

# Automatic Recording for KiSS Hard Disk Recorders

## Table of contents

1 Changelog.....	2
1.1 21 November 2006.....	2
1.2 19 November 2006.....	2
1.3 17 November 2006.....	2
1.4 7 September 2006.....	2
1.5 31 August 2006.....	2
1.6 24 August 2006.....	2
1.7 13-20 August 2006.....	3
2 Overview.....	3
3 Downloading.....	3
4 Configuring the crawler.....	4
4.1 Crawler configuration crawler.xml.....	4
4.2 Program configuration.....	4
4.3 Notification configuration.....	5
5 Installing and running the crawler.....	5
5.1 Standalone application.....	5
5.2 Web application.....	6
5.3 Source distribution.....	6
5.4 General usage.....	6
6 Examples.....	6
7 Contributing.....	6

**Warning:**

KiSS makes regular updates to their site that sometimes require adaptations to the crawler. If it stops working, check out the most recent version here.

## 1. Changelog

### 1.1. 21 November 2006

- Corrected the `config.xml` again.
- Corrected errors in the documentation for the web application. It starts running at 19:00 and not at 5:00.

### 1.2. 19 November 2006

- Corrected the `config.xml` file to deal with changes in the login procedure.

### 1.3. 17 November 2006

- Corrected the packed distributions. The standalone distribution had an error in the scripts and was missing libraries

### 1.4. 7 September 2006

- KiSS modified the login procedure. It is now working again.
- Generalized the startup scripts. They should now be insensitive to the specific libraries used.

### 1.5. 31 August 2006

- Added windows bat file for running the crawler under windows. Very add-hoc, will be generalized.

### 1.6. 24 August 2006

- The crawler now uses desktop login for crawling. Also, it is much more efficient since it no longer needs to crawl the individual programs. This is because the channel page includes descriptions of programs in javascript popups which can be used by the crawler. The result is a significant reduction of the load on the KiSS EPG site. Also, the delay between requests has been increased to further reduce load on the KiSS EPG site.
- The crawler now crawls programs for tomorrow instead of for today.
- The web based crawler is configured to run only between 7pm and 12pm. It used to run at

5am.

## 1.7. 13-20 August 2006

There were several changes to the login procedure, requiring modifications to the crawler.

- The crawler now uses the 'Referer' header field correctly at login.
- KiSS now uses hidden form fields in their login process which are now also handled correctly by the crawler.

## 2. Overview

In 2005, [KiSS](#) introduced the ability to schedule recordings on KiSS hard disk recorder (such as the DP-558) through a web site on the internet. When a new recording is scheduled through the web site, the KiSS recorder finds out about this new recording by polling a server on the internet. This is a really cool feature since it basically allows programming the recorder when away from home.

After using this feature for some time, I started noticing regular patterns. Often you are looking for the same programs and for certain types of programs. So, wouldn't it be nice to have a program do this work for you and automatically record programs and notify you of possibly interesting ones?

This is where the KiSS crawler comes in. This is a simple crawler which logs on to the KiSS electronic programme guide web site and gets programme information from there. Then based on that it automatically records programs for you or sends notifications about interesting ones.

In its current version, the crawler can be used in two ways:

- **standalone program:** A standalone program run from the command-line or as a scheduled task.
- **web application:** A web application running on a java application server. With this type of use, the crawler also features an automatic retry mechanism in case of failures, as well as a simple web interface.

## 3. Downloading

At this moment, no formal releases have been made and only the latest version can be downloaded.

The easy way to start is the [standalone program binary version](#) or using the [web application](#).

The latest source can be obtained from subversion with the URL <https://wamblee.org/svn/public/Utils>. The subversion repository allows read-only access to anyone.

The application was developed and tested on SuSE linux 10.1 with JBoss 4.0.4 application server. An application server or servlet container is only required for the web application. The crawler requires at least a Java Virtual Machine 1.5 or greater to run.

