

# Suivi du projet Video Parser

l'architecture du projet est tres simple :

la classe principale Vscmd lance un mini serveur web et integre la fonction process() qui va gerer les different appels de commandes en provenance de l'utilisateur via l'IHM web .

Le serveur est lance via un Thread , transformé en service ( daemon ) par la commande SetDaemon() . On n'oubliera pas de créer une methode run *runnable* dans le serveur et de lancer la methode start de la Thread associée .

```
//start web server as a detached thread
    // set it up as a daemon ( service ) listening on port 1012

    WebServer server = new WebServer(new
java.io.File("./gui/"),1012);
    Thread ThWebServer = new
Thread(null,server,"miniWeb");

    ThWebServer.setDaemon(true);
    ThWebServer.start();
```

le serveur écoute ensuite les requetes passées par le client : il recupere en effet le chemin demandé et l'analyse . Si il detecte une chaîne de type „GET“ , c'est a dire , url? cmd1=data1&cmd2=data2,...etc il la decompose et l'envoie a l'interpreteur de commandes , c'est a dire a la methode process de Vscmd

```
public static void Process(String path) {

    //The server sends us commands using the Get format
    //URL?COMMAND1=DATA1:COMMAND2=DATA2;...
    String[] Commands = (path.split("\\?")[1]).split("\\&");

    int len = Commands.length;

    for (int iCmd=0;iCmd<len;iCmd++)
    {
```

lorsque la commande eventuelle est executée ,le serveur effectue la requete pour l'url demandée .

Cet facon de procéder est en fait celle de tous les serveurs web qui gerent ce type de methode . L'avantage de travailler en 100% java est que l'on n'a pas a installer une architecture complexe avec serveur web dedié ( apache , tomcat , ... ) et un langage de script cgi ( php , perl , python ) ou des JSPs . Le fonctionnement est extremement simple .

Les requetes web se font via la classe getURL qui utilise simplement une methode (http) URLconnect derivée de URL .

Les données récupérées sont parsées et découpées avec la methode split puis stockées dans des

tableaux . Il serait preferable de creer un objet qui stocke ces informations .

```
VideoName[nV]=(VideoContent.split("var contentType = ")[1]).split(";")[0];
VideoLocation[nV]=(VideoContent.split("var contentHtmlFileName = ")
[1]).split(";")[0];
VideoChannel[nV]=(VideoContent.split("var channelFolder = ")[1]).split(";")
[0];
VideoImage[nV]=(VideoContent.split("var contentThumbnail = ")[1]).split(";")
[0];
```

D'une part les données sont stockees dans des fichiers en dur sur le disque en utilisant des templates HTML et la classe WriteHTML , d'autres part , elles sont stockées sur la base SQL via le mini requetteur sql ( tiny SQL ) et son driver jdbc intégré qui le connecte a une base DBF .

```
if(VideoText[iSqlTable].length()>135)
{
    TruncVT=VideoText[iSqlTable].substring(0,134);
    SQLcmd+="INSERT INTO "+Params.GetDBName()
+" (ID,NAME,LOCATION,CHANNEL,IMAGE,TEXT) VALUES ("+iSqlTable
+",\""+VideoName[iSqlTable]+"\", \""+VideoLocation[iSqlTable]
+"\", \""+VideoChannel[iSqlTable]+"\", \""+VideoImage[iSqlTable]+"\", \""+TruncVT
+"\" );\n";
}
else{
    SQLcmd+="INSERT INTO "+Params.GetDBName()
+" (ID,NAME,LOCATION,CHANNEL,IMAGE,TEXT) VALUES ("+iSqlTable
+",\""+VideoName[iSqlTable]+"\", \""+VideoLocation[iSqlTable]
+"\", \""+VideoChannel[iSqlTable]+"\", \""+VideoImage[iSqlTable]
+"\", \""+VideoText[iSqlTable]+"\" );\n";
}
SQLlog+="

La classe LogToFile et la methode Trace de cette dernière permettent de stocker les divers infos systemes dans une log dynamique HTML et une log texte .



Le systeme de recherche se trouve dans la classe SearchDatabase , il interroge la base SQL par requete simple sur une zone texte et spool le fichier de sortie .



La fonction WriteHTML construit le fichier de recherche en analysant le fichier de spool sql .



```
boolean slurp=true;

while(slurp)
{

mread=in.readLine();
if(mread==null){
    slurp=false;
}
}
```


```

```

else{

    //get the fields
    id = mread.split(tab)[0];
    mread=mread.replaceFirst(id,"");
    while (mread.startsWith(tab))
    {
        mread=mread.replaceFirst(tab, "");
    }
    name = mread.split(tab)[0];
    mread=mread.replaceFirst(name,"");
    while (mread.startsWith(tab))
    {
        mread=mread.replaceFirst(tab, "");
    }
}

```

Il eut ete préférable d'utiliser l'API de tiny sql pour effectuer des requetes directes mais en définitive , le résultat est le même .

On utilise une fonction de reload en javascript pour forcer la fenetre de terminal a se rafraichir et ainsi permettre une interaction en sortie avec l'utilisateur .

La page de recherche donne des resultats satisfaisants et malgres la limittion des composants , on peut sans problemes recuperer les quelques 4000 informations sur les videos du site break.com

d'autres methodes de scan restent a effectuer , mais elles donneraient sans doute le meme resultat . Une extension interessante serait de scanner les metadatas videos sur le web et de ne pas se limiter a un site en particulier .

