Graph Extractor Documentation

Abstract

Graph Extractor is a graph extracting tool from scanned PDF files which runs on Ubuntu Linux distributions. This document contains the user guide describing how to install and run the software and software documentation describing the algorithms, modules of the code/architectural diagram and test plan.

Preprint submitted to OpenSoft 2016 judges *March 10, 2016*

Contents

1 Installation Instructions **3**
- 1.1 Installing the GUI application . . . 3
  - 1.1.1 Requirements . . . 3
  - 1.1.2 Installation . . . 3
- 1.2 Installing the console application . . . 4
  - 1.2.1 Requirements . . . 4
  - 1.2.2 Installation . . . 5

2 User Manual **5**
- 2.1 Running the GUI application . . . 5
- 2.2 Processing a scanned PDF . . . 5
- 2.3 Running the console application . . . 6

3 Algorithm Description **6**

4 Software Architecture **6**

5 Test Plan **6**

1. Installation Instructions

Graph extractor can be installed as an application with **graphical user interface** and as a **console application**.

1.1. Installing the GUI application

Note that installing the required libraries for the GUI application takes much longer time (typically 30 to 45 minutes) as compared to the console application. In case of any error in installation please switch to installing the console application.

1.1.1. Requirements

- **Operating System**: Ubuntu
- **Architecture Type**: 64-bit (preferably)
- **Dependencies**:
  - build-essential
  - libopencv-dev
  - tesseract-ocr
  - tesseract-ocr-eng
  - tesseract-ocr-equ
  - unpaper
  - imagemagick
  - qt5-default
  - qt5-qmake
  - qtbase5-dev-tools
  - libqt5webengine5-dev

1.1.2. Installation

- Open the terminal and set the working directory to the submission folder.

```
:~/Submission$ ls
Console  Console_Installer  Documentation.pdf  GUI  GUI_Installer
```

- Change directory to the GUI_Installer directory.

```
:~/Submission$ cd GUI_Installer/
:~/Submission/GUI_Installer$ ls
dependencies.sh  graph_extractor_install.sh
```

- Make the file **graph_extractor_install.sh** executable if not already.

```
:~/Submission/GUI_Installer$ chmod +x graph_extractor_install.sh
```

- Run the installer with root privileges and export environment variables as follows:

```
:~/Submission/GUI_Installer$ sudo -E ./graph_extractor_install.sh
--- Installing Dependencies
...
...
make[1]: Leaving directory '.../Submission/build/graphextractor'
--- Installation Complete
```

Note that if you are behind a proxy server, setting the $http_proxy environment variable is a must. Please set the http proxy for the system.

- If you see –**Installation Complete** at the end without any errors, the installation is complete. If not, make sure to handle the errors as suggested by the installer. You will see a new folder named **build** in the parent directory which contains the installed software.

```
:~/Submission/GUI_Installer$ cd ../
:~/Submission$ ls
build  Console  Console_Installer  Documentation.pdf  GUI
GUI_Installer
:~/Submission$ cd build
:~/Submission/build$ ls
graphextractor  Graph_Extractor_run.sh  qpdflib
```

1.2. Installing the console application

1.2.1. Requirements

- **Operating System**: Ubuntu
- **Architecture Type**: 64-bit (preferably)
- **Dependencies**:
  - build-essential
  - libopencv-dev
  - tesseract-ocr
  - tesseract-ocr-eng
  - tesseract-ocr-equ
  - unpaper
  - imagemagick

1.2.2. Installation

2. User Manual

2.1. Running the GUI application

To run the GUI application:

- Change the directory to **build** directory created by installation.

```
:~/Submission/build$ ls
graphextractor  Graph_Extractor_run.sh  qpdflib
```

- Run the application as follows:

```
:~/Submission/build$ ./Graph_Extractor_run.sh
```

The application will execute.

