British Journal of Mathematics & Computer Science
3(2): 99-107, 2013

SCIENCEDOMAIN *international*
www.sciencedomain.org

Application of COTS Components: A Software Package for Crystallography

N. Gnanasankaran[1*], S. Natarajan[2], K. Alagarsamy[3] and K. Iyakutti[4]

[1]MCA Department, K.L.N. College of Information Technology, Pottapalayam-630 611, India.
[2]School of Physics, Madurai Kamaraj University, Madurai-625 021, India.
[3]Computer Centre, Madurai Kamaraj University, Madurai-625 021, India.
[4]Dept. of Physics and Nanotechnology,SRM University,Kattankulathur-603 203, India.

Research Article

Received: 22 December 2012
Accepted: 2 March 2013
Published: 13 March 2013

Abstract

Commercial Off-The-Shelf (COTS) technology is widely used in many industries and also in scientific computing. This paper will first explore the term itself and then tries to find how it typically manifests itself in the softwares used in physical sciences. Some details regarding one of the popular software used for crystal structure elucidation in Crystallography, viz., SHELX are furnished. Since this is a freely available program and a non-commercial one, it is better to consider this as an example of Scientific Off-The-Shelf (SOTS) components, a newly coined acronym.

Keywords: COTS, softwares for scientific work, crystallography, SOTS.

1 Introduction

1.1 Defining COTS

It is a significant challenge to relate COTS to the operation of your business. Every industry is faced with the "build or buy" dilemma. The homemaker can cook the food and side dishes from rice, vegetables and groceries – or can buy it. The scientist/programmer within a research institution can build a software application to study the data he has obtained – or he can buy it.

In the strictest definition [1] COTS systems have the following characteristics: the buyer has no access to source code; the vendor controls its development; and the software has a nontrivial installed base. A more relaxed definition encompasses the combination of any set of software components or applications purchased and combined into a large software system. Almost without

**Corresponding author: sankarn.iisc@gmail.com;*

exception, every software-related endeavour will utilize a significant percentage of COTS software components.

2 Component-Based Development

Component-based development offers several focus of reuse. Some common ways to acquire components are from:

I. External component libraries.
II. Commercial off-the-shelf (COTS) components.
III. Open source software (OSS) components.
IV. In-house component infrastructure.
V. Mining components from legacy systems.

2.1 External Component Libraries

The different ways of acquiring components comes from different approaches to software reuse, where external component libraries are in most cases only related to development with reuse, while using an in-house component repository and developing reusable components from legacy systems are also related to development for reuse.

2.2 Commercial Off-The-Shelf Components

COTS components are external executable software components being sold, leased, or licensed to the general public; offered by a vendor trying to profit from it; supported and evolved by the vendor and used by the customers without source code access or modification ("black box"). Examples are libraries to support 3D graphic modeling and CORBA middleware.

2.3 Open Source Software Components

An OSS component is an external component for which the source code is available ("white box") and the source code can be acquired either free of charge or for a nominal fee, and with a possible obligation to report back any changes done. An example of this is libraries for using and converting PNG-format images. According to preliminary results in a study of motivation for using off-the-shelf components (both COTS and OSS) by Li et al. [2] the main motivation for using either of these types of component is to achieve shorter time-to-market, reduced development effort and higher reliability. The main motivation to prefer COTS over OSS was the belief that COTS more strongly follows market trend and also the belief that the general quality and vendor support ability is better for COTS. The key motivation for selecting OSS components was access to free source code [2].

2.4 In-House Component Infrastructure

This view on components is based on the developers making components for their software that are general enough to be used in several development projects. This can either be done by having

a concrete component reuse policy with a target of producing reusable components or it can happen in a more haphazard way by using whatever components that is available when possible.

2.5 Mining Components from Legacy Systems

As reported by Sindre et al. [3] it is possible to obtain reusable components from existing, application-specific legacy code. This process is also called component reengineering.

3 Evaluation of COTS Products

For systems that depend on COTS products, the evaluation and selection of appropriate products is essential to the success of the entire system. An evaluation process defined by the SEI and National Research Council of Canada (NRC), called PECA, helps organizations make carefully reasoned and sound product decisions. PECA was named for the four main activities that make up the process, viz., Planning the evaluation, Establishing the criteria, Collecting the data and Analyzing the data.

Although the PECA process was derived in part from ISO 14598 [4], the process was freely adapted to fit the needs of COTS software product evaluation. The process begins with initial planning for an evaluation of a COTS product (or products) and concludes with a recommendation to the decision maker [5].

