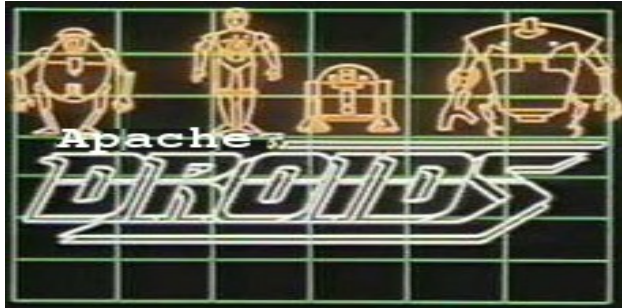


## Apache Droids



**Thorsten Scherler**

Sociedad Andaluza para el Desarrollo  
de la Sociedad de la Información S.A.U. (SADESI)

**[thorsten@apache.org](mailto:thorsten@apache.org)**

ApacheCon EU 2009, Amsterdam, 24 November 2008



## Incubator Disclaimer

Apache Droids is an effort undergoing incubation at the Apache Software Foundation (ASF), sponsored by the Lucene and HC project.

Incubation is required of all newly accepted projects until a further review indicates that the infrastructure, communications, and decision making process have stabilized in a manner consistent with other successful ASF projects.

While incubation status is not necessarily a reflection of the completeness or stability of the code, it does indicate that the project has yet to be fully endorsed by the ASF.



## Apache Droids

Droids is an intelligent standalone robot framework that allows to create and extend existing droids (robots).



## Apache Labs

- formed in October/November 2006
- place for innovation where committers of the foundation can experiment with new ideas
- provide the necessary resource to promote and maintain the innovative power within the Apache community without the burden of community building.
- proposed to the incubator as a seed for a new project when promoted





## droids@labs.apache.org

- formed in February 2007 out of personal curiosity and an usecase
- First version slimmed down Nutch
- Second version using Spring for configuration and API rewrite
- Third version community based rewrite (API, maven, http support, tika support)



## Apache Droids

A Droid can automatically seek out relevant online information based on the user's specifications and invoke custom handler on this information.



## why droids?

Droids (plural) is not designed for a special use case, it is a framework:

**Take what you need, do what you want.**



## Droids components

Queue, a queue is the data structure where the different tasks are waiting for service.





## droids components

Protocol, the protocol interface is a wrapper to hide the underlying implementation of the communication at protocol level.



## droids components

Parser -> Apache Tika, the parser component is just a wrapper for tika since it offers everything we need. No need to duplicate the effort. The Parser component parses different input types to SAX events.



## droids components

Handler, a handler is a component that uses the original stream and/or the parse (ContentHandler coming from Tika) and the url to invoke arbitrary business logic on the objects. Unless like the other components different handler can be applied on the stream/parse



## Getting started

- Before starting, you need to install
  - `jdk >= 1.6`
  - `maven >= 2.0.9`
- Download

```
mkdir droids
```

```
cd droids
```

```
svn co \
```

```
https://svn.apache.org/repos/asf/incubator/droids/trunk .
```

```
cd trunk
```





## Getting started

- Install

**mvn clean install**

- IDE

**mvn eclipse:clean eclipse:eclipse**

- IDE with sources add

**-Declipse.downloadSources=true**



## myDroid

A Droid (singular) is all about **ONE special usecase.**



## Types of Droids

- Runner
  - Predefined list of tasks, no task extraction
- Walker
  - Predefined entry point, task extraction via queue component
- Crawler
  - Task extraction in the parser stage



## helloCrawler

The helloCrawler is a wget style crawler. Meaning you go to a page extract the links and save the page afterward to the file system. The focus of the helloCrawler is this special usecase and to solve it hello uses different components.





## Examples

- SimpleRuntime
  - Adding max size
  - Changing the handler
  - Different taskmasters
  - Different parser



## Examples

- droids-spring

```
mvn package appassembler:assemble
```

```
cd target/appassembler/bin
```

```
chmod 755 cli
```

```
./cli hello http://localhost:8888/index.html
```



## Help droids

- Join the mailing list
- Get a wiki account
- Create plugins
- Report your usecase

**Become part of the community**



## droids resources

- <http://incubator.apache.org/droids>
- Mailing list
  - [droids-dev@incubator.apache.org](mailto:droids-dev@incubator.apache.org)
  - [droids-commits@incubator.apache.org](mailto:droids-commits@incubator.apache.org)
- <http://cwiki.apache.org/confluence/display/DROIDS/>
  - Installation
  - getStarted





# Thank you for your attention

**Thorsten Scherler**

Sociedad Andaluza para el Desarrollo  
de la Sociedad de la Información S.A.U. (SADESI)

**[thorsten@apache.org](mailto:thorsten@apache.org)**

ApacheCon US 2008, New Orleans, 07 November 2008

