Computer Languages for the Processing of Text
A Status Report by a Panel of Experts
From the Standards Development Project

Panel Chairperson - Charles D. Card
Chairman, Technical Subcommittee X3J6
Manager, Industry Standards
Sperry Univac
Blue Bell, Pennsylvania

Abstract

The subject of the panel is "Computer Languages for the Processing of Text". It is a project operating under the auspices of the American National Standards Institute, Inc. through its Committee on Computers and Information Processing (X3). Committee X3, in turn has organized technical sub-committee X3J6 which has the program of work to develop the standard.

The task is extensive, but may be simply summarized. conceptual statement.   The Computer Languages for the Processing of Text are intended to be an integrated set of notation used to portray the syntax and semantics of text and operations to be performed upon that text.   The scope of the standard ranges from applications on computer systems to free standing text processing systems.

How the complex work is being carried out is the topic of discussion. The panel, composed of key members of the standardization team, will discuss the project. The background, scope, objectives, and problems the project faces will be presented.

Text

Before we can dicuss the processing of text with any degree of success we must have a definition of the scope of our discussion.   An international group of experts in the field has offered the following definition:

"Text" is defined as any sequence of characters destined for storage or display. Text in storage contains two distinct forms of text: that which is to be displayed, and "Generic Identifiers" which can be used to control the display. Generic Identifiers are as device independent as possible.

That definition seems very parochial, but the dictionary does not help much either because it removes everything an author writes except the body of a document. In other words, the dictionaries exclude traditional material which is manipulated. Now, let us examine the offered definition above. In the realm of Computers and Information Processing, what is not text?   Probably nothing! Before elaborating upon that statement let us develop the scope of the work further.   The same experts who offered the definition of text offered the following in the same document.

"Text Processing" is defined as the systematic manipulation of text. It includes generation (text input), editing, printing, retrieval, and display of text.   A specific application of text processing is the transformation of Generic Identifiers to device dependent control text.

Following the strong comment above, let us see where the limits of scope are applied to the standardization project. The material below was exerpted from the charter and program of work of the standards development project.

The following is a brief and incomplete list to draw attention to the gamut of entities subject to text processing.

BILLS OF MATERIAL	MENUS
BOOKS	MUSICAL SCORES
CATALOGUES	NEWSPAPERS
CONTRACTS	PERIODICALS
COMPUTER PROGRAMS	PRESENTATION SLIDES
COMPUTER DEBUGGING	PROPOSALS
CROSSWORD PUZZLES	REPORTS
LEGISLATION	SCOREBOARDS
LETTERS	SPECIFICATIONS
MAILING LISTS	STANDARDS
MANUALS	

Please note that the list was intended to be representative. No bias or exclusionary intent is there. Yes, there is no recorded data in the field of computers and information processing that can not be properly considered to be text.

Does that imply that text processing will supplant or compete with data processing? Such a conclusion is hardly warranted. On the other hand, much of the data processing done to date has been mostly text processing with occasional computation.   Are we going to have another language which will be "THE" computer language?   Not at all!   Just as Fortran or Basic attack certain applications areas better than others, the Computer Languages for the processing of text (CLPT) will be better in some areas than others. At the same time, the CLPT will be used by a group of people who have never conciously used a computer. That implies notation that is much less complicated than traditional computer languages. Actually, there will be more use of functions than language in most processing of text.

Background

The project to develop standards for text processing was initiated in the mid 1960's. At that time, there were no word processors as we know them today. In fact, the processing of a full upper and lower case alphabet was rather rare in electronic processing. ASCII, the American National Standard Code for Information Interchange, had been extended by 1968 to handle much of what was needed to process text. As the work progressed ASCII was found to be a good base, but as publishers will attest, ASCII is limited. Accordingly, the text processing work was deferred until a methodology could be developed to handle the character sets needed in textual work. That work was completed in the 1970's and the text processing activity re-emerged.

