SimpleReader

by Mai Nguyen, NeXT Developer Support Team

Overview

This example shows how to build a customized language reader. The reader is organized as a bundle, so that it can be dynamically loaded by the Indexing Kit, if you are using NEXTSTEP applications such as Digital Librarian. You can also hardlink the reader into your code.

The reader bundle contains 3 files:

1. The reader class itself. In our example, this class is   IXFrenchReader.[hm]. Note that the bundle file has to be named French for the French language, or German for the German language, etc. You can set the proper name in the bundle by selecting Attributes in the Project Builder file, and type French for example into the Name textfield.

2. A stop words file with the extension **.swords**. The stop words file is a new line delimited ascii file (NS encoding) containing noise words to be stripped from queries and indexes.

3. A   domain file with the extension **.domain** which is generated from a word frequency histogram. The domain file can be generated programmatically from the IXWeightingDomain class, while the histogram file is generated from the IXAttributeParser class.

The domain file is essentially a filter that captures the relative frequencies of words in the language; this information is used to bias the weights assigned to hits returned from queries. This is done to improve the accuracy of the ranking. To generate a proper domain file, you must feed to the attribute parser a representative piece of the language you are building a reader for. Please see the ReaderDomain folder for more information on how to generate a **.domain** file programmatically.

The **.swords** and **.domain** files must be placed in the Other Resources folder in the Project Builder file of the Language Reader.

In NEXTSTEP Release 3.1, the default search path for Readers must be **/NextLibrary/Readers**. This may be changed in the future. Please check the 3.2 Release Notes on the Indexing Kit for an update.

Please send bugs/comments/suggestions to mai@next.com.

Program Organization

Major Classes in the Application

IXFrenchReader   A subclass of the Indexing Kit's IXLanguageReader. The purpose of subclassing IXLanguageReader is two-fold:
1) To dynamically load the proper reader depending on the system default language selected in Preferences for NeXT applications such as Librarian.
2) To finetune the reader for a specific language. In this example, only **foldPlural:inLength:** is used. Please note that the **reduceStem:inLength** and **foldPlural:inLength:** methods are mutually exclusive, i.e., if you use one, the other won't be in effect.

Other Peculiarities
In order to build a **French.reader** bundle, you must type in a terminal shell the following command line:

make PRODUCT_ROOT=French.reader

After building the French.reader bundle, you have to place it into **/NextLibrary/Readers**.

Other References

See also the Indexing Kit Technical Documentation on IXLanguageReader, IXAttributeReader, and IXAttributeParser.

Valid for 3.1

Change History

June 1993   Created for Release 3.1