Selection of the evaluated COTS components, their integration and how executions of the COTS components are orchestrated are the other important features to be considered in the process.

4 Softwares Utilizing COTS Components

In general, most of the software-related endevours utilize a considerable percentage of COTS software components. The application of COTS components in a software package, viz., SCILAB used in mathematics was evaluated by us earlier [6]. In another contribution, the case of a molecular modeling software, viz., GROMACS was taken up as a case study [7]. In this paper, one of the popular softwares (SHELX) used for elucidating crystal structures using X-ray diffraction data is considered for detailed study.

4.1 SHELX System of Programs

Several software program packages are available for elucidating the crystal structures, used in the various branches of science, viz., chemistry, physics, biology, mineralogy, material science, etc. In this section, some of the features of the SHELX system of programs are considered. SHELXS is often employed for structure elucidation and SHELXL for the refinement of the structure determined. The presently available advanced versions are known as SHELXS-97 and SHELXL-97. SHELXS-97 makes use of the ‘direct methods’ for the elucidation of small-molecule structures with up to about 100 unique non-hydrogen atoms. SHELXL-97 performs full-matrix, blocked full-matrix or conjugate-gradient least-squares refinement using a conventional structure factor summation with complex structure factors [8]. More information is available from the SHELX homepage at http://shelx.uni-ac.gwdg.de/SHELX/. The physics involved in crystal

structure analysis and the mathematics used are detailed in the following sections. This is to show how elegant programs developed by scientists are useful in several branches of science; in this case Crystallography.

5 Crystal Structure Analysis - An Example from Crystallography

5.1 General Introduction

X-ray structure determination is a method of precisely finding the positions of the atoms in a material, using experimental X-ray diffraction data. To carry out this procedure, two important parameters viz., (i) amplitude and (ii) phase of the diffracted X-ray beam have to be obtained for each hkl reflection. X-ray diffraction methods of data collection provide only the intensity data of Bragg reflections, hkl, but not the phase information.

Resolving the phase problem is central to structure determination by X-ray diffraction. The mathematical method for transforming a diffraction pattern into an image of the object is called Fourier transform and hence the structure determination [8-10]. The various steps in the process of crystal structure elucidation, with the mathematics involved are furnished in the following sections.

5.2 Structure Factor

It is convenient to express the positions of atoms in the unit cell with respect to origin in terms of fractional coordinates. If (x, y, z) are the fractional coordinates of an atom, the structure factor F_{hkl} is the resultant of waves scattered by N atoms in the unit cell in the direction of the reflection hkl and is given by the equation:

$$F_{hkl} = \sum_{j=1}^{N} f_j \exp\left[2\pi i(hx_j + ky_j + lz_j)\right] \tag{1}$$

This structure factor F_{hkl} is a complex quantity [9] involving both the amplitude and the phase of the diffracted reflection, as given below:

$$F_{hkl} = \left|F_{hkl}\right| \exp(i\alpha_{hkl}) \tag{2}$$

where, $|F_{hkl}|$ is the amplitude and α_{hkl} is the phase of the hkl reflection. Also,

$$F_{hkl} = \sum_{j=1}^{N} f_j \cos\left[2\pi(hx_j + ky_j + lz_j)\right] + i\sum_{j=1}^{N} f_j \sin\left[2\pi(hx_j + ky_j + lz_j)\right] \tag{3}$$

$$\text{Also, } F_{hkl} = A_{hkl} + iB_{hkl} \tag{4}$$

The phase angle associated with each F_{hkl} is given as,

$$\alpha_{hkl} = \tan^{-1}\left(\frac{B_{hkl}}{A_{hkl}}\right) \tag{5}$$

If the structure is centrosymmetric, B_{hkl} vanishes and the structure factor is no longer complex and the phase angle α_{hkl} can only be 0 or π. If the structure is non-centrosymmetric, then the structure factor becomes complex and the phases can have any value between 0 and 2π.

From the above equations, it may be seen that each atom in the unit cell contributes to the intensity of each reflection by an amount that depends on its position in the unit cell and its scattering power. Thus, each F_{hkl} contains information about the positions of the atoms in the unit cell and hence the entire crystal structure [9-11].

