

Инструкция по установке и пользованию AGFL 2.0 parser generator system for Windows

Установка

Все файлы дистрибутива

- **epforir.zip**
- **agfl-dev_v2-0_win32_sources.zip**¹
- **agfl-rt_v2-0_win32.zip**
- **cdl3-rt_v2.01_win32.zip**

должны быть разархивированы в одну директорию. Желательно создать для этой цели папку **c:\agfl_2.2**. В том случае, если установить систему в **c:\agfl_2.2** не является возможным, необходимо провести редактирование первых двух строчек файла **..\bin\setagfl.bat**.

```
SET GENBASE=c:\ agfl_2.2
```

```
SET GENCC=c:\agfl_2.2\bin\gcc
```

Вместо **c:\agfl_2.2** необходимо записать реальный путь к директории, в которую были разархивированы файлы AGFL на вашем компьютере. Например:

```
SET GENBASE=d:\folder1\folder2\agfl
```

```
SET GENCC= d:\folder1\folder2\agfl\bin\gcc
```

Работа с AGFL 2.0 for Windows

После того, как файлы были разархивированы, необходимо запустить файл **c:\agfl\bin\setagfl.bat** (будем далее считать, что установка произведена в каталог **c:\agfl_2.2**) для того, чтобы система прописала пути к своим файлам и каталогам в среде **Windows** (производится автоматическое редактирование **c:\autoexec.bat**).

Создание пользовательского парсера

Следующим шагом является создание собственного парсера. Он может быть создан только при наличии (и, естественно, корректном описании) следующих компонентов:

- Корневая грамматика;
- Грамматики, на которые существуют ссылки в корневой грамматике;
- Главный лексический **.lex** файл;
- Все лексические **.dat** файлы, на которые существуют ссылки в **.lex** файле;

Например, если грамматика пользователя называется **mygram**, то обязательно наличие файлов:

- **mygram.gra**;
- Другие файлы с расширением **.gra** (напр. **nounphrase.gra**, **verbphrase.gra**...);
- **mygram.lex**;
- Файлы с расширением **.dat** (напр. **noun.dat**, **verb.dat**...).

NB! Все эти файлы должны находиться в каталоге **c:\agfl_2.2\bin**.

Для создния собственного парсера (пусть он будет называться **mygram**) необходимо в командной строке задать следующее:

```
c:\agfl\bin>gen mygram
```

Выполнение этой команды должно привести к созданию трех файлов:

¹ Этот архив не является обязательным для правильной работы системы. Он создан на тот случай, если пользователь желает изменять исходные тексты программ (C++) в собственных целях и, кроме того, помогает лучше понять принципы работы AGFL for Windows. Таким образом, если вы не предполагаете вносить изменения в работу отдельных модулей, то файлы этого архива разархивировать не обязательно.

- Исполняемый модуль **mygram.exe**;
- Файл ввода/вывода лексикона **mygram.lif**;
- Машинный файл с лексической информацией **mygram.blf**.

Эти файлы являются **stand-alone** файлами (то есть только они необходимы для правильной работы сгенерированного парсера, и при желании могут быть перенесены на компьютер, где нет никаких других **AGFL** файлов).

Парсинг произвольного текста

После создания пользовательского парсера, парсинг произвольного текста производится путем выполнения команды

```
c:\agfl_2.2\bin\mygram >>
```

где **mygram** – имя пользовательской грамматики (и, как следствие, соответствующего ей исполнимого **.exe** файла). Дальнейший ввод предложений осуществляется с клавиатуры. При этом:

- На одной строчке должно быть одно предложение;
- Ввод заканчивается двойным нажатием клавиши **Enter**.
- Кроме того, у команды **mygram** существует набор параметров.

