
INSTITUTIONAL SOFTWARE REPOSITORY: ENABLING INTELLECTUAL PROPERTY AWARENESS ON ACADEMIC CAMPUSES

Gopi Krishna Garge

Malati Hegde

Geetha Anil Kumar

Abstract

Software is a critical digital content component that is generated on academic campus by way of mandatory course requirements. These software need to be protected with an assigned ownership and a specific terms of use. Software that is generated on campuses typically does not get archived or catalogued in campus-wide repositories. SALIS a prototype software repository addresses these needs and serves to augment typical institutional repositories. This paper attempts to bring out the need for protecting the software generated on campuses and the need to archive it in an orderly manner. Additionally SALIS also addresses Intellectual Property concerns such as ownership, protection and terms of use of the software produced and provide an effective way of archiving it.

Keywords: Institutional Repository; Software repository; Open Source Software; Software Licensing

1. Introduction

Academic campuses today have library information system that provides an idea of the library resources. Supplementing these systems are institutional repositories for archiving digital content generated on campus. Such system will eventually give way to campus-wide repositories, since access to journals and periodicals are provided online. All such repositories seem to exclude computer software generated in campus from these repositories. Such software are outputs of project work done by students, a mandatory pre-requisite for course completion.

Estimate conservatively the volume of such software generated across academic campuses. Among the 242 Universities in India, assuming 50 student projects per year per university and each such project generating 500 lines of software, the total output per year would be 60,50,000 lines of code. Note that we have considered only universities and not colleges affiliated to them.

Software is multidisciplinary and therefore the real numbers may be much higher. The potential for innovation in this context and the need for protecting the innovation emerge clearly with software being a part of the repository, the need for classification of the content arises. Therefore the infrastructure need translates into a repository that provides classification in functional, traditional and a concordance with an existing library information system.

In this paper, we provide a means of augmenting existing repositories with the capability of including software as digital content to be archived. SALIS, a prototype software repository, provides this augmentation capability.

2. SALIS: Software Repository

The software repository is a framework to provide a means of depositing software (digital content) developed within the institution. The software repository would provide a means of facilitating software resource exchanges for use by students and faculty within institution. This exchange will help catalogue software and provide appropriate guidance on software ownership and terms of use to ensure that the IP developed within the Institute is appropriately protected.

Therefore the repository, primarily an indexed database of information on software packages, can augment the traditional digital repository on the campus.

SALIS (Software and Licensing Information System) is a repository of information on software, principally free and open source software [1]. The primary objective is to promote the use of free or open source software by providing information to software developers about the availability of such software. Intellectual property information and a mention for software development reuse of its implications help in making a judicious selection of the base software. The secondary objective is to provide a databank for the Registrar of copyrights to speed up the copyrighting process, using this database as a reference for priorart. This is a pilot project and is limited to two subject areas: Data Communication Networks and Databases.

This repository is targeted at Software developers and who can choose software for strategic decision-making in their establishments, especially with respect to intellectual property issues involved with the reuse of open source software. The SALIS database provides information pertinent to the software from all aspects along with the source code and other value added information. The information is made available online via a web interface. URL: <http://www.salis.iisc.ernet.in>. There are various free and open source software repositories available on the Internet, which host the software source code and provide the minimal information of the software, such as sourceforge.net, freshmeat.net, tu cows.net, downloads.com etc. none of these provide a classification and information on Intellectual Property of the software.

3. Uses of SALIS Repository

3.1 Software Classification

Software repositories (the few mentioned above) do not have any means of classifying the software (e.g. Books are classified in a library). Software classification would either require a topic-base or a function-base for classification. In addition, such a classification should have a correspondence with patent classification schemes in use, since the database will be used in the copyright/patent office as a reference for priorart. The software compiled in the database have been classified using an internal classification scheme called SALIS and provided with a concordance against standard classification schemes such as the United States Patents & Trade Organization (USPTO), and the International Patent Classification [IPC] etc. The SALIS classification is the combination of the ACM and IPC classification system. The scheme was designed keeping in view the domains such as Data Networking and Databases [2].

