**Files: extensions and MIME[[1]](#footnote-1) type**

**Object Oriented Programming & Java project**

### Goals

The OOP - Java module project makes it possible to implement the main elements of the content of the module as part of the design of a small application. The work is to be done in pairs (exceptionally alone, for special situations[[2]](#footnote-2)).

OOP - Java模块项目可以将模块内容的主要元素实现为小型应用程序设计的一部分。这项工作将成对完成（特殊情况下，特殊情况下单独完成）。[[3]](#footnote-3)

### Context of the project

A common belief is that the extension of a file determines its type: as long as the effective type of a file and its extension are consistent, this statement is true, for against as soon as the extension is no longer consistent with the content, it is necessary to rely on the MIME type to evaluate a file (this is generally the case for content exchanges multimedia via mail or the way files are processed by browsers when visit a web page). Under Linux, the command "file" in command line allows to know the MIME type of a file.

一个普遍的看法是，文件的扩展名决定了它的类型：只要文件的有效类型与其扩展名一致，这种说法就是正确的，因为一旦扩展名不再与内容一致，就有必要依靠MIME类型来评估文件（这通常是通过邮件进行内容交换多媒体或文件处理方式的情况。通过浏览器访问网页）。在Linux下，命令行中的命令“file”允许知道文件的MIME类型。

The objective of the project is to create a small application allowing to detect anomalies on data files[[4]](#footnote-4) of the user, so either to check a specific file (detection of empty files, detection of inconsistency between extension and MIME type, in-depth analysis of content of a file), or to explore and analyze a set of files from a folder.

该项目的目标是创建一个小型应用程序，允许检测用户数据文件上的异常，因此可以检查特定文件（检测空文件，检测扩展名和MIME类型之间的不一致，深入分析文件的内容），或者探索和分析文件夹中的一组文件。[[5]](#footnote-5)

The main actions of your software are:

软件的主要操作是：

* checking a machine file (based on the extension and MIME type of the file)

检查计算机文件（基于文件的扩展名和 MIME 类型）

* the complete exploration (a folder and all of its sub-folders) of a directory with verification of each file.

目录的完整浏览（文件夹及其所有子文件夹），并验证每个文件。

* The validation of the extension and of the MIME type on the basis of a list of signatures to create (example: a file with extension ".html" and MIME type "text / html" should to start[[6]](#footnote-6) with "<! DOCTYPE html>"; a script file with extension ".sh" and type MIME "application / x-shellscript" or "application / x-shell" should start[[7]](#footnote-7) per "#! / Bin /" (for example #! / Bin / bash).

根据要创建的签名列表验证扩展名和 MIME 类型（例如：扩展名为“.html”和 MIME 类型为“text / html”的文件应以“<！DOCTYPE html>“;扩展名为“.sh”并键入MIME“应用程序/ x-shellscript”或“应用程序/ x-shell”的脚本文件应按“#！/ Bin /“（例如 #！ / Bin / bash）。[[8]](#footnote-8)[[9]](#footnote-9)

* For certain types of files, checking their integrity: principle: the program can start the decompression of a compressed file (zip type for example) in order to ensure that it is indeed a zip file, similarly, the files of office suites (docx, xlsx, pptx / odt, ods, odp) are actually zip files that can be checked from the same way. For images, we can retrieve the dimensions to ensure that it is indeed an image, etc.

对于某些类型的文件，检查它们的完整性：原则：程序可以开始解压缩压缩文件（例如zip类型），以确保它确实是一个zip文件，同样，办公套件（docx，xlsx，pptx / odt，ods，odp）的文件实际上是可以从相同方式检查的zip文件。对于图像，我们可以检索尺寸以确保它确实是图像等。

Files of size 0 (zero) bytes should be reported to the user, likewise, your program will have to detect if the extension of a file does not correspond to its MIME type; finally Yes the content of a file does not correspond to the information detected a priori ("extension and type MIME ”), the program should detect it and inform the user.

应向用户报告大小为 0（零）字节的文件，同样，您的程序必须检测文件的扩展名是否与其 MIME 类型不对应;最后是的，文件的内容与先验检测到的信息不对应（“扩展名和类型 MIME ”），程序应检测到它并通知用户。

In console (terminal) mode "CLI[[10]](#footnote-10) ": the parameters expected on the command line are: the name of the file to be processed or the name of the folder to explore: you will use a parameter additional to specify the type of entry ("-f" for a file, "-d" in the case of a directory). The program must display the result of its analysis directly in the console. If no parameter is specified (or option "-h" for "help"), the program displays the help and the possible options. An additional option "-s" allows to save the result of an analysis in a file to be specified by the user.