To this point the discussion has centered upon the work taking place in the public domain within the area of voluntary standards. The work going forward in the market place and academic circles is not to be discounted. In fact, the resulting myriad of ways to process text or process words has increased the urgency of the work that is being done. This report is from the project established to develop the standards and does not address the academic or theoretical constraints. Elaboration upon two concepts is felt to be important. First, the work is being done by a technical team organized as a sub-committee. That team and its deliberations are open to qualified participants and the documents and records of the group are available through the Secretariat. Secondly, the work is taking place in a program of voluntary standardization. This implies that very careful work must be done to obtain the consensus required for ratification. Much more rapid progress might be obtained if a single source were utilized to gain a standard which did not require consensus, but rather offered a dictum from the proposers as is occasionally seen where standards are on a mandated basis rather than voluntary. To cite an example, certain characters are prohibited from transmission across some national boundaries. Also, the ham radio operators were prohibited to send messages containing the character set used in this paper until very recently.

Technical sub-committee X3J6 (the accronym means the sixth group in the languages area working under committee X3) has as its program of work a single project entitled "Computer Languages for the Processing of Text". That program of work shows that there will be a document in the form of a proposed American National Standard by mid 1982. The standard will address the areas of Text Description, Text Editing, Text Formatting, and Text Processing. The subjects intertwine and can not easily be separated. The standard will treat them together.

What is intended is that a single standard for Text Editing and Text Formatting will emerge. In that standard will be a means to describe the text and elements to permit processes to be established.

The Project Team

The participants are drawn from all over North America. Each has a different background and their participation is based upon an area of expertise. We have people from academic institutions, the government, computer manufacturers, software developers, publishers, trade associations, etc. The user as well as the developer are represented. There are individuals participating who have earned their living as a secretary. It is important to understand that each participant in the project team acts as an independent individual expert.

Many have an impression that there may be a "mainframer" bias. In this work, the bias seems to be in the other direction. At least five of the participants have their own microcomputer. That offers a real advantage because there is a natural tendency to try out the features that get proposed. The net of the team is that it is a very good mix of backgrounds and temperments, gathered to do a difficult job. It does its work in the public view. We meet four or five times per year devoting three days to a meeting. The homework takes another significant portion of the time of the participants. In general, the participants devote about 20 to 25 per cent of their efforts on the project. That represents a significant expenditure. The savings and advantages offered by standardization will more than provide the justification for the effort.

The Standardization Process

There are rules and procedures governing our work. The administrative efforts take about half of the first morning of each meeting and consume much time for the officers outside of the meetings. Practically all of the meeting time and most of the evening time is devoted to technical discussions. Our project outline shows that we have concluded the data gathering phase and are into the digestion and drafting phase. We are not developing a radical new language. Instead, we are adapting features from many sources while working to make the result easy to use. In doing the data gathering we have assembled documents which would be a pile of paper over three feet high if a single copy of each of them were stacked. We are aware that we do not have all of the material. Many suppliers of word processing or text processing equipment or services have opted not to allow their material to be used in the effort. That is their perogative. They will not be forced to adopt any resulting standard. Voluntary standards must rely upon acceptance to keep then viable.

Goals

The project was authorized after much discussion and effort. The clamor to reduce the proliferation of ways to do the same thing was offset by those concerned that technological advances would be stifled with the advent of standards. The goals of the project are broad. Two exerpts from the charter follow:

2.2 Goals

The languages are intended to focus on the characteristics of text and its manipulation, a capability present in standardized computer laguages with emphasis only subordinate to computation. The primary goal of the Standardization project is to provide:

(1) A Language for the Description of Text that will permit documents to be processed by multiple applications and interchanged among users and processing systems.

(2) A Computer Language for the Processing of Text which includes a set of functions for the implementation of text processing applications.

