Education Document on Bar Code Technology

Automating document control and indexing with barcodes can dramatically improve user productivity and data integrity. In the transportation industry, bar coded tracking numbers are affixed to invoices or delivery documents, marked with notes on shortages or damage. In distribution or parts warehouses, bar codes can be affixed to a page as parts are used. Pharmaceutical companies bar code medicines and prescriptions.

In document imaging, automatic indexing using bar codes is becoming popular because OCR fails to deliver accurate conversions in typical forms-based document imaging applications. The forms get in the way, numbers and characters often are skewed, and documents are faint and sometimes dirty. Backgrounds and color interfere with the image, and it's difficult to locate the key fields automatically. As a result, OCR usually achieves successful recognition less than 80% of the time, leading to expensive index repair keying.

Bar Codes solve all these issues. As a result more and more companies are turning to barcodes to identify their documents. New bar codes such as PDF417 are 2-D bar codes that allow a user to create up to 2,000 characters of useful data. We believe because of the dramatic increase in the capacity of bar code technology, we will see a rapid increase in business adapting it into their workplace.

In fact, at the AIIM 2004 conference, Adobe Systems is introducing new forms processing technology the utilizes PDF, the free Adobe Reader and 2-D bar code technology to reduce cost and errors associated with manual data entry and provide a unified environment to support both paper and electronic forms processes. Adobe describes the functionality as follows:

> Once distributed to customers via the Web or email, the forms can be completed on or off-line using the free, Adobe Reader. As end users complete form fields, the 2-D bar code dynamically encodes the data in a format specified by the form author. Once completed and printed by the user, forms can be submitted by mail or fax. Upon receipt, organizations simply scan the bar code to capture the form data and deliver it to a back-end system for processing.

To prepare you for this market opportunity and to help you to understand bar code technology to sell our bar code software, we have created this education material for your use. Some of this information has been copied from other sources, so this material is for internal use only.

Overview

What is a bar code? A bar code is simply a series of stripes (usually black) on a light background (usually white) that can be scanned and read directly into a computer. They are interpreted virtually instantaneously and without errors by a bar code reading system. The elements (bars and spaces) in a bar code symbol must be of a consistent, proportional

thickness and thinness. The widest element could be as thick as a pencil or as thin as a business card, as long as the corresponding thin bars and spaces in the bar code remain proportionally thin.

Bar codes are read the same way that people read text from a page; the reflectance and absorption of light. A light of a given wavelength is beamed and moved across a bar code at a consistent speed. The reflected light is measured with a photoreceptor, tuned to look for light of the given wavelength. The off- and- on (white and black) pattern of the bar code creates an electrical wave that is sent on to a computer chip called a “ decoder.” The decoder then deciphers the signal into something the waiting computer understands. Imager and CCD (charge coupled device) bar code scanners read somewhat differently in that they “ take a picture” of a bar code symbol, analyze it, and create a conditioned electronic signal that basically mimics that from the reader types described in the paragraphs above.

The bar code “ symbology”

In other point of view, a bar code “ symbology” is to bar codes in much what a particular alphabet is to language. Different symbologies of bar codes use different combinations of bars and spaces to represent different characters. Bar code symbologies, like languages, are given different names, like Code 39, UPC, Codabar, PDF417, DataMatrix…

One dimensional vs two Dimensional barcodes

Before beginning a discussion on two- dimensional bar code symbols, some clarifications must be made regarding one- dimensional symbols...

Typical bar code symbols obviously have both height and breadth. To most people, they’d be considered a two- dimensional object. So, why do we refer to them as one-dimensional bar code symbols? Because, when read, it is only the width of the bars and spaces that is taken into account — the height of the bars is only to give the symbol some built- in redundancy. Typical bar code symbols, therefore, are only read in one dimension.