Параметры пользовательской программы парсинга

Параметр	Нач. установка	Назначение
-P число	не определено	Максимальное количество выдаваемых синтаксических разборов
-T число	не определено	Максимальное время, отводимое для работы программы (в секундах) ¹
-B	выключено	Вывод синтаксического разбора только с минимальным значением penalty ²
-b	выключено	Вывод в виде скобочной записи
-H	выключено	Включение процедур гибридного парсинга ³
-O	выключено	Отключение вывода результатов синтаксического разбора ⁴
-o файл	поток stdout	Вывод в файл
-t	выключено	Осуществление transduction ⁵
-A файл	выключено	Использовать алфавитный файл
-G	выключено	Показать лексический граф
-V		Показать версию программы
-s	выключено	Осуществлять синхронизацию ввода/вывода
-v	выключено	Включить дополнительные параметры. Возможные значения:
	Options	Показать параметры парсера
	Lexer	Показать результаты работы лексического модуля
	Parser	Показать результаты работы модуля парсинга (т.е. penalties)
	Totals	Показать время работы отдельных моделей

¹ Если программа не закончила обработку введенного текста по истечении указанного периода времени, ее выполнение будет остановлено принудительно.

² То есть наиболее “правильного”, по мнению парсера, варианта разбора.

³ При парсинге учитывает информацию о частотности тех или иных конструкций и схем на основе заранее размеченного текста.

⁴ Показываются только результаты работы отдельных модулей.

⁵ См. **Manual for the AGFL system version 2.0** (<http://www.cs.kun.nl/agfl/>)

Параметр	Нач. установка	Назначение
	All	Включение предыдущих четырех опций
	Timebreak	Показать, когда осуществится timebreak
	Penalties	Показывать penalties в синтаксических разборах
-h	выключено	Вывести эту справку на экран

Например:

c:\agfl\bin\epforir -tBv lexer -v parser >> Time flies like an arrow.

Приведет к следующему выходному результату (учитывая, что **epforir** – парсер английского текста на основе одноименной грамматики; комментарии выделены курсивом):

Time flies like an arrow. \$EOSS *результаты работы лексического модуля*

parsing 1 time 0.060 penalty 5 *результаты работы модуля парсинга*

{[N:flies ,N:time],[V:like ,N:arrow]} *Вывод синтаксического разбора только с минимальным значением penalty/Осуществление transduction*

Ввод/вывод из файла/в файл

Если необходимо осуществлять ввод текста не с клавиатуры, а из файла с последующим помещением результатов работы парсера в другой файл, нужно пользоваться командой:

c:\agfl_2.2\bin\mygram < infile.txt > outfile.txt

Здесь:

- **infile.txt** и **outfile.txt** могут иметь произвольные имена и расширения;
- после **mygram** может идти весь список параметров, описанных выше;
- Во входном файле исходный текст для парсинга должен быть представлен в виде “одно предложение на строке”.

Возможные проблемы

- Большинство ошибок может быть связано с установкой системы в каталог, отличный от **c:\agfl_2.2** (в этом случае необходимо внесение изменений в **setagfl.bat**)
- Для правильной работы системы необходимо пользоваться для распаковки архивов дистрибутива архиватором, поддерживающим длинные имена. Кроме того, оболочка, которой вы пользуетесь, также не должна обрезать длинные имена. В противном случае **.exe** файл не будет сгенерирован и компилятор выдаст ошибку вида:

ld: cannot find -liberty

В этом случае вы должны проверить формат имени библиотеки **libliberty.a** в каталоге **...\agfl\lib** и, при необходимости, другие имена.

Если компиляция парсера прошла успешно, но при его запуске выдается ошибка вида

Cannot find library glib-1.3.dll

То это означает, что произошло усечение имен файлов и вы должны заново переписать файл **glib-1.3.dll** из файлов дистрибутива (**..\agfl\glib\glib-1.3.dll**) в каталог **c:\agfl\bin**

Со всеми вопросами обращайтесь:

Ulysses@io8176.spb.edu (Илья Опарин)

Или напрямую к разработчикам **AGFL 2.0 parser generator system for Windows:**

agfl@cs.kun.nl