

A Mid-Term Progress Report
On
Text Extraction From Web Pages

**Submitted in partial fulfillment of the requirements for
the award of the degree of**

Bachelor of Technology
(Computer Science and Engineering)



Submitted By:
Navdeep Singh
1243678

**Guru Nanak Dev Engineering College
Ludhiana - 141006**

1	Introduction	1
1.1	Overview	1
1.2	Objectives	1
2	System Requirements	2
2.1	System Requirements	2
2.1.1	Hardware Requirements	2
2.1.2	Software Requirements	2
3	Software Requirement Analysis	3
3.1	Software Requirement Analysis	3
3.1.1	Functional Requirements	3
3.1.2	Non functional requirements	4
4	Software Design	5
4.1	Product Perspective	5
4.2	User Characteristics	5
4.3	System Design	6
4.4	Technologies Used	6
4.4.1	C++	6
4.4.2	Introduction To Qt	7
4.4.3	Introduction to Github	7
4.4.4	What is Git?	8
4.4.5	Installation of Git	8
4.4.6	Various Git Commands	8
	Create Repositories	9
	Make Changes	9

1.1 Overview

Automated Building Drawing is a project for creating two-dimensional drawings (front-view, top-view, side-view etc.) from a three-dimensional model.

The main purpose or objective of the project is to make it usable even by the layman. The main target users are the Civil Engineers who want their plans to be printed on the sheets. As of now, they have to create the drawings separately with different views in any CAD software and the 3D model separately. So to automate converting a particular three-dimensional model to the print-ready drawings (with different views), this project will be beneficial. The interface should be easy to use and pretty intuitive. Because the interface is a thing that makes user experience better and to make the user use it.

The Drawing module allows you to put your 3D work on paper. That is, to put views of your models in a 2D window and to insert that window in a drawing, for example a sheet with a border, a title and your logo and finally print that sheet. The drawing may consist of different views like top, front, side and orthographic views.

1.2 Objectives

- To put views of your models in a 2D window and to insert that window in a drawing,.
- Automatically creates orthographic views of an object.

CHAPTER 2

SYSTEM REQUIREMENTS

2.1 System Requirements

2.1.1 Hardware Requirements

- Operating System: ubuntu 12.04 or windows 7
- Processor Speed: 512KHz or more
- RAM: Minimum 256MB

2.1.2 Software Requirements

- Software: Xampp or lampp(in case of ubuntu)
- Programming Language: C++, Python, Qt
- Database: MySQL or some Object-oriented database

3.1 Software Requirement Analysis

A Software Requirements Analysis for a software system is a complete description of the behaviour of a system to be developed. It includes a set of use cases that describe all the interactions the users will have with the software. In addition to use cases, the SRS also contains non-functional requirements. Non-functional requirements are requirements which impose constraints on the design or implementation.

- **General Description:** The Drawing module allows you to put your 3D work on paper. That is, to put views of your models in a 2D window and to insert that window in a drawing, for example a sheet with a border, a title and your logo and finally print that sheet. The drawing may consist of different views like top, front, side and orthographic views. It is developed using FreeCad, Qt and C++.
- **Users of the System:** The main target users are the Civil Engineers who want their plans to be printed on the sheets. As of now, they have to create the drawings separately with different views in any CAD software and the 3D model separately. So to automate converting a particular three-dimensional model to the print-ready drawings (with different views), this project will be beneficial. So to decrease the efforts, time and cost, it would be really beneficial.

3.1.1 Functional Requirements

- **Specific Requirements:** This phase covers the whole requirements for the system. After understanding the system we need the input data to the system then we watch the output and determine whether the output from the system is according to our requirements or not. So what we have to input and then what we'll get as output is given in this phase. This phase also describes the software and non-function requirements of the system.
- **Input Requirements of the System**
 1. Three Dimensional Model.
 2. Dimensions Of Different Views.

3. Scaling Parameters.
4. Name Of Particular Object.

- **Output Requirements of the System**

1. 2Dimensional Model.
2. Generation Of Front View.
3. Generation Of Top View.
4. Generation Of Side View.

- **Special User Requirements**

1. Exporting the 2D model in pdf or svg format.

- **Software Requirements**

1. Programming language: C++
2. Framework: Qt
3. Documentation: Doxygen 1.8.3
4. Text Editor: Gedit, Notepad++, Sublime
5. Operating System: Ubuntu 12.04 or up
6. Web Server: Apache 2.4

3.1.2 Non functional requirements

1. Scalability: System should be able to handle a number of users. For e.g., handling around hundred users at the same time.
2. Usability: Simple user interfaces that a layman can understand.
3. Speed: Speed of the system should be responsive i.e. Response to a particular action should be available in short period of time.