3.2 Software Reuse information

SALIS database contains comments on the software reusability based on the software license. Open source software (OSS) provides scope for reusing all or portions of the readily available source code either on an "AS-IS" basis or with modifications. However extreme care should be taken to understand both the copyright and license terms of the software in question before reusing all or portions of the OSS, regardless of the intent of use (commercial use, non-commercial use, developing another OSS, etc.) so there is no infringement or misuse. The SALIS database provides such information (value addition) for software that is part of the database. Such reuse information is supplemented by the license composition of the software, thereby clearly indicating what portions of the software can be reused and what cannot be, as appropriate [3].

4. Selection of Software for SALIS Database

Keeping in view the usefulness of the database to the target audience, following criteria were used for selecting for inclusion in the SALIS:

4.1 Free and Open Source Software

Software whose source is freely available for download and whose licensing conditions permit any user to download the software and reuse it with no restrictions.

The advantages of choosing open source software are:

- Source Code is available for download and it is possible to study the software for possible reusable components. It also helps in providing significantly better technical and IP expert comments for the software that will be helpful to anyone interested in using the software.
- Usually the IP protection of free software allows reusability and modification of the source code with little or no restrictions. This makes the software interesting and useful to software developers.

4.2 Commercial Software

Some commercial packages are also included in SALIS, mainly to bring about the contrast of the IP protection involved. Usually the source code and technical information is not freely available for most of the commercial software. Hence it is not possible to give extensive technical and IP reviews for software.

The area of Communication Networks and Databases encompass a very large range of subtopics. To focus on the software included with the database, the current topics in both of these areas were chosen, for example, Network Management, Network Security, Network Monitoring, Protocols, etc. Such a choice provides for homogeneity of content and enables us to provide feature comparisons and technical reviews, which will be of value to developers when they need to identify their candidate software.

5. SALIS Database Design, Structure and Content

The database is implemented using MySQL, an RDBMS and an OSS as well. The design of the database of the repository is the most important part of the work. What attributes are important in identifying a software and giving full information and the different access point to the software database were considered while designing the schema. Each of the Software included in the database is described with a set of attributes (fields). These help in uniquely identifying the software which facilitates browsing and searching database and provide value added information about the software.

The database is designed after thorough analysis of other software repositories; the various fields in the database are chosen to provide maximum and relevant information. As Fig. 1 shows SALIS database is divided into two databases namely technical and application database. There are totally 550 software in the SALIS database [4].

- Technical Database: Software included in the SALIS Technical database fall broadly into the following domains:
 - Communication Networks
 - Database and Information Retrieval

They are usually widely used software in the areas of networking, database, and information retrieval domain. It also houses library packages that are considered to be defacto for building some applications e.g. 'libpcap' a most common packet capture library, 'libpal' a packet assembly library.

- Application Database: Software listed in the SALIS Application Database are
 - Generic free tools for Linux and windows operating systems, including Word Processor, Spread sheet etc.
 - E-Technology related software (e-governance, e-security, and e-learning) tools.

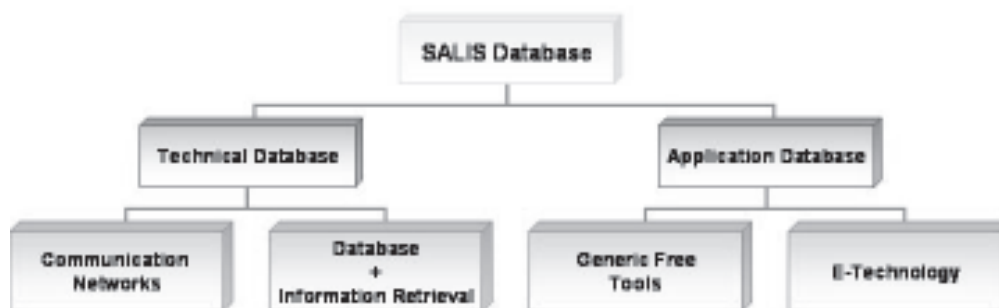


Fig. 1 SALIS Database Structure

5.1 Updating the SALIS Database:

All processes such as software uploading, downloading and other maintenance tasks are done by using web based Content Management Interface. The content management interface is designed keeping in mind the needs of the end user. You can connect to database from remote location and can perform all of the maintenance tasks on the online database. It hides the SQL queries and other tasks to make it easy for the end user in maintaining the database.

