# CHAPTER 2

OPEN SOURCE CONCEPTS

Learning objectives

After learning this chapter the student will be able to:

- Understand about OSS/FOSS/FLOSS

- Cite different examples of OSS

- Describe open document format

- Understand character encoding in Indian languages

- Know about open type/true type/static/dynamic fonts

Puzzle [2]

Write the numbers 1 to 8 in the given circles so that no two numbers joined by a line differ by 1.
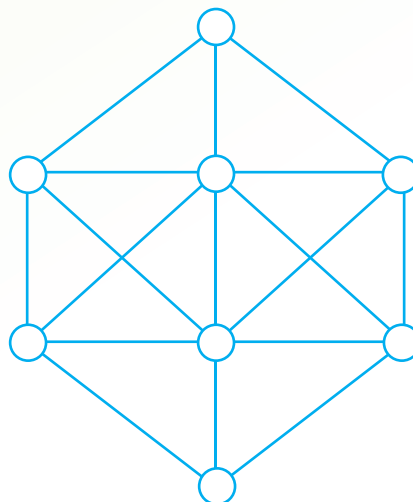
Introduction

Computers and internet have transformed our lives. Software is required to work on a computer but the software that we buy or download only comes in the **compiled** ready-

to-run version. It is next to impossible to modify the compiled version of the software. At times we feel the need to change certain features of the software but are unable to do so.

In this chapter we will study about software which are developed collaboratively and they can be modified as well. Such software are available in many forms like Free Software, Open Source Software (OSS), Free Open Source Software (FOSS) and Free/Liberal Open Source Software (FLOSS).Well, have you noticed something common in all these terms. That's right! It is the word free. People often correlate this word with the cost. They think that these software are available for free. Practically, these software can be acquired at very little or no cost. But, here, "free" means freedom to use.

These software can be studied, copied, redistributed freely and even modified according to one's need without seeking any kind of permission. In order to modify such software the developers also provide the source code to the users.

There do exist software which are actually "free" in the sense of price. These are known as Freeware. Lots of freeware can be downloaded from the internet for various different purposes such as currency converters, drawing graphs and charts etc. But freeware may not come with the source code. Therefore freeware differ from free software. The focus in this chapter is on free software rather than freeware.

The first formal definition of "free software" was given in 1983 by Richard Stallman, a long time member of the hacker community at the MIT Artificial Intelligence Laboratory. He insisted that a free software should give the following four freedoms to users:

- Freedom 0: The freedom to run the program for any purpose.

- Freedom 1: The freedom to study how the program works, and change it to make it do what you wish.

- Freedom 2: The freedom to redistribute copies so as to help your neighbour.

- Freedom 3: The freedom to improve the program, and release your improvements (and modified versions in general) to the public, so that the whole community benefits.

Examples of free software include the Linux Kernel, MySQL Relational Database, Apache web server, OpenOffice.org office suite and TeX and LaTeX typesetting systems.

Later on, the term "free software" was recoined as "open source software (OSS)" and soon after as "free open source software (FOSS)". In order to avoid the ambiguity in the word "free", in 2001, FOSS was termed as FLOSS, short form of "free/liberal open source software".

> **Know More!**
>
> You can get more information on open source software from opensource.org.

NRCFOSS :

National Resource Centre for Free and Open Source Software (NRCFOSS) is an initiative of the Department of Information Technology, Ministry of Communications & Information Technology, Government of India. NRCFOSS contributes to the growth of FOSS in India through Research & Development, Human Resource Development, Networking & Entrepreneurship development, as well as serves as the reference point for all FOSS related activities in the country.

Examples of FLOSS/FOSS :

Operating systems and Desktop environments

- **Linux** - Operating system kernel

- **Ubuntu** - Linux distribution with full complement of software for everyday use.

- **Google Chrome OS** - Lightweight operating system based around the web browser

- **Android smart** - phone operating system - by Google / Open Handset Alliance

- **Symbian smart** - phone operating system - by Nokia

Graphics and multimedia

- **GIMP** - Bitmap graphics editor, similar to Adobe Photoshop

- **Blender** - Advanced 3D modelling and rendering application.