4.1 Product Perspective

Drawing automation refers to the varied computer machinery and software used to digitally create, collect, store, manipulate, and relay information needed for accomplishing basic drawing tasks. Creating two dimensional models, electronic transfer, and the management of automated models comprise the basic activities of an drawing automation system. Drawing automation helps in optimizing or automating existing drawing procedures.

Automated Building Drawing is a project for creating two-dimensional drawings (front-view, top-view, side-view etc.) from a three-dimensional model. The main purpose or objective of the project is to make it usable even by the layman. The main target users are the Civil Engineers who want their plans to be printed on the sheets. As of now, they have to create the drawings separately with different views in any CAD software and the 3D model separately. So to automate converting a particular three-dimensional model to the print- ready drawings (with different views), this project will be beneficial. The interface should be easy to use and pretty intuitive. Because the interface is a thing that makes user experience better and to make the user use it.

4.2 User Characteristics

The objective of this system is to provide an efficient and effective service to the students and teachers or any other person related to drawing directly or indirectly. It is aimed to encourage people to use computers and internet instead of paper and pen for their daily work. The Drawing module allows user to put your 3D work on paper. That is, to put views of your models in a 2D window and to insert that window in a drawing, for example a sheet with a border, a title and your logo and finally print that sheet. The drawing may consist of different views like top, front, side and orthographic views.

4.3 System Design

Systems design is the process or art of defining the architecture, components, modules, interfaces, and data for a system to satisfy specified requirements. One could see it as the application of systems theory to product development. There is some overlap with the disciplines of systems analysis, systems architecture and systems engineering.

- External design: External design consists of conceiving, planning out and specifying the externally observable characteristics of the software product. These characteristics include user displays or user interface forms and the report formats, external data sources and the functional characteristics, performance requirements etc. External design begins during the analysis phase and continues into the design phase.
- Logical design: The logical design of a system pertains to an abstract representation of the data flows, inputs and outputs of the system. This is often conducted via modeling, which involves a simplistic (and sometimes graphical) representation of an actual system. In the context of systems design, modeling can undertake the following forms, including:
 - Data flow diagrams
 - Entity Relationship Diagrams
- Physical design: The physical design relates to the actual input and output processes of the system. This is laid down in terms of how data is input into a system, how it is verified/authenticated, how it is processed, and how it is displayed as output.

4.4 Technologies Used

4.4.1 C++

C++ is one of the most popular programming languages and is implemented on a wide variety of hardware and operating system platforms. As an efficient compiler to native code, its application domains include systems software, application software, device drivers, embedded software, high- performance server and client applications, and entertainment software such as video games. Several groups provide both free and proprietary C++ compiler software, including the GNU Project, Microsoft, Intel and Embarcadero Technologies. C++ has greatly influenced many other popular programming languages, most notably C# and Java. Other successful languages such as Objective- C use a very different syntax and approach to adding classes to C.

Bjarne Stroustrup began his work on C with Classes in 1979. The idea of creating a new language originated from Stroustrups experience in programming for his Ph.D. thesis. Stroustrup found that Simula had features that were very helpful for large software development, but the language was too slow for practical use, while BCPL was fast but too low-level to be suitable for large software development. When Stroustrup started working in AT&T Bell Labs, he had the problem of analyzing the UNIX kernel with respect to distributed computing. Remembering his Ph.D. experience, Stroustrup set out to enhance the C language with Simula-like features. C was chosen because it was general-purpose, fast, portable and widely used. Besides C and Simula, some other languages that inspired him were ALGOL 68, Ada, CLU and ML. At first, the class, derived class, strong type checking, inlining, and default argument features were added to C via Stroustrups C++ to

C compiler, Cfront. The first commercial implementation of C++ was released on 14 October 1985.