5.2 Using the SALIS Database:

The SALIS database has a Web based front end through which the database can be accessed. Fig. 2. shows the structure of the search results. These three sections provide factual information on the software collected from various sources of information including the Internet, documentation and downloaded software.

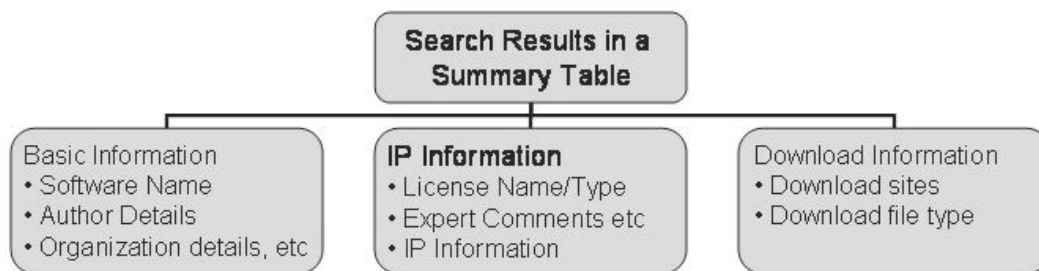


Fig. 2 Schematic representation of search results

The database comprises more than 35 attributes that are grouped into three categories, namely

- a. **Basic Information:** This includes basic factual information about the software for example Title, Description, Author details, type of IP, Operating system etc. similar information.
- b. **Intellectual property information:** This gives information on the kind of IP protection given to the software, information about the license, the license structure and components and the details of reuse.
- c. **Download Information:** Providing information on downloading the particular software

These attributes were arrived after (i) Going through similar sites providing software directory and search services on the Internet (ii) By experts in particular areas (iii) By obtaining feedback from the industry through questionnaires (iv) By discussions and deliberations about the fields by the team members.

5.3 Value addition to the database

On locating candidate software for the database, the basic information about the software is gleaned and is followed by a series of value additions.

- a. Classification: the SALIS contains software and their metadata. The repository uses a modified version of the CCS (Computing Classification Scheme) by ACM (Association of Computing Machinery) to classify the software. A need was felt to organize the information using the USPTO classification scheme to facilitate users familiar with this scheme to access the repository contents. To enable this, a mapping between the two schemes was developed.
- b. Expert Comments: This section on expert comments provides two type of value addition unique to SALIS: the technical and IP comments.
 - Technical Expert Comments: This is provided by studying the documentation, features, functionality and source code of the software. It also involves installing and running the package. The analysis is done with the help of experts in the field.
 - Intellectual Property Expert Comments: IP comments is the unique feature of SALIS, it not only provides the license type but also lists the various necessary fields relevant to Intellectual Property such as license name, type of IP, etc. commenting on the IP of the software again involves a long procedure of (a) commenting on the license of the whole package. (b) Parse through source files for embedded components, codes that are under different licenses, copyrighted by people other than the author of the package/software as a whole. (c) Providing a pictorial image, showing Graphical representation of embedded components and their licenses, (if the software has embedded components).
- c. Comparison Charts: Feature comparison charts that compare the functional features of similar packages in the database.

These value additions help the end user to perform a comprehensive survey of software available in a specific category and provide with as much information as possible to help make decision on the means of developing a product whether a product has to be built, ground up or is there a potential for reuse of OSS and build upon that.