It is expected that fulfillment of this goal may lead to the development of high-level application languages for the editing, analysis and indexing, search and retrieval, and formatting of text. It is an objective that it be possible to define and implement these languages using the Computer Language for the Processing of Text (Item 2), but suppliers are not constrained in the implementation methods actually used.

2.3 Scope

It is intended that the proposed standards encompass the definition, recording, recall, entry, transformation, searching, manipulation, and display of text.

Computer text processing has an acknowledged overlap with word processing and computer graphics. In particular, word processing is expected to be a significant portion of that which might be interfaced to or use the proposed languages.

The scope of the standards may further be defined by including the description of operations which may be performed upon text. These operations, when included into languages for the processing of text, contain certain features,which permit "Programming" to be possible. The operations may be divided into five categories. These categories are:

1) Manipulation of text.
2) Control of visual form in the display of text. 3) Control of interfaces with other portions of the overall system in which the text processing is or will be taking place.
4) Administrative activities involving the creation, processing, transfer, and storage of text.
5) Those operations, which permit the text processing to become procedural in nature.

These standards must describe what is meant by text in both the philosophical and technological domains. The domain of text could include such fundamentals as word, sentences, books, editions, and the associated concepts implied.

From the above it is hoped that our goals and constraints are broadly understood. Our overlap with computer graphics is particularly a problem to us. A subordinate goal is to permit the development of documents which can be represented by a single stream of characters. This should include figures. We will not be addressing pictures, but will defer to the computer graphics standards for that material. This implies that the groups work together and they are doing so. Similarly, the computer graphics people are not hoping to do the more sophisticated composition work we will be addressing. There is an area both groups recognize, but are not yet handling. That area is the pictorial matter which is only representable by dot rasters. A half tone picture is a good example of such pictorial matter. This leads to a paradox because the so-called other end of word processing includes facsimile processing which is no more than the rasters mentioned above. We expect that the standards will address the full spectrum of text, but initially we will rely on other technologies to support our work.

A notion that has gained much attention recently is the "Open Systems Interconnection" model. This is a conceptual model of interconnected processes which function as an entity. That model is an implicit part of our design and implies that the CLPT will exist as an application within many computer environments. At the other extreme, we expect that the CLPT can be implemented on an eight-bit microprocessor with 50 000 bytes of RAM, two flexible disks, a keyboard, and a display. The only visible difference to a user of the CLPT in those extreme environmental situations is intended to be throughput or storage resource limited. Interaction with the implementation environment is a critical area. In order to maintain transportability and interchange, we are defining the languages with their own structures and not relying on the host systems. In fact, the imposition is intended against the host. By that is meant that the implementor is expected to gain the standard features in the form of the standard regardless of the features of the host system. We do not expect that the majority of the users of the CLPT will be interested in programming or the implications of integrated data processing. Compatibility with the host is expected. Sacrifice of CLPT features in order to exist in a given operating environment is not expected.

Decision Guidelines

We were favored with very few guidelines in our charter, but those we were given are very significant. Again, an exerpt is given below:

It is intended that the development project be bounded by the following guidelines:
1) Trade-offs must be weighed in favor of human factors when compared against technical generality.
2) The dominant majority of features, elements, or operations must not be restricted from the interactive or free-standing environments described above. Where implementation environmet

constraints may impact a language feature or element, the development project is expected to acknowledge such a restriction in the standards.

3) The languages must be desribed in terms which do not require educational background beyond that of a typical American High School commercial course graduate.

4) Rigorous formal descriptions of the languages should be included as a separate portion of the standards. In such portions only, the educational constraint above should be relaxed. This implies deliberately redundant sections.

Perhaps the importance of the guidelines is clear. We place special emphasis on the concept of "User Friendliness". Trade-offs between elegance and ease-of-use will always fall to the ease-of-use side.