### **Parametrages :**

- **nombre de videos par page**
- **nombre de requetes**
- **port du serveur**
- **etc ...**

### **fonctions Params.get/Set value**

**etc ...**

http://localhost:1012/ - Windows Internet Explorer

http://localhost:1012/

Google

Go

Bookmarks

48 blocked

Check

AutoLink

AutoFill

Send to

Settings

http://localhost:1012/

http://localhost:1012/

<< a very SIMPLE VIDEO PARSER 1.0

Listeof Videos

441 vidéos trouvées --- , page 9 on 23

<< Previous Page -- Next Page >>

0...1...2...3...4...5...6...7...8...9...10...11...12...13...14...15...16...17...18...19...20...21...22...

#160

|                      |   |
|----------------------|---|
| <b>_TITLE_</b>       | „Indonesian Plane Crash   |
| <b>_LOCATION_</b>    | <a href="http://www.break.com/index/indonesian_plane_crash.html">http://www.break.com/index/indonesian_plane_crash.html</a> |
| <b>_DESCRIPTION_</b> | <i>moments after the crash catching people escaping the wreckage.</i>   |
| <b>_CONTENT_</b>     |    |

#161

Fertig

Internet

VideoParserDoc.doc - OpenOffice.org Writer

http://localhost:1012/ - Windows Internet Explorer

http://localhost:1012/

Google

Go

Bookmarks

48 blocked

Check

AutoLink

AutoFill

Send to

Settings

http://localhost:1012/

http://localhost:1012/

his face into ...

|                  |   |
|------------------|---|
| <b>_CONTENT_</b> |  |
|------------------|---|

#123

|                      |   |
|----------------------|---|
| <b>_TITLE_</b>       | „Round And Round We Go  |
| <b>_LOCATION_</b>    | <a href="http://www.break.com/index/round_and_round_we_go.html">http://www.break.com/index/round_and_round_we_go.html</a> |
| <b>_DESCRIPTION_</b> | <i>e got tossed like a rag doll.</i>  |
| <b>_CONTENT_</b>     |                                        |

#124

|                   |   |
|-------------------|---|
| <b>_TITLE_</b>    | „Car Launched By Road Block   |
| <b>_LOCATION_</b> | <a href="http://www.break.com/index/car_launched_by_roadblock.html">http://www.break.com/index/car_launched_by_roadblock.html</a> |

Fertig

Internet

100%

http://localhost:1012/ - Windows Internet Explorer

http://localhost:1012/

VideoSearch:process START\_FIND\_DATABASE=START\_FIND\_DATABASE

## SIMPLE VIDEO PARSER 1.0

[Setting---Log](#)

### Scanner La Cible

|  |  |
|--|--|
| Informations                                       | 1 Enregistrements en base                        |
| <a href="#">consulter la liste TOTALE archivée</a> | <input type="button" value="START_LIVE_SEARCH"/> |

### rechercher une vidéo dans la base de donnée

|   |  |
|---|--|
| Mot-clef :                                    | <input type="text"/>                                     |
| <input type="button" value="START_FIND_DAT"/> | Soumettre une requete SQL a la base <input type="text"/> |
|   | <a href="#">resultat de la recherche</a>                 |

Fertig

http://localhost:1012/ - Windows Internet Explorer

http://localhost:1012/

VideoSearch:process START\_FIND\_DATABASE=START\_FIND\_DATABASE

## SIMPLE VIDEO PARSER 1.0

[Setting---Log](#)

### Scanner La Cible

|  |  |
|--|--|
| Informations                                       | 1 Enregistrements en base                        |
| <a href="#">consulter la liste TOTALE archivée</a> | <input type="button" value="START_LIVE_SEARCH"/> |

### rechercher une vidéo dans la base de donnée

|   |  |
|---|--|
| Mot-clef :                                    | dude <input type="text"/>                                |
| <input type="button" value="START_FIND_DAT"/> | Soumettre une requete SQL a la base <input type="text"/> |
|   | <a href="#">resultat de la recherche</a>                 |

http://localhost:1012/index.vdp

resultats de Recherche sur „dude“

The screenshot shows a Windows Internet Explorer browser window with the address bar set to `http://localhost:1012/`. The page content displays search results for the term "dude". At the top, it says "VideoSearch:process START\_FIND\_DATABASE=START\_FIND\_DATABASE". Below this, there are four search results, each consisting of a small thumbnail image, a title, a short description, and a URL:

- #88...Lion Doesnt Like Visitor**  
*Some **dude** goes to the zoo with his girlfriend and the lion behind the glass for some reason cant stand the **dude**.*  
[http://break.com/index/lion\\_doesnt\\_like\\_visitors.html](http://break.com/index/lion_doesnt_like_visitors.html)
- #138... Spotter Gets Kicked Across Gymnasium**  
*hard that the **dude** goes flying across the mat.*  
[http://break.com/index/spotter\\_gets\\_kicked\\_across\\_gymnasium.html](http://break.com/index/spotter_gets_kicked_across_gymnasium.html)
- #142... Slingshot From Point Blank**  
*Another group of friends grab a water ballon and sling shout launcher and nail some **dude** from about 10 feet.*  
[http://break.com/index/slingshot\\_from\\_point\\_blank.html](http://break.com/index/slingshot_from_point_blank.html)

The browser's taskbar at the bottom shows the word "Fertig" on the left and system icons for Internet and a 100% zoom level on the right.