在控制台（终端）模式“CLI”中：命令行上预期的参数是：要处理的文件的名称或要浏览的文件夹的名称：您将使用附加参数来指定条目类型（“-f”表示文件，“-d”表示目录）。程序必须直接在控制台中显示其分析结果。如果未指定参数（或选项“-h”表示“帮助”），程序将显示帮助和可能的选项。附加选项“-s”允许将分析结果保存在用户指定的文件中。[[11]](#footnote-11)

In GUI graphic mode[[12]](#footnote-12) ": exploring any tree structure[[13]](#footnote-13) files will list in the graphical interface all the files (but not the directories and sub-directories) and their locations. For each file, the level of compliance between the extension, MIME type and content will be indicated (problematic files will be visually identifiable).

在GUI图形模式下“：浏览任何树结构文件将在图形界面中列出所有文件（但不包括目录和子目录）及其位置。对于每个文件，将指示扩展名，MIME类型和内容之间的合规性级别（有问题的文件将在视觉上可识别）。[[14]](#footnote-14)[[15]](#footnote-15)

The signature database, ie: extensions, associated MIME types and identification content if there

if applicable, will be stored as a CSV file.

签名数据库，即：扩展名、关联的MIME类型和标识内容（如果有）

如果适用，将存储为 CSV 文件。

Some execution scenarios (fictitious examples of launching your 2 programs[[16]](#footnote-16)):

一些执行场景（启动 2 个程序的虚构示例）：[[17]](#footnote-17)

java -jar cli.jar

java -jar cli.jar -d.

java -jar cli.jar -f test.html

java -jar cli.jar -d. -s analysis

java -jar gui.jar

***Explanation:*** the first 4 commands concern console mode (also called mode terminal or command window); the last command launches the graphical interface.

前 4 个命令涉及控制台模式（也称为模式）终端或命令窗口）;最后一个命令启动图形界面。

* the first line displays the modes of use of your software in console mode (ie possible options and their role);

第一行显示软件在控制台模式下的使用模式（即可能的选项及其角色）

* the second line lists and analyzes all the files **from** the specified folder ( *“-d” = directory* ) [here from the current folder (".")] by browsing the whole tree structure sub-folders, in console mode;

第二行列出并分析指定文件夹（“*-d”=目录*）**[**此处来自当前文件夹（“.”）]的所有文件，方法是在控制台模式下浏览整个树结构子文件夹

* the third line takes as input the file “test.html” ( *“-f” = file* ) and displays on the screen the result of its verification, in console mode;

第三行将文件“test.html”（*“-f”=文件*）作为输入，并在控制台模式下在屏幕上显示其验证结果

* the fourth line takes as input the current folder ". "And save the result of the verification in an “analysis” file ( *“-s” = save* ) in console mode;

第四行将当前文件夹“.并将验证结果保存在控制台模式下的“分析”文件中（*“-s”=保存*）

* the last line corresponds to the launch of the graphical interface.

最后一行对应于图形界面的启动。

***NB:*** the format for saving the results of an analysis is left to your initiative, but it is

necessary to be able to exploit this result outside of your software: special attention is

therefore, to bring to this choice.

***注意：*** 保存分析结果的格式由您主动决定，但它是为了能够在软件之外利用此结果所必需的：特别注意的是因此，带来这个选择。

In graphical mode, it is requested to be able to keep a serialized version of the analysis between 2 program executions in order to be able to dissociate the analysis (possibly long), from display of identified problems. When the software is launched in graphic mode, it is therefore the last result obtained which is displayed, with the possibility of launching a new analysis.

在图形模式下，要求能够在 2 个程序执行之间保留分析的序列化版本，以便能够将分析（可能很长）与已识别问题的显示分离。当软件以图形模式启动时，显示的是最后获得的结果，并有可能启动新的分析

Possible extensions: you can plan improvements to your solution but **only** if everything else is complete. You will therefore favor the quality of the realization over the quantity of features.

