extractor. The plugins folder contains the English language plugin and its dictionary and stopword list.



Before running the application the "NewsClassifier.exe.config" should be edited. The .config file has two options: Store and Timeout. The "Store" option tells the program where to look for new news articles. The "Timeout" option tells the program the maximum allowed time for processing an article and is milliseconds.

```
NewsClassifier.exe.config - メモ帳
ファイル(E) 編集(E) 書式(Q) 表示(V) ヘルブ(H)
<?xml version="1.0" encoding="utf-8" ?>
<configuration>
<appSettings>
<add key="Store" value="z:¥¥" />
<add key="Timeout" value="60000" />
</appSettings>
</configuration>
```

#### 3. Program Execution

The program consists of a single binary "NewsClassifier.exe" that does all the work. Once the program is started it automatically starts to process articles if any are available. It is important that "Store" location from the config file is accessible or the program will crash. Any errors or other problems will be logged to "log.txt" in the bin directory.

\*\*Document preparer David B. Bracewell