- **Songbird** - similar to iTunes with built-in browser.

- **Audacity** - digital audio editor
- **Ardour** - digital audio workstation
- **F-Spot** - Photo manager

Office software

- **OpenOffice.org** - office productivity software. Comparable to Microsoft Office. It uses an open file format and can read and write Microsoft Office files
- **NeoOffice** - Mac OS X version of OpenOffice
- **PDFCreator** - creates PDFs from any Windows program.

Internet related software

- **Apache webserver** - web server
- **Mozilla Foundation**
  - **Mozilla Firefox -** web browser
  - **Mozilla Thunderbird** - mail client
- **Google Chrome** - Google's web browser

Other

- **Celestia** - 3D space simulation software.
- **Flight Gear** - flight simulator
- **Second life** - virtual world viewer
- **Wine** - a compatibility layer for computers running Linux that enables them to run many applications that were originally written for MS Windows

Programming related

- **Eclipse** - software framework and Java IDE
- **PHP** - server-side programming language
- **PERL** - Dynamic programming language

- **Python** - versatile, clean and powerful programming language used for cross-platform desktop applications, server-side scripting for websites, and scripting within java and .net environments

- **MySQL** - Database management system

- **Java** - Programming language

More comprehensive list of open-source software can be found at :

http://en.wikipedia.org/wiki/List_of_open_source_software_packages

GNU/Linux - The GNU Project was launched in 1983 by Richard Stallman of Free Software Foundation (FSF) to develop a complete Unix-like operating system which is free software: the GNU operating system. Unix-like operating systems are built from a collection of libraries, applications and developer tools, plus a kernel to allocate resources and to talk to the hardware. GNU is often used with the Linux kernel. The combination of GNU and Linux is the GNU/Linux operating system, now used by millions.

Firefox - Firefox is a free and open source web browser produced by Mozilla Foundation. Firefox runs on various versions of GNU/Linux, Mac OS X, Microsoft Windows and many other Unix-like operating systems.

Open Office - Open Office is the leading open source office suite for word processing, spreadsheets, presentations, databases etc. It is available in many languages. It stores files in open document format (ODF) for data interchange that is its default file format.

43

NetBeans - NetBeans began in 1996 as verify, a Java Integrated Development Environment (IDE) student project, under the guidance of the Faculty of Mathematics and Physics at Charles University in Prague. In 1999 it was bought by Sun Microsystem which open-sourced the NetBeans IDE in June of the following year. The NetBeans community has since continued to grow, thanks to individuals and companies using and contributing to the project. NetBeans refers to both a platform framework for Java desktop applications, and an IDE for developing applications with Java, JavaScript, PHP, Python, Ruby, C, C++, and others.

BOSS(Bharat Operating Systems Solutions) - BOSS is a free Indian Operating System based on GNU/LINUX developed by C-DAC(Center for Development of Advance Computing). BOSS makes it easier for a Microsoft Windows user to shift to GNU/LINUX platform which is there with variant features. Currently BOSS GNU/Linux Desktop is available in many Indian Languages which also enables the non-English literate users in the country to be exposed to Information and communication technology and to use the computer more effectively.

Open Source Software Security

A commonly voiced concern about open source software:

"If anyone can contribute to open source software, doesn't it become a free-for-all full of loopholes?"

The Answer is while anyone can propose a contribution to an open source project, any actual change must go through a small core group of maintainers first. Getting a change incorporated into an open source project is thus rather like getting an article published in a scientific journal.

Open source software is more heavily tested than their commercial counterparts as it can be downloaded by anybody around the world and any one of them can discover a bug or security flaw and submit those reports back to the project.

Also since it's an Open Source, if any bug or flaw is found, project's core maintainers, while potentially embarrassed, have no further reason to cover up the flaw like their proprietary counterpart. Also the loop-hole or bug is available to every one and the customer can take the preventive measure accordingly and fix comes faster.

