An Internship in Information Systems: Combining Computer Science Education with Realistic Problems

John Buck and Ben Shneiderman
Department of Computer Science
Indiana University
Bloomington, Indiana 47401

Introduction

Computer science graduates who become professional programmers will have a direct and substantial influence on the impact of applications, but little in traditional computer science training curriculum prepares them for this serious responsibility. Recognizing this situation, we designed a two term sequence for advanced undergraduates and masters students which would not only provide them with the required academic knowledge, but would develop the pragmatic skills and professional attitude necessary for dealing with realistic problems and working environments. The educational atmosphere that we tried to create resembles the internship phase followed in teacher training, medical schools, law schools, clinical psychology and other disciplines.

Theory vs. Practice

The dual aspects of computer science, theory and practice, have been discussed by a number of commentators (Wegner, 1970; Shneiderman, 1971; Finerman, 1975). The theoretician works in greater isolation, deriving feedback, through the journal literature, from the small number of specialists familiar with the particular subfield. Objective criterion for creative ideas are more feasible for the theoretician and the impact of a poorly conceived paper is limited. By contrast, the practitioner develops programs by intense interaction with diverse individuals: programmers with a wide range of skills and subject area specialists who have called the programmer analysts to develop an implementation. Feedback from users and management is more immediate and more complex. Practitioners must respond to the diverse demands of users who have little concern for elegant programming techniques but expect systems to meet their needs. Management also has little interest in programming details but is concerned with economic issues, adherence to a schedule, satisfaction of higher management and a wide range of personnel problems. Success is difficult to measure because of the multiplicity of conflicting demands. The programmer/analyst is caught in the middle; trying to do a proper technical job while satisfying the wishes and personal quirks of users and management.

Finerman (1975) strongly supports the development of skills other than the purely academic ones and McFarlan and Nolan (1973) stress that:

> "Our research has increasingly led us to believe that the effectiveness of using the computer in organizations is often influenced more by understanding and resolution of the management issues ... than by a detailed insight into the underlying issues of technology."

Entwined in these conflicts is the question of social and professional conscience. Programmer/Analysts often make decisions about details which critically influence the new system's impact. They must deal with the desires of those directly concerned with a new system while adhering to professional standards and considering the impact of their work on society. The ACM report on "Guidelines for Humanizing Computerized Information Systems" (Sterling, 1974) should be required reading for all professional programmers and analysts.

Motivation for the Internship in Information Systems

The lay population and informed professionals derive their impressions of computers, computer programmers and computer scientists mainly from interaction with the products of computer practitioners. These impressions have not always been positive, partially due to the ineptness of some practitioners. To produce a more positive impression, practitioners must be taught to be not only intellectually able, but sensitive to their product's impact and their own behavior in a professional environment.

Keeping these concerns in mind we developed the internship in Information Systems with the following goals:

1) to teach students the academic knowledge relevant to designing and implementing realistic medium scale information processing projects (1 to 10 person-years of effort),

2) to develop communication skills necessary for interviewing users and making presentations to management,

3) to give students experience in writing complete feasibility studies, design specifications, systems manuals and users manuals and

4) to expose students to the complexity of interacting with others on a project development team.

Two further goals above the level of direct concern to our students were:

5) to serve the university and civic community by providing a free systems analysis and programming ability to organizations which could not afford it or were unsure of making an investment for such services (some of our projects were pilot projects which were intended to lead to full scale applications later).

6) to gain favorable recognition for the Computer Science Department and for computer science in general.

Detailed Course Specifications

Before entering the Information Systems sequence the students must have had a minimum of 2 semesters of heavy programming coursework using at least 2 high level languages. They have had exposure to basic machine structures, symbolic coding and assembly systems, data analysis, sorting, searching, and string and text manipulation. An additional prerequisite is a course on Data Structures covering the structure and use of storage media, methods of representing data, and techniques for operating on data structures (typically includes Knuth Vol. I, Chapter 2, The Art of Computer Programming, Addison-Wesley, 1968).

The initial semester of our sequence is titled Information Systems Design. Its customary enrollment is about 40 students. There are three lectures a week with additional required homework and a project assignment.

The course begins with a three week discussion of file indexing techniques, indexed sequential files, randomization, data base concepts, and related topics. As a text we have used Data Management for On-Line Systems (Hayden Boon, 1974) by David Lefkovitz. While the book is well done and extensive in its coverage we use only a portion of the text. The students are required to develop a program for the maintenance of a large scale file for credit reference and balance purposes, as would be maintained by a retail business, using the techniques discussed. This may be done in any language on whatever equipment they choose. A CDC 6600, a DEC-10 and an IBM 370-158, as well as a number of mini computers, are available for student use.

