

READING:

Retrieving nEws Around the world for dIScovery and kNoledge mininG

News Classifier User Guide

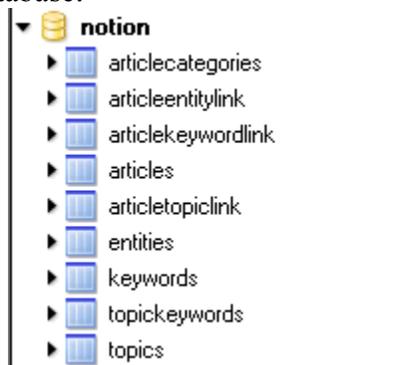
1. Requirements

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

2. Setup

Database Setup

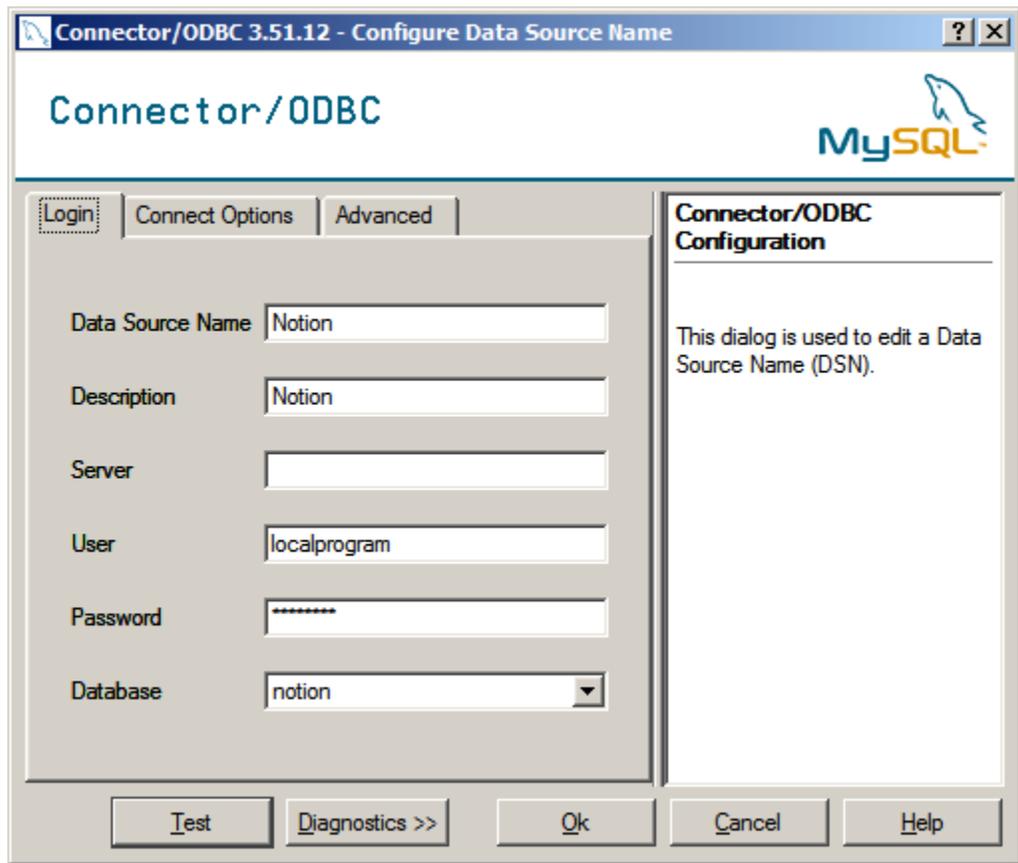
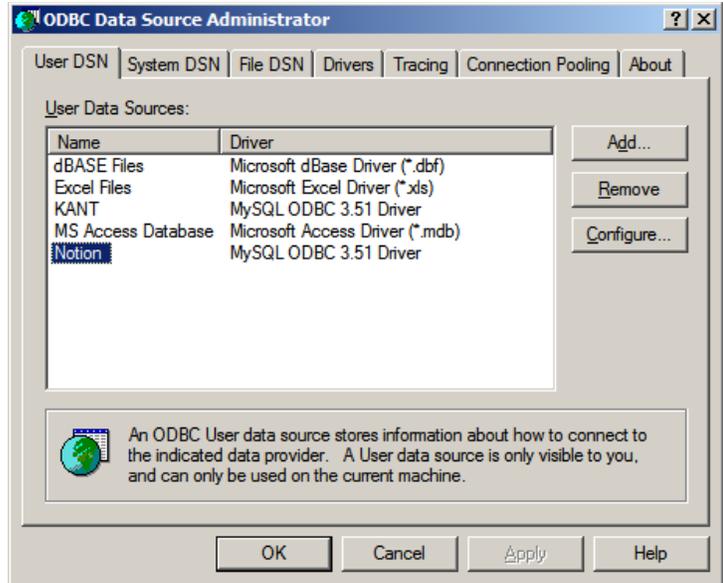
The file “notion-backup.sql” is the current backup of the database. The figure below shows the tables in the database.



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.



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 stopwords 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(F) 編集(E) 書式(O) 表示(V) ヘルプ(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