Common open standards

Open Document Format

The Open Document format (ODF) is a format for office documents, such as spreadsheets, databases, presentations and word-processing documents. Open Document is a free and open format. For governments, businesses, archivists and others, it's critical to store documents in a way that can be read for years to come. Proprietary digital file formats are typically changing with every new version of the software, so there should be some format which supports files created in any application. The data should be the center of importance not the application. Office suite applications, as always, need a file format that is designed to organize the data when it moves away from the application. People with different machines in different places should be able to open and edit the data in a file. ODF offers an open alternative to the formats used by all of the existing Office application versions for text, spreadsheet, presentation, and other kinds of documents. Open Document's main file extensions are .odt (for text documents), .ods (for spreadsheets), and .odp (for presentations). These will be more commonly recognized when more people and organizations adopt Open Document-ready software.

Ogg Vobris

Ogg Vorbis is a new audio compression format developed by Xiph.org. It is an open, patent-free, professional audio encoding and streaming technology with all the benefits of Open source. It is comparable to other formats such as MP3, VQF, AAC etc. used to store and play digital music.

Character Encoding

A character encoding system consists of a code that associates each character from a given system with something else, such as a sequence of natural numbers, binary numbers or electrical pulses, in order to facilitate the transmission of data (generally numbers and/or text) through telecommunication networks or for storage of text in

45

computers. ASCII, EBCDIC, and UNICODE are the most widely used character encoding systems for computers.

Indian Language Computing

Let us recall from Annexure I of Class XI that in order to communicate with the computers some kind of a binary code is required. In this regard, a detailed study was made on BCD code which is a 4 bit code. However, BCD code is not sufficient enough to accommodate even all the characters of English alphabet and digits. To accommodate all these, a higher bit code is required. Among such codes, the most popular is the ASCII code (American Standard Code for Information Interchange). It is a 7 bit code that can store $2^7 = 128$ characters. In the earlier days most computers were using an 8 bit system. This extra bit gave computer developers lot of empty spaces which was used for different purposes. In order to work with Indian languages, these unspecified spaces were used for Indic characters. The efforts were very creative and gave good results despite the fact that there was no or very little support from the operating systems. Since no general rules and methodologies were adopted in developing Indic characters, different developers developed these characters in their own ways. This created compatibility issues across different programs and across different operating systems. For example browsers like Internet Explorer did not support all Indic characters and displayed distorted texts.

On one hand, more and more work had been undertaken to facilitate the use of Indian languages on computers but on the other hand it was becoming difficult to maintain consistency across different programs developed for the same set of Indic characters. In such a scenario, it was important to have a common standard for coding Indian scripts. In 1991, the Bureau of Indian Standards (BIS) adopted the Indian Script Code for Information Interchange (ISCII), the ISCII standard that was evolved by a standardization committee, under Department of Electronics during 1986-88.

ISCII is an 8 bit encoding system as compared with the 7 bit ASCII. The lower 128 code points are plain ASCII while the upper 128 code points are ISCII specific containing the basic alphabets required for the 10 Indian scripts (Assamese, Bengali, Devanagari, Gujarat, Gurmukhi, Kannada, Malayalam, Oriya, Tamil and Telugu) which have originated from the Brahmi script.

> **Know More!**
>
> The **Indic scripts** are a family of abugida (alphabetic-syllabary) writing systems. They are used throughout South Asia, Southeast Asia, and parts of Central and East Asia, and are descended from the Brahmi script of the ancient Indian subcontinent. They are used by languages of several linguistic families: Indo-European, Dravidian, Tibeto-Burman, Mongolic, Austro-Asiatic, Austronesian, Tai, and possibly Korean (hangul). They were also the source of the dictionary order of Japanese kana.

UNICODE

For a long time ASCII has been the standard code used worldwide. ISCII was the Indian contribution to work with Indian languages. Similarly, to work with other languages of the world, people were making efforts to use ASCII along with the extra 8th bit. Efforts were continued to conceive a single standard code which could cater to all world's languages and it was UNICODE.

Unicode was developed with the aim to conceive a single standard code which could manage to represent all languages of the world.

Unicode characters are represented in one of the three encoding forms: a 32-bit form (UTF-32), a 16-bit form (UTF-16), and an 8 bit form (UTF-8). The 8-bit, byte oriented form, UTF-8, has been designed for ease of use with existing ASCII-based systems.