The remainder of the course centers around the methods of Systems Analysis as used in business for the development of computer related systems. At this point sample projects are presented and students are encouraged to select one from the list given or create one of their own. The majority of these projects are tasks which have practical application and have been solicited from the University and local community. Simultaneously students form teams of 2 to 5 members and accept the project responsibility jointly. The student teams get real exposure to the problems of personal relationship and system development which go beyond the basic computer problem. Examples of projects undertaken in the past appear in the Appendix. The text we have used for this portion of the course is Information Systems Analysis (SRA,1974) by M. J. Alexander. It provides not only a good discussion of the various techniques employed, reports required, and criteria of performance but also a selected group of systems are described which allow the student to see these techniques in application to realistic problems. A new text which looks like a good alternate is Systems Analysis: Definition, Process, and Design (SRA,1976) by Philip C. Semprevivo. It has a clear presentation of the material necessary and an orientation towards Data Base Systems.

During this segment of the course we place great emphasis on the human relationship problems encountered when doing a system study. The importance of interview techniques and presentations is stressed. Flowcharts, Decision Tables, Gantt Charts, Simulation and other methods of system modeling are discussed as well as the documentation required, in normal practice, to design and define a system. Cost analysis is introduced in the preparation of a preliminary survey which the "customer" must sign off on before the student may continue. System design, complete with input and output specification and the design of the forms associated with those functions, is the next phase of the work. At the conclusion of this period the student teams have developed a complete system with full documentation, so that it can be handed to a programming team for implementation. The project teams do a live presentation to the class and to the "customer" for whom the system is being done. As stated earlier, each of these projects are done, as much as possible, with an actual user thereby maximizing the practicality of the experience for the student. After presentation of the final design and acceptance by the user, the students are ready to proceed to the second semester

when they take on the role of programmers. Of course, the progress of the students and their design is monitored by the faculty, and suitable performance by the students is assured.

In the second semester, Information Systems Development, hardware organization and software constructs such as sequential access, direct access, hash coding, indexing strategies, inverted files, rings, trees, and multilinked structures are discussed in depth. These concepts are applied by the students in programming their projects. Detailed mathematical comparisons of storage structures and retrieval algorithms are presented to assist students in making implementation decisions for their projects. Informal discussions about the practical problems of implementing large systems are held occasionally to deal with problems raised by particular projects. In a form of "group therapy" students are asked to reveal the difficulties they are having with their projects and to frankly discuss the unexpected problems that they have encountered.

During the second half of the term a thorough overview of contemporary database concepts is presented. Codd's relational model of data and numerous query languages developed for it are discussed. The hierarchic organization of IBM's Information Management System (IMS) or MRI's System 2000 is shown as a second conceptual framework. Finally, the network structure provided by the CODASYL Data Base Task Group (DBTG) Report is surveyed.

Texts used in the second semester are The Art of Computer Programming, Vol. III (Addison-Wesley, 1973) by Knuth and Introduction to Data Base Systems (Addison-Wesley, 1975) by C. J. Date. A good alternative in a single text would be Computer Data Base Organization (Prentice-Hall, 1975) by James Martin. Students must complete their project and have a running system by the end of the semester.

The thrust of this sequence is to provide a real system development experience for the student. Delays in design decisions, problems in development i.e. test time on the computer, availability of equipment, etc. are all met in the course of their endeavors. Still the deadline approaches just as it would in a real world situation.

We are pleased by the encouraging feedback about our internship program. Local organizations are eager to use the services of our students and their products are now functioning to aid the community. Some of the projects are of such a nature that they may eventually receive national prominence. Others are serving much needed areas of purely local interest. Students who have gone on to professional programming careers point to the sequence as the most helpful course they had while at IU since it so closely parallels the experiences that they were now finding themselves in. Many attribute their ability to get their job to the fact that they could show concrete accomplishments on their projects.

Summary

The goal of this sequence was to develop the pragmatic skills and professional attitude necessary for dealing with realistic problems and working environments. In addition to lectures on academic topics, students were required to complete substantial realistic projects, thereby creating a highly motivating internship atmosphere.