4.4.2 Introduction To Qt

Qt Creator is a complete IDE for creating applications with Qt Quick and the Qt application framework. Qt is designed for developing applications and user interfaces once and deploying them across several desktop and mobile operating systems. One of the major advantages of Qt Creator is that it allows a team of developers to share a project across different development platforms (Microsoft Windows, Mac OS X, and Linux) with a common tool for development and debugging. In addition, UI designers can join the team by using Qt Quick tools for creating fluid user interfaces in close cooperation with the developers. The main goal for Qt Creator is meeting the development needs of Qt Quick developers who are looking for simplicity, usability, productivity, extendibility and openness, while aiming to lower the barrier of entry for newcomers to Qt Quick and Qt. The key features of Qt Creator allow UI designers and developers to accomplish the following tasks:

- Get started with Qt Quick application development quickly and easily with examples, tutorials, and project wizards.
- Design application user interface with the integrated editor, Qt Quick Designer, or use graphics software to design the user interface and use scripts to export the design to Qt Quick Designer.
- Develop applications with the advanced code editor that provides new powerful features for completing code snippets, refactoring code, and viewing the element hierarchy of QML files.
- Build and deploy Qt Quick applications that target multiple desktop and mobile platforms, such as Microsoft Windows, Mac OS X, Linux, Symbian, MeeGo, and Maemo.
- Debug JavaScript functions and execute JavaScript expressions in the current context, and inspect QML at runtime to explore the object structure, debug animations, and inspect colors.
- Profile your Qt Quick applications with the QML Profiler. You can inspect binding evaluations, signal handling, and painting operations when running QML code. This is useful for identifying potential bottlenecks, especially in the evaluation of bindings.
- Deploy applications to mobile devices and create application installation packages for Symbian and Maemo devices that can be published in the Ovi Store and other channels.
- Easily access information with the integrated context-sensitive Qt Help system.
- It has different modes such as Welcome, edit debug, design, analyze and help

4.4.3 Introduction to Github

GitHub is a Git repository web-based hosting service which offers all of the functionality of Git as well as adding many of its own features. Unlike Git which is strictly a command-line tool, Github provides a web-based graphical interface and desktop as well as mobile integration. It also provides access control and several collaboration features such as wikis, task management, and bug tracking

and feature requests for every project.

GitHub offers both paid plans for private reposito to handle everything from small to very large projects with speed and efficiency. ositories, and free accounts, which are usually used to host open source software projects. As of 2014, Github reports having over 3.4 million users, making it the largest code host in the world.

GitHub has become such a staple amongst the open-source development community that many developers have begun considering it a replacement for a conventional resume and some employers require applications to provide a link to and have an active contributing GitHub account in order to qualify for a job.

4.4.4 What is Git?

Git is a distributed revision control and source code management (SCM) system with an emphasis on speed, data integrity, and support for distributed, non-linear workflows. Git was initially designed and developed by Linus Torvalds for Linux kernel development in 2005, and has since become the most widely adopted version control system for software development.

As with most other distributed revision control systems, and unlike most clientserver systems, every Git working directory is a full-fledged repository with complete history and full version-tracking capabilities, independent of network access or a central server. Like the Linux kernel, Git is free and open source software distributed under the terms of the GNU General Public License version 2 to handle everything from small to very large projects with speed and efficiency.

Git is easy to learn and has a tiny footprint with lightning fast performance. It outclasses SCM tools like Subversion, CVS, Perforce, and ClearCase with features like cheap local branching, convenient staging areas, and multiple workflows.

4.4.5 Installation of Git

Installation of git is a very easy process. The current git version is: 2.0.4. Type the commands in the terminal:

```
$ sudo apt-get update
```

```
$ sudo apt-get install git
```

This will install the git on your pc or laptop.

4.4.6 Various Git Commands

Git is the open source distributed version control system that facilitates GitHub activities on your laptop or desktop. The commonly used Git command line instructions are:-

Create Repositories

Start a new repository or obtain from an exiting URL

\$ git init [project-name]

Creates a new local repository with the specified name

\$ git clone [url]

Downloads a project and its entire version history

Make Changes

Review edits and craft a commit transaction

\$ git status

Lists all new or modified files to be committed

\$ git diff

Shows file differences not yet staged

\$ git add [file]

Snapshots the file in preparation for versioning

\$ git reset [file]

Unstages the file, but preserve its contents

\$ git commit -m "[descriptive message]"

Records file snapshots permanently in version history

BIBLIOGRAPHY

- [1] FreeCAD Code, <https://github.com/FreeCAD/FreeCAD>
- [2] FreeCAD forum, <http://forum.freecadweb.org>
- [3] Libdxf read write, <https://github.com/rvt/libdxfw>
- [4] Module, http://www.freecadweb.org/wiki/index.php?title=Drawing_Module
- [5] Wikipedia, <http://en.wikipedia.org/wiki/>