Advantages of Unicode

- With the usage of Unicode, single versions of software were developed instead of language-specific versions that reduced the complexity.

- UNICODE is supported by most OS and application vendors. This ensures platform, vendor and application independence.

- Incorporating Unicode into applications and websites offers significant cost savings than proprietary solutions.

- It allows data to be transported through several different systems without distortion.

• Since every number and character combination is unique, the representation results in a true standard.

Different Types of Fonts

Post Script

This font format was developed by Adobe in 1980's. This font consists of two parts which are both necessary for the font to be properly printed and displayed on screen. With most operating systems, these fonts can be installed simply by being placed in the system's folder. However, for Micro Soft operating systems predating windows 2000, they need to be installed using ATM (Adobe Type Manager) utility.

True Type

This format was jointly developed by Apple and Microsoft in the late 80s. These fonts contain both the screen and printer font data in a single component, making the fonts easier to install. And that is why these are a good choice for those who find the installation of fonts difficult.

Open Type

This is the latest font format which is a joint effort by Apple and Microsoft. Like True Type fonts, this contains both the screen and printer font data in a single component. However, open type fonts support multiple platforms and expanded character sets. Additionally, open type format allows the storage of upto 65,000 characters. This additional space provides freedom to include add-ons such as small caps, old style figures, alternate characters and other extras that previously needed to be distributed as separate fonts.

However, not all open type fonts contain additional characters. Many fonts have been converted from either PostScript or TrueType formats without expanded character sets to take advantage of the cross-platform functionality benefits of Open Type. OpenType fonts that do contain expanded character sets are referred to informally as OpenType Pro fonts. Support for OpenType Pro fonts is increasing, but the format is yet to be fully supported by all applications.

Static and Dynamic fonts

Static fonts - In these types of fonts the characters are designed and digitized and then stored in font files. Every time printing takes place, same character will appear with same shape e.g. Times New Roman, Arial etc.

Dynamic fonts - Dynamic font is a web browser technology used when visiting any website that uses fonts which are not installed on client's machine. The web browser would not be able to display the page properly, but will select one of fonts available on client machine. In this the characters are redefined at each occurrence (everytime they are displayed or printed). All hand written fonts such as handwritten alphabets, calligraphic letters, graffiti etc are dynamic fonts because of individual variations.

This technology is helpful in displaying multilingual websites.

Entering Indian language text

Many tools have been developed to facilitate the typing of Indian language text. These tools broadly support two types of text entries:

- Phonetic Text Entry
- Keymap based Text Entry

Phonetic Text Entry (Transliteration)

In this type of text entry, traditional keyboards with English keys are used. But while typing, the Indian alphabets are written phonetically (i.e., the way they sound as per the pronunciation) in English Script and then converted to corresponding language word. For e.g. we will type "mera desh mahaan" from English keyboard and the relevent phonetic key entry software will transliterate it in the language selected eg. Hindi.("मेरा देश महान")

Keymap based Text Entry

In this method the keyboard keys are mapped to specific characters using a keymap. The whole arrangement of mapping the keyboard keys to specific language characters is known as keymap. A keymap is internally stored as a table. Multiple keymaps are used to store complete keymapping e.g. a table to represent the keymappings without any additional key press(such as shift) ; a table representing keymappings along with Shift key press; along with Ctrl key press and so on.

Indian language keymaps are known as INSCRIPT keymaps or Indian Script Keymaps.

ENGLISH KEYBOARD WITH INSCRIPT OVERLAY

The ASCII characters of a standard QWERTY keyboard are on the left half of a key. The INSCRIPT (Indian Script) overlay characters are shown on the right half of a key. CAPS LOCK is used to select the INSCRIPT overlay.

NUKTA CHARACTERS IN INSCRIPT OVERLAY

When Nukta "-" is typed after character, the character shown to its left on the key, is obtained.

INSCRIPT OVERLAY FOR ASSAMESE

Know More!

You can get more information on Indian script keymaps from http://tdil.mit.gov.in/isciichart.pdf.