One other important aspect of one- dimensional bar code symbols is that they seldom represent more than a dozen characters. Therefore, the bar code does not contain any data, per se. Rather, the bar code represents the key to a record in a database, where related information is stored. The best example is a car license plate which, by itself, doesn’t mean much but, when entered into a motor vehicle database, can access all sorts of information regarding the car it is attached to.

In the Two- dimensional (2-D) symbols, data are encoded in both the height and width of the symbol, and the amount of data that can be contained in a single symbol is significantly greater than that stored in a one-dimensional symbol. In fact, over thousands alphanumeric characters can potentially be placed in a single symbol the size of a large postage stamp! Obviously, the main advantage of using 2D bar codes is that possibly a large amount of easily- and accurately- read data can " ride" with the item to which it is attached. There are new applications being created for 2D bar code technology every day.

One of the amazing (and beneficial) aspects of two- dimensional symbols is their potential durability. To sabotage the readability of a conventional 1D symbol, one only has to add another bar to the beginning or end of the symbol or draw a line through the symbol, parallel to the stripes. This throws off the checks and a balance built into the decoding algorithms of a 1D bar code decoder and makes the symbol unreadable. By comparison, many degrees of redundancy can be built into a 2D symbol. While it makes the symbol somewhat larger, the remaining symbol is remarkably secure. We have experimented with vandalizing 2D symbols with holes, black marker and tearing. The symbol has remained readable through all of this abuse!

About PDF417

It's PDF417. PDF stands for " Portable Data File." A two- dimensional symbology, a single PDF417 symbol carries up to 1.1 kilobytes of machine- readable data in a space no larger than a standard bar code. And, unlike traditional one- dimensional bar codes, which depend on real- time links to a larger database, PDF417 symbols are the database. PDF417 symbols travel on paper. Moreover, PDF417 is recognized as the standard between two-dimensional symbologies by leading organizations worldwide.

In addition, PDF417 is an error- correcting symbology designed for real- world situations where portions of labels can get destroyed or damaged in handling. It performs error correction by making calculations, if necessary, to reconstruct undecoded or corrupted portions of the symbol.

Encoding data into a PDF417 bar code is a two- step process. First, data is converted into codeword values of 0 - 928, which represent the data. This is " high- level encoding." Then the values are physically represented by particular bar/ space patterns, which is " low- level encoding." Decoding is the reverse process.

At first glance, a PDF417 symbol looks like a set of stacked bar codes. When we look closer to analyze how the symbol is put together, there are several key elements. Those will be explaining below.

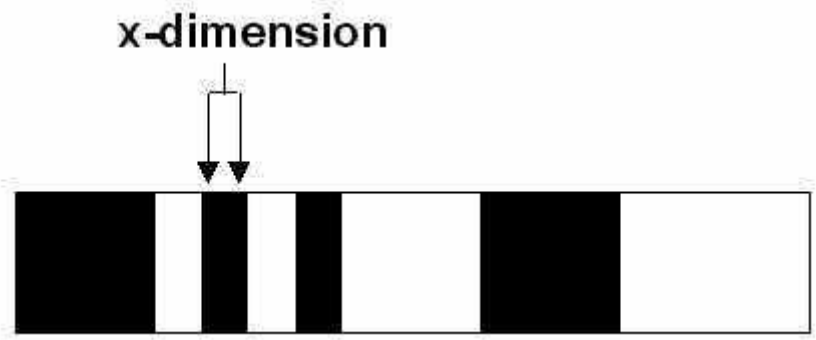
Bars and Spaces

All bar codes are comprised of bars and spaces (dark and light regions). A bar is a continuous dark area; a space is a continuous light area. From here on, we will refer to bar and/ or space as simply bar. Most bar codes have a fixed number of possible bar widths. That is, each and every bar must be one of a fixed number of sizes. Each bar's width must also be a multiple of the smallest bar width. For instance, if the narrowest bar is 10 mil, then possible bar widths can be 10 mil, 20 mil, 30 mil, 40 mil, etc... But not, 15mil, or 25mil since these sizes are not multiples of the narrowest bar size of 10 mil. (mil, stands of thousands of an inch).