可能的扩展：您可以计划对解决方案的改进，但**前提是** 其他所有内容都已完成。因此，您将更喜欢实现的质量而不是功能的数量

### Planning and advice for the presentation

***Planning***

* Identify the main sub-tasks of the project to be carried out, their level of priority, the distribution of roles within the pair as well as the corresponding schedule for each task (period and duration). You can create a small summary table or (better), create a GANTT diagram[[18]](#footnote-18) of your project. In all cases, the table or diagram of Gantt must be returned in png format (capture or export) before **(Date & Time to be determined)**: **1 point**

确定要执行的项目的主要子任务，其优先级，配对中的角色分配以及每个任务的相应时间表（期间和持续时间）。您可以创建一个小的汇总表，或者（更好的是）创建项目的甘特图。在所有情况下，甘特图的表格或图表必须在（日期和时间待定）之前以png格式（捕获或导出）返回[[19]](#footnote-19)：**1分**

* progress points in December (compulsory presence): **3 points** (for information, one level of achievement of approximately 50% is expected at the last progress point)

12月进度点（强制存在）：**3分**（供参考， 预计在最后一个进度点达到约50%的一个水平）

* Defense (5 min) and demonstration (5 min): **6 points** : **(Date & Time to be determined).**

答辩（5分钟）和演示（5分钟）：**6分**： **（日期和时间待定12月）**

***Defense: 5 slides maximum, 5 minutes maximum***

* the title slide will present the binomial, the context, the subject. (ie: the cover page will be compact),
* the other slides must present the specific features of the project team’s work, therefore no "obvious" information (ex. details of the subject, personal progress, ...) should be mentioned,
* the concluding slide will highlight the level of completion of the project (points treated and untreated specifications and extensions if any),
* you will need to provide a pdf version of your slideshow in case.

The following information should be present: the distribution of tasks, the main design elements.

1. 标题幻灯片将呈现二项式、上下文和主题。（即：封面将紧凑），
2. 其他幻灯片必须呈现项目团队工作的具体特征，因此不应提及“明显”的信息（例如主题的详细信息、个人进展等），
3. 最后的幻灯片将突出项目的完成水平（处理和未处理的要点规范和扩展，如果有的话），
4. 您需要提供幻灯片的 PDF 版本以防万一。

应提供以下信息：任务分配，主要设计元素。

***NB:*** to be avoided ABSOLUTELY: unreadable UML class diagrams (too loaded, ...), programs (Java code), screenshots (since there is also a demonstration), the list of tools (eg Eclipse, etc.), ...

***注意：*** 绝对要避免：不可读的UML类图（加载太多，...），程序（Java代码），屏幕截图（因为也有演示），工具列表（例如Eclipse等），...

***Important:*** you must have your portable machine turned on, ready with all of the software already launched BEFORE entering the room. Your laptop should have a VGA port or you will need to provide a VGA adapter that matches your situation.

***重要提示：在*** 进入房间之前，您必须打开便携式机器，准备好所有软件。您的笔记本电脑应具有 VGA 端口，或者您需要提供符合您情况的 VGA 适配器。

***Important:*** you will ensure an equitable distribution of your speaking time within the pair for both presentation and demonstration.

***重要提示：*** 您将确保在两人中公平分配您的演讲时间，以进行演示和演示。

***Demonstration: 5 minutes maximum***

***演示：最多5分钟***

1. console mode,

2. graphic mode.

1.控制台模式，

2.图形模式。

You will make sure to plan a scenario for the demonstration.

您将确保为演示计划一个方案。

### Expected results and evaluation criteria

Completeness and quality of the project: **7 points** ("readme.txt" file, java code, javadoc, files

jar). Deliverables to be deposited on the educational platform before **(Date/Time to be determined)**.

项目的完整性和质量： **7 分**（“自述文件.txt”文件、java 代码、javadoc、文件罐子）。在（**日/时间待定）之前存放**在教育平台上的可交付成果。

* "readme.txt" file containing the names / first names / TD group / of the members of the project as well as specific useful information,
* project report (minimum 5 pages, maximum 10 pages): **3 points** (content and form will be evaluated). The following 2 files are to be returned: 1) the word processing document ( **docx or odt** ), 2) the **pdf** version of your report
* sets of project source files (.java),
* javadoc,
* The 2 jar files in java 1.8 compatible version.

1. “自述文件.txt”文件包含项目成员的姓名/名字/TD组/以及 特定的有用信息，
2. 项目报告（最少5页，最多10页）：**3分**（内容和形式将进行评估）。将返回以下 2 个文件：1） 文字处理文档（**docx 或 odt**），2） 报告的 **pdf** 版本
3. 项目源文件集 （.java），
4. javadoc，
5. Java 1.8兼容版本中的2个jar文件。

***Important :*** All files and sub-folders to be submitted must be placed in a directory unique with the 2 names of the binomial (in the form NOM1\_NOM2) to compress into one file in zip format which will be uploaded to the educational course platform *(moodle)*.