Text Description

The Language for the Description of Text is intended to be a notation which can be added to the text in order to identify the elements of text. Of themselves, they are intended to have no particular effect. They are intended to be interpreted and serve as triggers to the text processing procedures used to present the text to readers. A term often used for this purpose is "Mark-up". The Generic Identifiers are the flags used for the purpose of marking the text. With such marking, the processing and formatting may be accomplished independently of the other processes. No attempt will be made to standardize all of the elements of documents. Such an effort would frighten away many users with the sheer enormity of the lists. Further, much research is required to find the true set of common document elements. Starting with title, author, and subject some insight is offered. Such matters as highlighting, list entry, security code, etc. offer an indication of the complexities of this portion of the work. While we do not expect to offer standard lists of mark-up flags, we do expect to standardize the syntax of the identifiers. We also expect to offer several lists or templates which will represet some document types.

A major fraction of our effort is devoted to this area. We are also coordinating with the GCCA section of the Printing Industries of America. The current expectation is that the standard will offer the same capability that they offer in our fundamental module. We exect to offer a more complicated version in our higher level of modularity. An interesting sidelight is that the mark-up technique can lead to text processing of standard marked-up text by procedures other than the CLPT. A FORTRAN or COBOL program could well be utilized if felt necessary. Such a capability is not the intent, but is inescapable when we force the generality to a Generic Identifier. It is noteworthy that the breakdown of the prototypes we have investigated have all been due to particularization rather than any problem in the flagging technique.

The Text Processing Language (TPL)

The CLPT or TPL covers all of the action functionality that is usually associated with the composite subject of text editing and text formatting. Many contemporary text processors deal primarily with one or the other of those areas. Very few offer clear deliniation between the functions. Those which do are inconsistent in where the line is drawn. As an exercise you are asked to consider how you would draw the lines between the two functional areas for your own work. After doing so test your model on the differing processes you can envision from what has been presented to this point. If you still have a clean split and have not sacrificed convenience, let us know how you made the division.

In order to simplify our labor we have made a three way ad hoc split of the work on the TPL. The split is between editing, formatting, and procedures. No division of the standard may be implied. It simply happened that there were participants whose interests fall more into one area than another. More significantly, progress is made more rapidly when the working party is small. The overlaps and missing links are the responsibility of all three groups. As our work progresses we will form other ad hoc tasks to do such work as editing the draft standard, formalize the syntax and semantics, write the glossary, etc. For the present we are taking text editing to mean the creation, manipulation, revision, alteration, and storage of text. Text formatting is oriented around the manipulation of text for the purposes of display, presentation, or printing. This are breaks down into Typewriter-like devices and Typesetter-like devices. In this context a device must be considered to be generic. With the advances in technology we may find only the more complicated of the two areas of interest by the date of publication of the standard. Only time will tell, and our work is addressed at both methods. The procedures function is a misnomer, but the terms we are using are not too much more enlightening. We are calling the ad hoc group the "Nucleus and Services" group. This title was adopted to call attention that a Text Processor would be implemeted in an environment. That implementation would require functionsality to provide the services necessary to do the editing and formatting tasks. In other words the nucleus and services activity is providing the core of the processor. As part of that core there will be a procedures function. This comes about because of the idea expressed in the work statement regarding being able to express the language in itself.

There is a mild controversy over a standard which allows options to the user. Opposition to flexibility comes from administrators and training officers for the most part. Their argument seems to be based upon the idea that rigid specifications make their jobs easier. We counter that the standard itself is not intended as a tutorial, nor is it intended to be a specification for a complete implementation. Instead, we expect to describe a set of upward compatible modules

which constitute an increasingly sophisticated text processing system. In all of the modules the idea of user customization within the standard by the use of procedures written according to the standard, is present.