These value additions provided by studying the documentation, features, functionality and source code of the software and also involve installing and running the package. The analysis is done with the help of experts in the field. The Fig.3 shows the Procedure followed in identifying the software and providing the comments. In this procedure the first step of the process is to identifying the sub domain, i.e. on which domain we should identify the software, after identifying second step is to identifying the software and mining it on internet. In third step collected software are given to experts to selection process, after this in fourth step selected software are uploaded to database. In fifth step, for uploaded software team will start working on it by going through its source code, documentation and study that software feature, functionality and prepare reports on Installation, Reusability, and comparison table for each software. And in final step, Technical and IP comments are written.

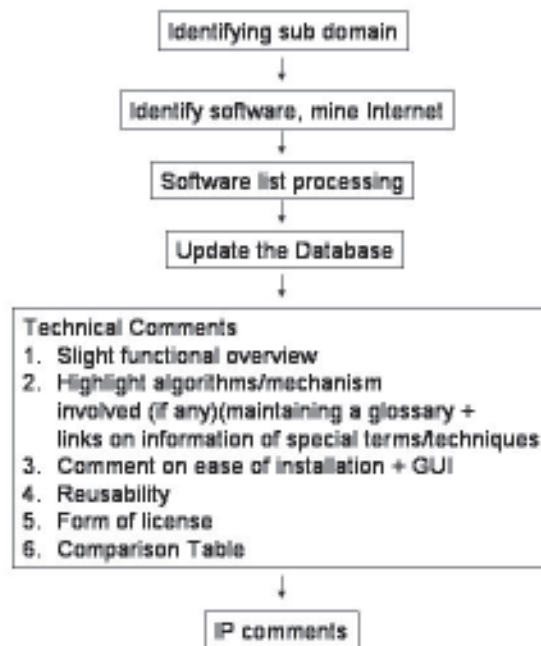


Fig. 3 Procedure to identify the s/w and provide comment

6. Adapting SALIS to augment an Institutional Repository

In general Institutional repository is a set of services that an academic campus offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members. An effective institutional repository of necessity represents collaboration among librarians, information technologists, archives and records managers, faculty, and academic administrators and policymakers [5]. At any given point in time, an institutional repository will be supported by a set of information technologies, but a key part of the services that comprise an institutional repository is the management of technological changes, and the migration of digital content from one set of technologies to the next as part of the organizational commitment to providing repository services [6].

Like Dspace, ePrints SALIS is also a digital repository [7]. It functions as a repository for the software source code produced by developers. Fig. 4 shows the architecture of the SALIS information system. SALIS provides stable long-term storage needed to house the digital products (software) of users.

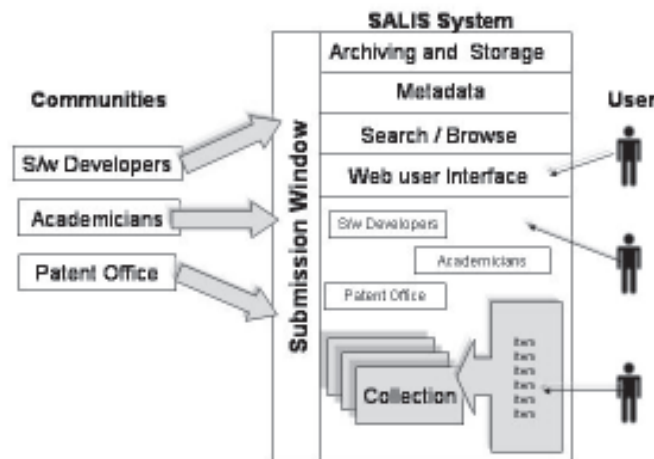


Fig. 4 SALIS Technical Architecture

7. Functional components of SALIS

The SALIS user interface is web-based. There are two interfaces: one for users who submit content and the other for end-users looking for information.

7.1 Content management Interface

The content management interface is an interface to gather content into the database. Information related to the software is collected and put into the database through this interface. It is a web based interface, which will allow content to be sent to the database via HTML forms. The connectivity to the database is done using PHP.