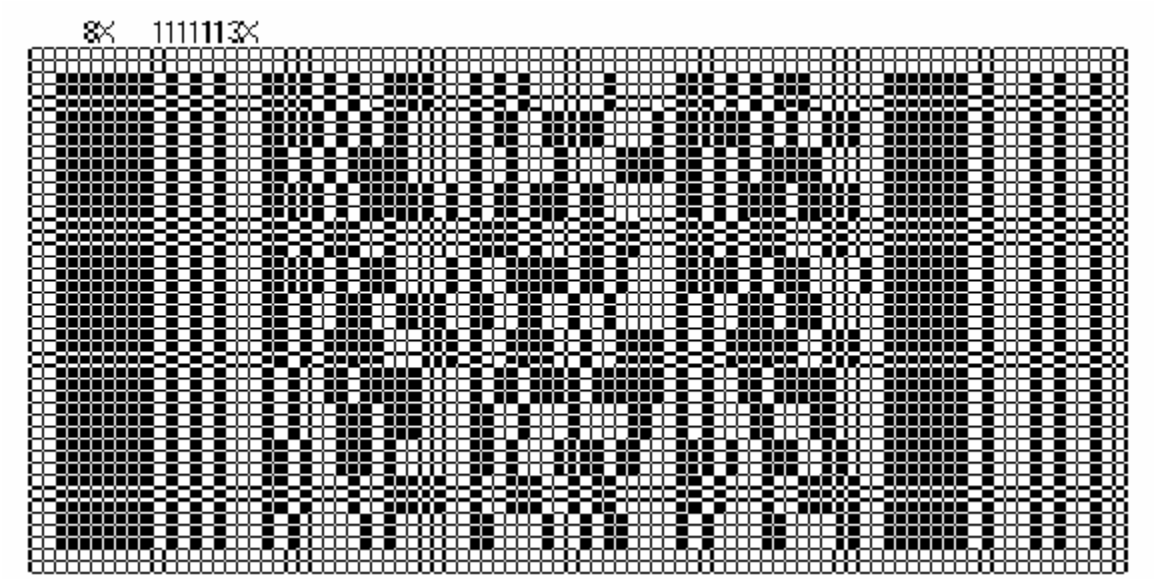
X Dimension

The width of the smallest bar is defined as a bar code's ' X' dimension. Each X' dimension is sometimes referred to as a module. In the following picture, you can easily pick out the smallest, or narrowest bar. By measuring this bar, we can determine the bar codes ' X'

dimension. You can also see that all of the other bar widths is multiples of the smallest bar or ' X' dimension. Each bar's width is often expressed relative to the ' X' dimension; for instance, 3X refers to a bar that is 3times as wide as the narrowest bar.



The following picture is a PDF417 bar code that looks like it is printed on graph paper. This image makes it easier to count the number of ' X' dimensions, or modules in each bar. If we look closely, we will see that each and every bar is an exact multiple of the minimum bar width. The first bar is 8X wide, the following space 1X wide, etc... In a validly printed code, without ink spread, you should never see a bar that is 1.5X, 4.2X, or not a whole multiple of X. In a PDF417 symbol, you will always see 1X, 2X, 3X, 4X, 5X, 6X, 7X, or 8X bars. Keep this in mind, you will see later how to detect, and avoid printing problems related to this topic.