References

1. Finerman, A. "Professionalism in the Computing Field" CACM 18, 1 (January 1975).
2. McFarlan, F.W. and R.L. Nolan. "Curriculum Recommendations for Graduate Professional Programs in Information Systems: Recommended Addendum on Information Systems Administration: CACM 16, 7 (July 1973).
3. Shneiderman, B. "Computer Science Education and Social Relevance" SIGCSE Bulletin Vol. 3, No. 1 (March 1971).
4. Sterling, T.D. "Guidelines for Humanizing Computerized Information Systems: A Report from Stanley House" CACM 17, 11 (November 1974).
5. Wegner, P. "Three Computer Cultures - Computer Technology, Computer Mathematics and Computer Science: in Advances in Computers 10, (1970).

Appendix

On-Line Optometry Clinic Patient Database

This highly successful project, carried out by two students, provided the Optometry Clinic at Indiana University with an on-line information system to keep track of patient data. The system is in use by clerical personnel for maintaining patient addresses, appointment dates, appointment histories and limited clinical information for the 8,000 patients seen by the clinic each year. In addition to coping with the file management issues and interactive terminal dialogues, the implementors had to deal with the privacy questions related to medical information.

Voter Registration and Precinct Mapping System

Making use of detailed street and address information collected by volunteers, two students implemented a system for maintaining block by block voter registration information for Monroe County, Indiana. Immediate benefits were a complete, up-to-date and accurate street guide, which could be updated regularly with a minimal effort. Future efforts are designed to increase the accuracy of voter registration books and simplify office procedures at the Board of

Voter Registration.

International Trade Data Economic Report Generator

A large 12 reel sequential tape file containing international trade data was made available to economic researchers at Indiana University. The data base contains import/export information from 22 reporting countries and 240 partner countries covering 10,000 commodities over 15 years. Querying the sequential file was a costly and time consuming affair requiring multiple scans of the entire file, when only a fraction of the information was necessary for a given research direction. To remedy this situation, a direct access file was constructed from the relevant information and a simple querying facility was developed to allow non-programmers to access the database.

African Studies Slide Collection

A four person team worked with the photo-librarian of the African Studies program at Indiana University to develop a computerized index to a collection of 10,000 slides of Africa. A complex indexing and cataloguing scheme was used to enable users to get output describing the holdings by photographer, topic, region or by a series of keywords which were stored in a glossary. The manager of the slide collection could add and delete slides, add to the glossary and receive status reports.

Traffic Information System

Working with an applied psychology professor several students wrote a program to maintain detailed information about street intersections in Bloomington, Indiana. The system will enable researchers to study patterns of signs and street conditions along heavily traveled routes with the hope of making suggestions for improving safety at intersections. The eventual plan is to correlate this information with traffic accident data and get support from the U.S. Department of Transportation.

Wrestlers Information Management Package

In answer to a request from the National Amateur Athletic Union Office one of our teams developed a package to maintain a file of wrestlers and their records at all levels of competition. Reports could be obtained on all wrestlers from one state, school, coach, age group, weight class etc. Other programs facilitated selection for international team trials by tabulation of accomplishments in various tournaments and levels of competition. The package has been designed to make it easily adaptable to other sports.

Bill of Materials Processor

One group working with a local manufacturer and the Graduate School of Business developed a much more efficient BMP system than had been thought possible by either the School of Business or the manufacturer. Capitalizing on the specificity of the product line they were able to effectively reduce run time by an estimated 60%.

Court Case Record History System

The local court system is overloaded with cases and experiences great difficulty in scheduling the cases, lawyers, and courtrooms. This group worked with the court clerks and has designed a system to ease not only that burden but to maintain vigilance on open cases and clean up the backlog.

Indiana State Geological Survey: Petroleum and Gas Well File

The Geological Survey Petroleum File contains extensive data on 50,000 oil and gas wells in the state of Indiana. Two students prepared a system to aid in the preparation of statistics for yearly reports on drilling trends, reserves on hand etc., and provide wildcat drillers with information on other wells in the area of a projected drilling site. Searches can be performed by specifying areas in the state, geologic features or other critical variables.

Building Permits System

Working with the School of Business, two students constructed a system to support the collection, maintenance, analysis and distribution of current and past information concerning building construction activity in Indiana. The system provides monthly, quarterly and annual reports as well as supporting research queries on types of construction activity, for example residential or industrial.

Archaelogy Information System

Complex descriptions of the sites and artifacts found at particular sites are maintained by the Glenn Black Laboratory of Archaelogy at Indiana University. Unfortunately searching for specific information from the handwritten forms is a prohibitively time consuming experience. The proposed system will simplify the coding techniques and allow machine searching to facilitate research.