- Technology platform: SALIS is developed in such a way that institutions and organizations could run with minimal resources. The system comprises other open source middleware and tools, and programs written by the SALIS team. All original code is in the HTML and PHP. Other pieces of the technology stack include a relational database management system (MySQL), a Web server and Apache, (from the Apache Foundation). All leveraged components and libraries are also open source software.
- Database: SALIS database will be of interest to software developers who can, prior to developing a product, can refer to this database to check for readily available Open Source Software components that can potentially be used in their product development.
- Search Interface: The SALIS database has a front end search interface. The search facility permits keyword searches as well as search with a list of attributes such as license, functionality, operating system the software works on etc. the search results are brief and provided as a listing with a provision to as for additional information classified into Basic, Intellectual Property and Download Information.

7.2 Feature comparison of few such repositories

Table – 2 lists the similar software repositories features comparison. Here we have compared them on the basis of some parameters which SALIS is giving as value added information for software packages. Here we can notice that among below mentioned parameters, only description is provided by other repositories, where as in SALIS its giving information on Software functionality, Comments on technical aspects, Comments on IP, Reusability, Comparison Report, Installation Report, Component analysis, where non of the other software repositories provide these information.

Features	Sourceforge.net	Freshmeat.net	Downloads.com	SALIS
Description	✓	✓	✓	✓
Software functionality	x	x	x	✓
Comments on technical aspects	x	x	x	✓
Comments on IP	x	x	x	✓
Reusability	x	x	x	✓
Comparison Report	x	x	x	✓
Installation Report	x	x	x	✓
Component analysis	x	x	x	✓

8. How to integrate Solutions using SALIS ?

In order to demonstrate the typical use of the SALIS database in a developer environment, the development of a typical product prototype was chosen. The target product was a low cost small office home office (SOHO) router. The router is expected to work in a small office, say a Panchayat office, being able to provide connectivity to the upstream government network infrastructure as well as be able to connect local subnetworks and have them interwork. The typical product design cycle brought out the functional requirement of such a router. Given the target segment, it would be ideal to have a host based router that can perform a multiple role as an IP router as well as an access terminal for a user. Such a router ought to have firewalling capabilities, an Intrusion Detection System (IDS) and an element of Network Monitoring. The SALIS Database was searched to find the candidate software. Software that was free and had unrestrictive terms of use were chosen. Most of these software worked on the Linux OS, a free and unrestrictive use software. The product was to run on commodity hardware available in the market.

SALIS SOHO router is built using the RedHat Linux 8.0 (standard, downloadable) distribution and packaged with Open Source Software (OSS) that performs other important functions required by a typical SOHO unit. The router has a powerful firewall and intrusion detection system installed that need to be customized to the network it is installed on.

9. Conclusion

SALIS is a repository of software packages. It is used to archive software content developed on campuses. SALIS augments a typical institutional repository and serves to include IP information for the software generated on campus. Thus, when a user his software in SALIS, S/He would become

aware of the IP issues also. In several decisions with the Industry, the perception of SALIS also has been that of a knowledge management tool. As a prototype deployment, SALIS is being deployed on the campus at IISc, to capture software developed on campus.

References

1. Technical Report: Software And Licensing Information System (SALIS), Malati Hegde, Gopi Krishna Garge, Geetha Anil Kumar, Suvarsha Walters, Anand S. Katti.
2. Minj, Suvarsha, Rajashekhar, T B. Mapping of two schemes of classification for software classification. *Cataloging & Classification Quarterly*, Volume 41, no. 1, 2005.
3. Mathew, Mary, Hegde, Malati, Garge, Gopi. "Intellectual Property in Software: Insights for Indian Business", *Journal of Intellectual Property Rights*, Vol 9, November 2004, pp 515-532.
4. SALIS website <http://salis.iisc.ernet.in>
5. Lynch, Clifford A. "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age" *ARL*, no. 226 (February 2003): 1-7.
6. Lynch, Clifford A., "Institutional repositories: essential infrastructure for scholarship in the digital age", *ARL BiMonthly Report*, 2003, No. 226, 1-7.
7. DAEDALUS repository for published and peer reviewed material <http://eprints.gla.ac.uk>