Start / Stop Patterns

Every bar code has a start pattern on the left, and a stop pattern on the right. These patterns are unique for each type of bar code. PDF417' s unique start and stop patterns are:

Codewords
In a PDF417 bar code, each bar and space does not store data. Data is actually stored in codewords. A codeword is a consecutive sequence of 4 bars and 4 spaces totaling 17X wide. The 417 in PDF417 refer to this codeword structure. Codewords reside between the start pattern on the left, and the stop pattern on the right. This region contains several types of codewords including data codewords, control codewords, and row indicator codewords. Each of these codewords abides by the rules described above. The picture below points out a single PDF417 codeword:

Data Codeword Region
User data is first encoded into codeword values. These codeword values are then converted into physical codewords represented by bars and spaces as described above. Data codewords are physically located between the left and right row indicator codewords. Below, you can see the Data Codeword Region:

Rows
If you look closely at a PDF417 symbol, you will notice that it appears to be made of many " 1D- like" bar codes. In reality, it is made up of multiple rows. A PDF417 bar code can have anywhere from 3 to 90 rows. This allows a PDF417 symbol to be reshaped by adjusting the number of rows. The following PDF417 symbol has 5 rows:

Columns

A PDF417 symbol is made up of multiple data columns, which are sometimes referred to as the data column area. The number of data columns can vary from 3 to 30, to accommodate user's real estate requirements. These columns contain encoded data, as well as error correction information. Within the data column area, a single PDF417 can contain no more than 928 codewords. An example of a 3 column PDF417 symbol appears below:

Error Correction

The PDF417 symbology has error correction capability. This capability enables scanners to read the bar code even if it has been torn, written on, or damaged in other ways. How much damage a symbol can withstand depends on the amount of error correction in each PDF417 symbol. The user has the ability to select 1 of 9 error correction levels for each symbol printed. Error correction is specified by selecting a level from 0 to 8. At level 0, a damaged PDF417 cannot be read, but the damage can be detected. At levels 1 through 8, a PDF417 symbol can still be read, even when damaged. As the error correction level increases, more damage can occur to the symbol and still be read. Consequently, the higher the error correction level, the larger the symbol becomes, while the data capacity goes down. The following table illustrates these facts:

Error Correction Level	Error Correction Codewords	Error Correction Capacity	Maximum Text Capacity
0	2	0	1850
1	4	1	1846
2	8	3	1838
3	16	7	1822
4	32	15	1790
5	64	31	1726
6	128	63	1598
7	256	127	1342
8	512	255	830

Error Correction Capacity
PDF417 symbols can be damaged and still decoded. The amount of damage that a symbol can withstand is its error correction capacity. For example, at level 5-error correction, 64 codewords of error correction are used. At this level, 31 codewords can have errors, while still being read correctly. If more than 31 errors exist, the symbol is unreadable.

X and Y Dimensions
The X dimension is the width of the narrowest bar in a printed codeword. The Y dimension is the height of each row within the PDF417 symbol. PDF417 is usually printed with an X to Y ratio from 1: 2 to 1: 5. By changing the X to Y ratio to 1: 2 or 1: 3, a significant amount of space can be saved.

Row and Column Limits
The number of rows and columns can be selected to place the symbol in a specific form. The PDF417 symbol is limited to 30 columns and 90 rows. When selecting columns, keep in mind that you are only selecting the number of data columns in the symbol. The normal PDF417 symbol has two row start columns and two row stop columns. Truncated PDF417 contains only two row start columns.

Truncated PDF417
A truncated PDF417 symbol is more area efficient than normal PDF417. By selecting this option, the right hand side of the symbol is removed or " truncated". This option should only be used in clean environments, since it is less immune to damage.

Compaction Modes and Amount of Data Encoded
We recommend limiting the amount of data encoded in 2D symbols to 1700 characters if possible. Although the AIM PDF417 specifications state that " up to 1100 bytes or 1800 ASCII characters can be encoded in a PDF417 symbol", we have found that these numbers are not realistic. The amount of data that can be encoded will vary depending

upon the type of data, the compaction type, the error correction level chosen and what your scanner can read. In text compaction mode, the amount of compaction varies due to mode switching between different types of characters, such as between numbers, upper case, lower case and punctuation.