***重要提示：***所有要提交的文件和子文件夹必须放置在具有二项式 2 个名称（以 NOM1\_NOM2 形式）的唯一目录中，以压缩成一个 zip 格式的文件，该文件将上传到教育课程平台 *（moodle）。*

***Some indications for the realization***

***实现的一些迹象***

The objective of the project is to allow you to implement, as part of an achievement concrete, the concepts of OOP and Java discussed during the module. There is no need to want be exhaustive in dealing with the many situations present in the files handled. Of even, it is possible to use external libraries corresponding to your needs.

该项目的目标是允许您实现，作为具体成就的一部分，模块期间讨论的 OOP 和 Java 概念。没有必要详尽无遗地处理所处理文件中存在的许多情况。甚至，也可以使用符合您需求的外部库。

First part

To start the project, it is recommended to deal with a single case first file by relying first only on the extension, then on the MIME type and finally on a signature to be defined. Checking the integrity of certain files should only be covered in a second time.

要启动项目，建议首先处理单个案例文件，首先仅依赖扩展名，然后依赖 MIME 类型，最后依赖要定义的签名。检查某些文件的完整性应该只在第二次中介绍。

Second part

Then, you can generalize your solution to a set of files placed in a directory and its subdirectories.

然后，可以将解决方案通用化为放置在目录及其子目录中的一组文件。

***Some indications for the report***

This document written using a word processor (MS Word or free office for example) must

allow to provide a complete report and an assessment of your work and its outcome. In particular, you will place the UML class diagram for your application there. Information planning and distribution of tasks are expected. Explanations on specific aspects of your solution. You can add some representative screenshots but don't overdo it not. You will detail the level of achievement of your achievement with also a look criticize the strengths and weaknesses that you have identified.

使用文字处理器（例如MS Word或免费办公室）编写的文档必须允许提供完整的报告和对您的工作及其结果的评估。特别是，您将在此处放置应用程序的 UML 类图。预计会进行信息规划和任务分配。有关解决方案特定方面的说明。您可以添加一些具有代表性的屏幕截图，但不要过度。您将详细说明您的成就水平，并批评您确定的优势和劣势。

### Resources

* <https://www.iana.org/assignments/media-types/media-types.xhtml>
* <https://developer.mozilla.org/fr/docs/Web/HTTP/Basics_of_HTTP/MIME_Types>
* <https://developer.mozilla.org/fr/docs/Web/HTTP/Basics_of_HTTP/MIME_types/Common_types>
* <https://tools.ietf.org/html/rfc2045>

1. MIME: Multipurpose Internet Mail Extensions [↑](#footnote-ref-1)
2. Case of students in Terminal control, AJAC or odd number of students on the whole promotion by example [↑](#footnote-ref-2)
3. 学生在终端控制，AJAC或奇数学生的案例中，整个促销示例 [↑](#footnote-ref-3)
4. We will not consider the analysis of programs (of type .exe, .jar, .class, etc.). [↑](#footnote-ref-4)
5. 我们不会考虑程序分析（类型.exe、.jar、.class等）。 [↑](#footnote-ref-5)
6. After possibly comments of type “<! -… ->” [↑](#footnote-ref-6)
7. After possibly lines of comments preceded by the character “#” [↑](#footnote-ref-7)
8. 在可能键入“<！-... ->" [↑](#footnote-ref-8)
9. 在可能以字符“#”开头的注释行之后 [↑](#footnote-ref-9)
10. CLI: Command Line Interface [↑](#footnote-ref-10)
11. CLI：命令行界面 [↑](#footnote-ref-11)
12. GUI: Graphical User Interface [↑](#footnote-ref-12)
13. With exploration of all the sub-directories [↑](#footnote-ref-13)
14. 图形用户界面：图形用户界面 [↑](#footnote-ref-14)
15. 探索所有子目录 [↑](#footnote-ref-15)
16. cli: command line interface / gui: graphical user interface [↑](#footnote-ref-16)
17. CLI：命令行界面/GUI：图形用户界面 [↑](#footnote-ref-17)
18. cf. <https://www.ganttproject.biz/download/free> [↑](#footnote-ref-18)
19. 参看 <https://www.ganttproject.biz/download/free> [↑](#footnote-ref-19)