5.3 Structure Solution Methods

Structure determination is the method of finding the positions of the atoms in the unit cell, which represents the distribution of electron density of the atoms in the crystal. Since X-rays are scattered predominantly by electrons, the crystal structure is, in essence, the representation of the electron density distribution around the atoms in the unit cell. As the electron density in a crystal varies continuously and periodically, the electron density, ρ(XYZ), at position X, Y, Z in the unit cell of volume, V, is

$$\rho(XYZ) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l} F_{hkl} \exp[-2\pi i(hX + kY + lZ)] \tag{6}$$

Accordingly, the electron density at any point in the unit cell can be computed from Fourier synthesis. Both amplitudes $|F_{hkl}|$ and phase angles α_{hkl} of Bragg reflections hkl, are required to compute the electron density. But the above equation can not be used directly as only $|F_{hkl}|$ are obtained from diffraction intensities and the corresponding phase information is lost. This is called the phase problem in crystallography. Hence, the above equation can be rewritten as

$$\rho(XYZ) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l} \left|F_{hkl}\right| \exp[-2\pi i(hX + kY + lZ) + i\alpha_{hkl}] \tag{7}$$

In order to compute the electron density in the unit cell, it is necessary to find the phase angles of each of the diffracted beams. To obtain an initial set of phases for hkl reflections, few procedures have been employed in crystallography. They are: (i) Direct methods (ii) Patterson and heavy atom method and (iii) Isomorphous replacement method [9]. Among these, direct methods are the popular choice to solve the crystal structures of the compounds. In the following section, a brief account of this method is given.

5.4 Direct Methods

"Direct methods" is employed in determining the initial phases of a set of reflections from their internal phase relationships. These methods are extensively used for structure determination of small molecules, especially of light-atom structures (with H, C, N and O atoms). The success of these methods is automation and packaged programs for structure determination of molecules consisting of about 100 non-hydrogen atoms. The basis of direct methods imposes two conditions

that restrict the relative values of the phase angles to be estimated from the values of $|F_{hkl}|$. The electron density in the unit cell is always either positive or zero but never negative and the electron density maps should have large values at and near atomic positions and nearly zero values elsewhere. The various steps involved in the application of direct methods are as follows:

Normalized structure factors $|E_{hkl}|$ are calculated from the magnitudes of observed structure factor amplitudes F_{hkl}, using the following equation:

$$\left|E_{hkl}\right| = \frac{\left|F_{hkl}\right|}{\left[\sum f_j^2\right]^{1/2}} \tag{8}$$

Only $|E_{hkl}|$ values greater than 1.5 are used as they are the main contributors to the E-map. The probability for each triple product is high if each $|E_{hkl}|$ value is very high. The tangent formula [12] is used to give a consistent set of relative phases with probability as given below:

$$P_{hkl} = \tfrac{1}{2} + \tfrac{1}{2}\tanh\left(\left|E_{hkl}E_{h'k'l'}E_{h-h',k-k',l-l'}\right|\right) \tag{9}$$

The best phase set is used to find an E-map, calculated the same way as an electron density map, with E_{hkl} instead of F_{hkl} as the Fourier coefficient. The area around each peak in the E-map is searched for evidence of atomic connectivity and hence molecular fragments and respective positions selected are used as an initial trial structure. The above processes could be achieved using the program: SHELXS-97.

5.5 Structure Refinement

SHELXL-97 [8] is a program for the refinement of crystal structures from diffraction data and is primarily intended for single crystal X-ray data of small moiety structures.

In crystallography, the method of least-squares is employed for structure refinement. Since the number of observations exceeds the number of variables, the best fit according to the principle of least-squares is obtained by minimization of the quantity D, defined in the following equation:

$$D = \sum_{hkl} w_{hkl}\left(\left|F_o\right| - k\left|F_c\right|\right)^2 \tag{10}$$

where, F_o is the observed structure factor, F_c, the calculated structure factor, k, the scale factor and w, the weight assigned to the reflections based on $\sigma(F_{hkl})$ and the summation is extended over all the observed reflections. As the functional form of the structure factor is transcendental, a truncated Taylor series approximation is used to set up the normal equations, the solution of which gives a better approximation of the refined quantities. This process is repeated until convergence is attained as indicated by a low shift in estimated standard deviation values.

All the non-hydrogen atoms are assigned isotropic temperature factors during initial stages of structure refinement. Later on, the anisotropic nature of the thermal vibrations of the atoms is introduced for the non-hydrogen atoms described by the factor:

$$\exp\left[-2\pi^2\left[\begin{matrix}U_{11}h^2(a^*)^2+U_{22}k^2(b^*)^2+U_{33}l^2(c^*)^2+2U_{12}hka^*b^*+\\2U_{23}klb^*c^*+2U_{13}lha^*c^*\end{matrix}\right]\right] \tag{11}$$

The parameters U_{ij} s are refined until convergence is reached. Using the SHELXL-97, the positioning and refining of hydrogen atoms in crystal structures can also be done. The hydrogen atoms in various functional groups like methyl, methylene, hydroxyl, ammonium etc., can thus be located.