## 4. Configuring the crawler

The crawler comes with three configuration files:

- `crawler.xml`: basic crawler configuration tailored to the KiSS electronic programme guide.
- `programs.xml`: containing a description of which programs must be recorded and which programs are interesting.
- `org.wamblee.crawler.properties`: Containing a configuration

For the standalone program, all configuration files are in the `conf` directory. For the web application, the properties file is located in the `WEB-INF/classes` directory of the web application, and `crawler.xml` and `programs.xml` are located outside of the web application at a location configured in the properties file.

### 4.1. Crawler configuration `crawler.xml`

First of all, copy the `config.xml.example` file to `config.xml`. After that, edit the first entry of that file and replace `user` and `passwd` with your personal user id and password for the KiSS Electronic Programme Guide.

### 4.2. Program configuration

Interesting TV shows are described using `program` elements. Each `program` element contains one or more `match` elements that describe a condition that the interesting program must match.

Matching can be done on the following properties of a program:

Field name	Description
<code>name</code>	Program name
<code>description</code>	Program description
<code>channel</code>	Channel name

keywords	Keywords/classification of the program.
----------	---

The field to match is specified using the `field` attribute of the `match` element. If no field name is specified then the program name is matched. Matching is done by converting the field value to lowercase and then doing a perl-like regular expression match of the provided value. As a result, the content of the match element should be specified in lower case otherwise the pattern will never match. If multiple `match` elements are specified for a given `program` element, then all matches must apply for a program to be interesting.

Example patterns:

Pattern	Examples of matching field values
<code>the.*x.*files</code>	"The X files", "The X-Files: the making of"
<code>star trek</code>	"Star Trek Voyager", "Star Trek: The next generation"

It is possible that different programs cannot be recorded since they overlap. To deal with such conflicts, it is possible to specify a priority using the `priority` element. Higher values of the priority value mean a higher priority. If two programs have the same priority, then it is (more or less) unspecified which of the two will be recorded, but it will at least record one program. If no priority is specified, then the priority is 1 (one).

Since it is not always desirable to try to record every program that matches the criteria, it is also possible to generate notifications for interesting programs only without recording them. This is done by specifying the `action` element with the content `notify`. By default, the `action` is `record`. To make the mail reports more readable it is possible to also assign a category to a program for grouping interesting programs. This can be done using the `category` element. Note that if the `action` is `notify`, then the `priority` element is not used.

### 4.3. Notification configuration

Edit the configuration file `org.wamblee.crawler.properties`. The properties file is self-explanatory.

## 5. Installing and running the crawler

### 5.1. Standalone application

In the binary distribution, execute the `run` script for your operating system (`run.bat` for windows, and `run.sh` for unix).

## 5.2. Web application

After deploying the web application, navigate to the application in your browser (e.g. `http://localhost:8080/wamblee-crawler-kissweb`). The screen should show an overview of the last time it ran (if it ran before) as well as a button to run the crawler immediately. Also, the result of the last run can be viewed. The crawler will run automatically starting after 19:00, and will retry at 1 hour intervals in case of failure to retrieve programme information.

Since the crawler checks the status at 1 hour intervals it can run for the first time anytime between 19:00 and 20:00. This is done on purpose since it means that crawlers run by different people will not all start running simultaneously and is thus more friendly to the KiSS servers.

## 5.3. Source distribution

With the source code, build everything with maven2 as follows:

```
mvn -Dmaven.test.skip=true install
cd crawler
mvn package assembly:assembly
```

After this, locate the binary distribution in the `target` subdirectory of the `crawler` directory. Then proceed as for the binary distribution.

## 5.4. General usage

When the crawler runs, it retrieves the programs for tomorrow.

### Note:

If you deploy the web application today, it will run automatically on the next (!) day. This even holds if you deploy the application before the normal scheduled time.

## 6. Examples

The best example is in the distribution itself. It is my personal `programs.xml` file.

## 7. Contributing

You are always welcome to contribute. If you find a problem just tell me about it and if you have ideas am I always interested to hear about them.

If you are a programmer and have a fix for a bug, just send me a patch and if you are fanatic enough and have ideas, I can also give you write access to the repository.