Future Trends

At present very few open source endeavors which involve few people are there and are largely unstructured and undisciplined. But as this concept matures in future more OSS projects will come up and will adopt more structured methodologies to code, control and coordinate. In future this concept will be adopted at a big level by private and government sector. This will force companies to embrace open source and will force the open source community to innovate in line with more complex need of government and business.

Summary

- Free and open source software, also OSS, FOSS, or FLOSS (free/libre open source software) is software that is liberally licensed to grant the user right to use, study, change, and improve its design through the availability of its source code.

- NRCFOSS - National Resource Centre for Free and Open Source Software.

- ODF (Open Document Format) offers an open alternative to the formats used by all of the existing software.

- Ogg Vorbis is open audio compression format.

- A character encoding system consists of a code that associates each character from a given system with something else, such as a sequence of natural numbers, binary numbers or electrical pulses, in order to facilitate the transmission of data (generally numbers and/or text) through telecommunication networks or for storage of text in computers

- ISCII - Indian Script Code for Information Interchange is common standard for coding Indian Scripts.

- UNICODE - Universal Code is a single standard code which can cater to all the world's languages.

- Post Script font consists of two parts which are both necessary for the font to be properly printed and displayed on screen.

- True type fonts contain both the screen and printer font data in a single component.

- Open type fonts support multiple platforms and expanded character sets.

- On the basis of their configuration, fonts are categorized as static fonts and dynamic fonts.

- For entering text in Indian script keymap based or phonetic text based entry can be done.

EXERCISES

MULTIPLE CHOICE QUESTIONS

1) Open Document's main file extensions are

   (a) .odt   (b) .ods

   (c) .odp   (d) All of these

2) Technology no longer protected by copyright, available to everyone, is considered to be

   (a) Proprietary   (b) Open

   (c) Experimental   (d) in the public domain

3) Which of the following codes uses 7 bits to represent a character

   (a) ISCII   (b) ASCII

   (c) UNICODE   (d) All of the above

4) How many bits are used by UTF-8 to represent a character

   (a) 7   (b) 8

   (c) 16   (d) 32

5) ASCII stands for

   (a) American System Code for Information Interchange

   (b) American Standard Code for Information Interchange

   (c) American Standard Code for Interchange Information

   (d) American System Code for Interchange Information

6) ISCII stands for

   (a) Indian System Code for Information Interchange

   (b) Indian Standard Code for Information Interchange

   (c) Indian Script Code for Interchange Information

   (d) International Standard Code for Interchange Information

7) ISCII is

   (a)   An 8 bit character code   (b)   A 16 bit character code

   (c)   A 7 bit character code   (d)   Same as ASCII

8) Which of the following is open source operating system

   (a)   DOS   (b)   Windows 2010

   (c)   BOSS   (d)   Mac

9) Following operations are possible with OSS

   (a)   Free download

   (b)   Source code is available

   (c)   You can redistribute the modified version

   (d)   All of the above

10) Platform independence is provided by

   (a)   ASCII   (b)   ISCII

   (c)   UNICODE   (d)   All of the above

ANSWER THE FOLLOWING QUESTIONS

1) Define Open source software. Give at least two examples of OSS. Justify that free software is not open source software.

2) Give examples of two OSS in each of the following categories :

   (a)   OS   (b)   Graphics and animation

   (c)   Office software   (d)   Internet related software

   (e)   Programming related software.

3) What is a character encoding system?

4) Differentiate between true type font and open type font.

5) What do you understand by ODF?

6) What is Ogg Vorbis?

7) Expand the following terms:

OSS, FLOSS, ISCII, ASCII, BOSS, NRCFOSS

8) What is UNICODE?

9) Write short notes on GNU, BOSS, ISCII

10) What is the difference between static and dynamic fonts?

LAB EXERCISES

1) Find out which software in your school lab are open source.

2) Note down the category of software (system software or application software) to which they belong to.

3) If any of them is application software then specify its area of application.

4) Search on internet about the features of mySQL and Netbeans.

TIME BOUND TEAM BASED EXERCISE
(Team size recommended: 3 students each team)

1) Download open office on your computer. Write down its components, their areas of applications, their file extensions.

2) Procure BOSS and install it (with the permission of your teacher) on your computer and write down the features which are similar to windows and also which are not similar to windows.