The residual or reliability factor R of the crystal structure determination is given by the equation:

$$R=\frac{\sum\left\|F_o\right|-\left|F_c\right\|}{\sum\left|F_o\right|} \tag{12}$$

5.6 Computations Involved

From the details given in the above paragraphs, it is easy to understand that one has to go through the following calculations, step by step to arrive at the precise structure of the molecule, using the X-ray data collected.

(i) structure factor calculations for the various data points (reflections: h, k, l) for the particular space group of the crystal system, (ii) phase estimation based on the results of the direct methods used to solve the phase problem, (iii) calculation of the electron densities using Fourier series calculations and (iv) refining the positions of the atoms and thermal factors using the least-squares procedures.

Suitable computer programs have been written and are available for the different steps of the calculations, stated above. The modern software packages such as SHELXS-97 and SHELXL-97 incorporates all the above programs in a unique way and helps the crystallographers in elucidating crystal structures easily. The readers may refer a recent paper published by Natarajan et al. [13] to see the results obtained using the above programs in the case of an organic molecule(L-Asparagine-L-Tartaric acid).

6 Conclusions

It is easy to visualize the usage of **Commercial Off-The-Shelf (COTS)** technology in the above example. The modern packages are very versatile in their functioning and the quality is not compromised in utilizing the above software package. Hence, the above example could be considered as a positive usage of COTS components, in designing a software package for the usage of a group of scientists. Since the program package discussed above is freely downloadable from websites and a non-commercial one, it is better to consider this as an example of **Scientific Off-The-Shelf (SOTS)** components, a newly coined acronym.

Acknowledgements

One of the authors (NG) thanks the Management and the Principal of the K.L.N. College of Information Technology, for the encouragements.

Competing Interests

Authors have declared that no competing interests exist.

References

[1] Basili V, Boehm B. COTS-Based Systems Top 10 List. IEEE Computer. 2001;34(5):91-93.

[2] Li J, Conradi R, Slyngstad OPN, Bunse C, Torchiano M, Morisio M. Preliminary results of a state-of-the-practice survey on motivations of using off-the-shelf component, 6th Int'l Conference on Product Focused Software Process Improvement (PROFES'2005), Oulu, Finland. June 2005;13-16.

[3] Sindre G, Conradi R, Karlson EA. The reboot approach to software reuse. Journal of Systems and Software. 1995;30(3):201-212.

[4] International Organization for Standardization (ISO). ISO/IEC 14598-1:1999 – Information Technology–Software Product Evaluation. Geneva, Switzerland: ISO/IEC; 1999.

[5] Comella-Dorda S, Dean J, Lewis G, Morris E, Oberndorf P, Harper E. A Process for COTS Software Product Evaluation, Technical Report CMU/SEI-2003-TR-017 ESC-TR-2003-017, Carnegie Mellon Software Engineering Institute, Pittsburgh, USA; July 2004.

[6] Gnanasankaran N, Natarajan S, Iyakutti K, Alagarsamy K. Commercial off-the-shelf (COTS) components in software engineering: The Software Package SCILAB. Int. J. Computer Technology & Applications. 2013;4(1):68-71.

[7] Gnanasankaran N, Natarajan S, Iyakutti K, Alagarsamy K. A case study of the Application of COTS Components in a Molecular Dynamics Software. Proceedings of ICECT-12(IEEE), Kanyakumari, India:6-8. April 2012;1:115-117.

[8] Sheldrick GM. A short history of SHELX. Acta Cryst A. 2008;64:112-122.

[9] Stout GH, Jenson LH. X-ray structure determination: a practical guide, 2nd ed. New York : John Wiley and Sons; 1989.

[10] Narayanan P. Essentials of Biophysics, New Delhi: New Age International Publishers; 2000.

[11] Giacovazzo C, Monaco HL, Viterbo D, Scordari F, Gilli G, Zanotti G, et al. Fundamentals of Crystallography, New York: Oxford University Press; 1992.

[12] Karle J, Hauptman H. Semiinvariants for non-centrosymmetric space groups with conventional centered cells. Acta Cryst. 1961;14:217-223.

[13] Natarajan S, Hema V, Kalyana Sundar J, Suresh J, Lakshman PLN. L-Asparagine-L-tartaric acid(1/1). Acta Cryst. *E* 2010; 66:o2239-o2240.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
www.sciencedomain.org/review-history.php?iid=206&id=6&aid=1091