Text Editing

The text editing group has studied many existing implementations. The group has the largest population of users. Their preoccupation is with ease-of-use. They face several paradoxes. The first is that cryptic notation or function keys are very difficult for a beginner, while natural language is very bothersome to an experienced user. The expected solution is to provide at least two levels of notation and transformation procedures between them. Actually a third level is contemplated, but not for routine human processing. It will be a dense "compiled" form of procedures and other processed text intended to minimize storage space and interpretation time. Another problem faced by the editing group has to do with the very different problems confronting a user when the input device varies from a hard copy keyboard and printer similar to a typewriter, through a video display having cursor controls, to a sophisticated buffered video display having much local intelligence. Luckily, the Open System Interconnection Architecture Model simplifies the latter item into the issue of buffering. So far, none of the functionality the editing team have desired will be out of place in either interactive device. Much of the material discussed by the editing group has been extensions of functionality at a lower level. The decision now facing the team is where to draw the line between intrinsic functionality and functionality derived by user or supplier written procedures.

Text Formatting

One of the largest problems facing the formatting group has been the issue of sufficiency when addressing the matter of presentation style. Typography is clearly an art form. Art forms do not lend themselves to the rigidity of standards. Luckily, much critical research has been done by Professor Donald Knuth at Stanford. He has documented much of the historical evolution of typographic practice. He has also provided insights toward techniques offering capabilities never possible in the past. As an example, text justification has been traditionally a matter taken into account on a line by line basis. Most of the reason for that was the mechanical limitations in the composing equipment. Knuth reasoned that the desire was for the most pleasing zone of appearance and the unit of that visual impact is most often the paragraph. As a result he developed a spacing system based upon the positioning of rectangular arrays.

The acceptability of some Knuth concepts in the more traditional circles may be slow in coming. Our standard will include the capabilities of his work, but will not impose them on the formatted results. A similar problem is the ongoing controversy over the issue of hyphenation. A low frustration susceptability is necessary in a worker attempting to properly address the hyphenation problem. The "standard" dictionaries are not consistent from version to version and edition to edition and phoneme driven techniques will easily split ring into r-ing. Luckily again, we have been aided by much human factors research. On the other hand, the standard will not be able to settle the controversy. What we plan is to offer the flexibility to do whatever formatting the user desires via procedures. Obviously, the capabilities of the presentation device are critical here and we are not so naive as to attempt to standardize the interface to such devices. Whatever capabilities are available must be addressed. What we will attempt to standardize is the notation and processes needed to accomplish the formatting tasks.

The previous paragraph may lead to a conclusion that we are concentrating on typography. Far from it! Much of our work in formatting is aimed at the more conventional unit spacing devices like a typewriter or video display. We are also looking at such matters as an easy way to describe a table and keep the matter of whether the table fits on a single panel isolated from the data of the table. Even within tabular matter we are attempting to offer sufficient generality that repositioning of rows and columns is easy.

Such traditional matters as run-arounds, figures, font, kerning, etc. are not being ignored. We have even found that the printer's widow has become split into a widow and an orphan to distinguish between the bottom or top of a panel.

Text procedures

There are to be procedures or programs possible in the language. We doubt that every user will wish to write them or even read them. As procedures deal with increasingly complex subjects, the level of skill needed to write or understand them will increase. We are not revolutionizing the field. At the basic level the procedures will consist of individual commands or functions which could be singly invoked in an interactive session. The next level of sophistication we see is the user who wants to take advantage of a personal bias either in their work or in their approach. Customization along parameter lines is what is envisioned here.

The full fledged programmer supplied procedures as application packages are also envisioned. We would not be surprised to see the supplier of COBOL and RPG packages to write checks in "Words" offer a similar package in TPL. It is important to realize that everybody is not expected to be able to write procedures. We do plan to make it easy for beginners to write them. In fact, we expect to permit "Unstructured Programming" because that is how many people approach their work and we do not intent to impose upon their style any more than necessary.

A procedure as we envision it is a command or function built from other commands and functions. It is much as a MACRO in assembly languages. It

will be more because it will be possible to prepare procedures which generate no text which is contrary to the conventional concept of a MACRO. Much discussion has centered upon how to describe procedures. Will they be structured according to the latest from the field of computer science? We rather doubt it. While there is much dispute over notational brevity and complexity, there is much to be said for a notation that reads closely to the way we talk or write.