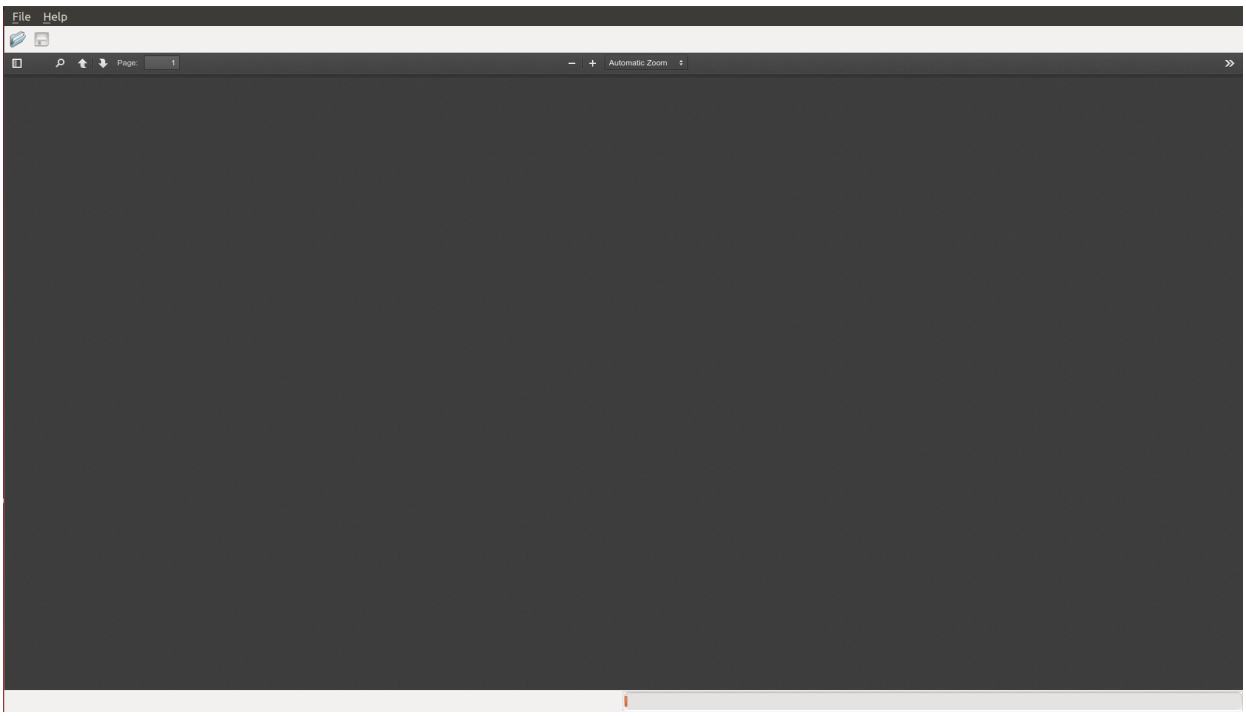

Figure 1: Graph Extractor: Starting the application

2.2. Processing a scanned PDF

To extract graphs and export them to tables:

- Click the open button or press Ctrl+O or go to File→Open.

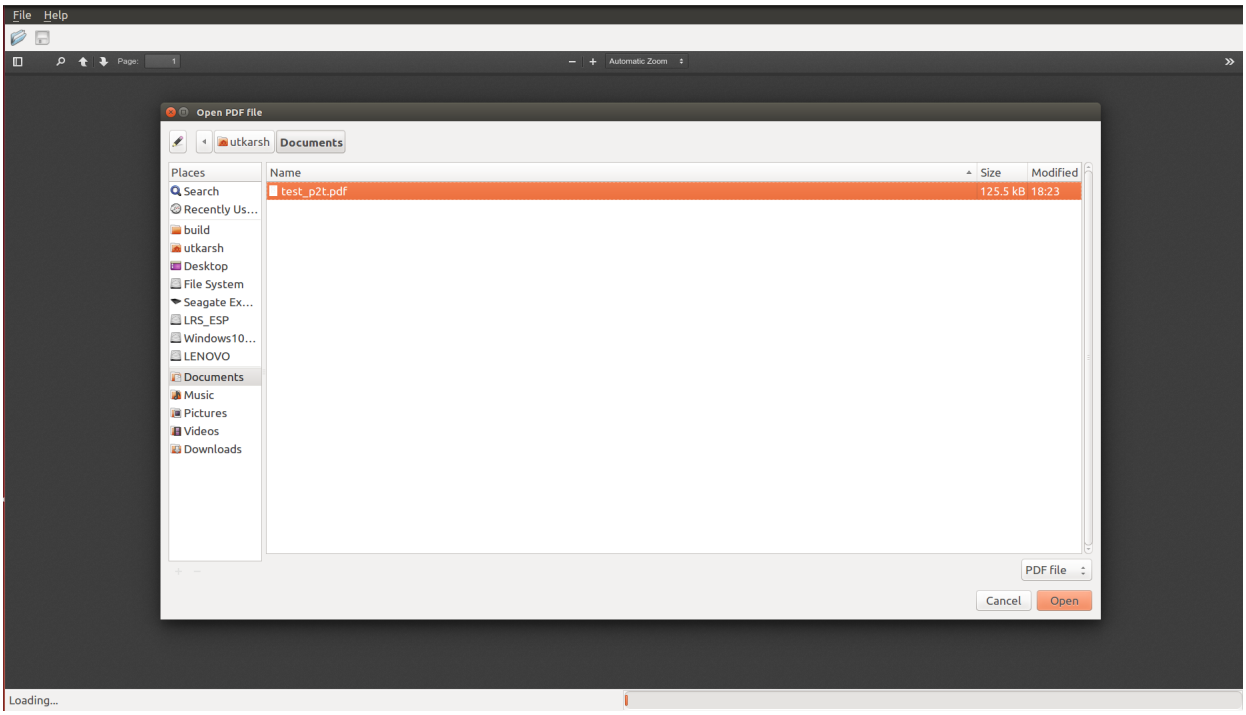

Figure 2: Graph Extractor: Opening a scanned PDF

- Select a PDF file and click Open to begin the graph extraction.

2.3. Running the console application

3. Algorithm Description

4. Software Architecture

5. Test Plan

References