System Interface

How can we discuss system interface if we are talking about a system that is implemented on a free standing microprocessor? That is dependent upon the other implementaions and our strategy. Some want a "bit level" definition to make implementations identical. That is impractical in a voluntary standards program. We do expect that a user who ends a session on a given implementation and carries the files to another implementation of equal or higher level, will be able to continue as if the continuation were on the first implementation. There will be some features that may prove difficult in some environments. If the process calls for the date or time of day and there is no capability in the system to provide that data the implementer will be forced to provide a means of entering the information via the user. On the other hand, if the data is present in the system, but formatted differently, the implementer will be expected to provide the information in the format of the standard. Similarly, if the host system desires a format different from that used by the standard, the implementer interface routines will be forced to do the transformation.

Much information about the virtual machine (the model CLPT processor) will be needed as a user session progresses. This will be provided to the processor from the nucleus process.

The Panelists

Each of the panelists is a key participant in the area the are covering. We had hoped to bring the leaders of each sub area together, but the costs were prohibitive. They will amplify the statements in this summary. More significantly they will be able to bring these remarks up to date. Our work is very dynamic and the publication logistics of a major conference make topical discourses difficult. It is hoped that the interaction of a question period will add to your understanding of our task and our better understanding of the needs of the using public.

Summary

The standardization project entitled "Computer Languages for the Processing of Text" was initiated by the American National Standards Institute Committee on Computers and Information Processing (X3). The project goals are to provide an American National Standard which will serve the broad area of text processing. The standard is scheduled to be published in early 1983 after public review and comments in the fall of 1982. While this seems to

be a long time away, the work to be done will require all of the remaining time alotted. Participation in the project is open to qualified experts in the field. The criteria for participation are simple, expertise coupled with availability of the necessary time and the financial backing to do the necessary travel to the meetings. The current members of the project team are from all over North America. Their backgrounds and employment are varied. The participants act as individual experts and are not permitted to represent corporate positions.

The standard which is planned will deal with the subjects of Text Description, Text Editing, Text Formatting, and a Procedural Language to accomodate those three subject areas. None of the work of the project team represents truly new techniques. Rather it will be a blending of functionality seen in many contemporary implementations of text processing.

The impetus for the work came from the widely varied techniques seen in the available products. There has been little capability for interchange of information and virtually no capability to interchange processing methods. The project is intended to correct those deficiencies and offer a contemporary way of processing text which will be convenient and effective. It is not intended that the user interface to text processing will become standardized. Instead, a standard notation for procedures and functionality will be established.

A significant problem raised by current techniques is that they have implicit final format specifications which are mutually incompatible. When their actions are driven by "functions" imbedded in the recorded text, the transportability to another implementation is often made extremely difficult. A text description technique and notation is planned in order to separate the description of text (semantics) from the display or presentation. This notation is allied to the notion of editorial mark-up in use outside the computerized text processing field. The differences are primarily in the scope of the notation rules and the concept that there would be procedures and system states associated with the mark-up.

Most people experienced in text processing tend to think first of a text editor when they think about the subject. Text editors have a family resemblence more in name than function. The editing funtionality associated with the standard will gather the necessary and better functions from many of the existing editors. It will offer flexibility and will permit both string orientation and line number orientation without precluding screen editing. Some of the more advanced and complex functions will be available via procedue calls rather than being inherent in the language. This strategy is being adopted in order to keep the syntax of the commonly used functions rather simple. It is felt improper to impose a clumsy syntax on users in order to allow the occasional use of a high powered technique.

The conventional functions of creating, changing, moving, duplicating, storing, inserting, deleting, comparing, and searching will all be available. Operations will be possible upon characters, words, strings, fields, lines, records, units, groups, files and collections of files.

Text formatting has been subjected to the greatest inconsistency in the implementations researched to date. This is primarily due to the motivation of the implementer. Few, when faced with the size of the effort, were willing to provide broad generality. Most often a narrow publishing style was addressed. Also notable has been a close paralleling of existing typographic practice. That is no surprise, but leads to the sudden changes in styling that Knuth propounds. The standard will permit formatting ranging from the single page or screen image through very sophisticated multiple column illustrated text. Again, as in text editing, the user will not be burdened with a notation which makes the routine difficult in order to accept the infrequent special case. Here we are fortunate. Formatting tends to be designed by specialists and used by less demanding people. Here the tools provided for formatting will be blended in order to permit those specialists to be as persnickety as they may choose without forcing the office typist to deal with the intricate details of obscure techniques. Of course, where curiosity develops the path will be there to grow without going to another process.

There is an aura of everything for everyone in the above. So far there has not been any indication that there will be a block against reaching the goal. It is quite clear that the details will stimulate new more particular techniques. That will lead to more detailed extension in order to satisfy the stylists who will finally have the capability to quickly and economically experiment with new techniques. It will be exciting to watch as years of practice involving minimizing the human effort to prepare a document are offset by the means to do easily what was previously prohibitive. Will hyphenation, abbreviations, and acronyms persist? Should they? Would such a question have been reasonable just ten years ago?

Users will not be confronted with mammoth lists of terminology to memorize. It is intended that they use only that which is necessary for the immediate purpose. The standard will, unfortunately, contain an extensive list of very technical terminology in order for the implementers to be capable of producing their copies of the process. The reference to the unfortunate circumstance is derived from past experiences with the use of standards by producers. Too often, the standard is offered as the tutorial vehicle for a new user. Many intended users of the standard will not have any reason to deal with a substantial portion of its features. In fact, a sizeable intended user base will be unaware that they are in any way involved with the standard. This will come from their use of processes programmed according to the standard without their direct access to the functions themselves.

It is intended that customization of the interactive process on an individual user basis be possible within the standard. That is not to say that such customization will be necessary. Instead, the capablity will be there when desired. The above implies an open-ended approach. Such is the intent. As time goes on, the standard can then be extended to take the more popular features into the language. That is entirely consistent with the policy that American National Standards be reviewed at least every five years for reaffirmation, revision, or withdrawal.

The work progressing in the United States is coordinated with work going forward in the worldwide standardization area. There is an Experts Group established to work on Computer Language for the Processing of Text under the Programming Languages Sub-Committee of the Computers and Information Processing Technical Committee of the International Standards Organization. In reverse order (the usual hierarchy) the acronym is: ISO/TC 97/SC 5/EGCLPT for those readers who are fans of alphabet soup. Many of the participants in the United States activity are also in the international Experts Group.

The work is coordinated also with other areas. Particular attention is being paid to the work in text communication and what is called a "Page Image Format". This work is under the Character Sets technical group with assistance from the Word Processing standards sub-committees. The work in Computer Graphics is also being coordinated with the hope that it will be possible to publish entirely from "Standard recorded digital media". Such publishing would include anything currently printed.

Will such a goal be reached? Yes! There is little doubt. Will there be a major impact on current office procedures? Yes! Again there is little doubt. That being the case, will the impact be disruptive? No! We are working very hard to provide an evolutionary way to reach the new ways of doing business. The historic failures of some major integrated data processing systems has influence our work by reminding us to permit steady change without forcing discontinuities.

This is an exciting area. We will see revolutionary new systems in the next two decades and will wonder how we lived without them when we reflect upon today's developments at that later time.

Will we see our school children taught keyboarding rather than script penmanship? Will text processing deal with spoken input? Those are both strong possibilities and not very far fetched. Look at what has happened to the Kanji writen language of the Far-East in the few years that scholars have had access to computers. We are all caught up in a movement which is exciting and thrilling. In text processing we can exploit technology toward extending our intellectual pursuits with far less risk of harmful results than in many other fields.