Because each codeword represents 1 of 929 possible values, data can be compacted into the codewords to save space. PDF417 has three data compaction modes: byte, text and numeric. Byte compaction encodes actual bytes of data, text mode encodes most characters on the US keyboard and numeric compaction encodes only numbers. The chart below shows more details about these modes.

Compaction Type	What can be encoded	Maximum Compaction
Byte Compaction	ASCII 0 to 255 plus Data Bytes	1.2 bytes per codeword
Text Compaction	ASCII 9,10,13 & 32-127	2 characters per codeword
Numeric Compaction	only numbers 0-9	2.9 digits per codeword

PDF417 translates into productivity
PDF417 technology translates into productivity. Easy to use and easy to integrate, a PDF417 application delivers the benefits of digital data communications with the simplicity of bar coding. It's very low in cost… printable in various sizes on a wide variety of media using traditional printing technologies…easy to support…and highly robust in its error correction capabilities. Moreover, PDF417 has been recognized as the standard among two-dimensional symbologies by leading organizations worldwide.

PDF417: Much More than Meets the Eye
PDF417 answers the need to capture, store and transfer large amounts of data inexpensively. It can exchange complete data files (such as text, numerics or binary) and encode graphics, fingerprints, shipping manifests, electronic data interchange (EDI) messages, equipment calibration instructions and much more. It provides a powerful communications capability – without the need to access an external database. And, for virtually no incremental cost, you can add a PDF symbol to the documents and labels you are already printing.

Think of PDF as an independent database with complete freedom of movement, traveling together with a person or on an item, object, package, form, document, card or label. It does what wired networks can't: allows you to immediately access your data regardless of location. Plus, encryption is available as an option when additional security is required.

Moreover, because PDF417 is a machine-readable method of transporting data, it eliminates time-consuming and error-prone manual data entry. It functions as a paper-based computer memory that can be written once and read over and over again. And, as a universal machine language, it communicates with all host operating systems. PDF417 encodes full ASCII, numeric or binary data and it uses sophisticated error correction algorithms to keep intact 100 percent of the data – even when as much as half the symbol

is damaged. And it's self-verifying, so data errors can be detected and data integrity maintained.

What a Difference a Dimension Makes
One-dimensional bar codes contain an access code that serves as a real-time key for opening a database. A PDF417 symbol contains a complete data record and requires no access to an external database. Data, text, graphics, biometrics and voice records are immediately applied to the application transaction by simply scanning the symbol.

A Technology You Can Trust
PDF417 is the de facto standard 2-D symbology. In addition to its performance capabilities and application value in a wide range of industries, PDF417 technology is in the public domain–placed there by Symbol–and it conforms to industry and international "open" standards. For all of these reasons, PDF417 is emerging as the standard 2-D symbology by leading standards-setting organizations:

- The American National Standards Institute (ANSI) has just published a new 2-D standard: ANSI MH 10.8.3M unit loads and transport packages for two-dimensional symbols. It recommends the use of PDF417 for all shipping, receiving and supporting EDI documentation. Most standards associations will use this ANSI standard as the basis for their specific industry's application specification.
- AIM USA and AIM Europe, accredited ANSI standards-developing organizations, which have approved and published PDF417 as a Uniform Symbology Specification (USS) standard.
- The U.S. Department of Defense, which has designated PDF417 as the standard 2-D bar code for logistics applications and EDI formatting on paper labels.
- The American Association of Motor Vehicle Administrators (AAMVA), which has approved and published PDF417 for a broad range of driver and motor vehicle applications.
- The Automotive Industry Action Group (AIAG) 2-D Applications Committee, which has selected PDF417 for its key 2-D production and logistics applications, is completing their B-10 standard for shipping and receiving.
- The Telecommunication Industry Forum (TCIF) is reviewing PDF417 as the 2-D standard for product marking.

In addition, the U.S. Department of Defense, Office of Personnel and Readiness, has issued millions of military identification cards with PDF417 symbols for use on a global basis.

Verified in Independent Testing
Not only is PDF417 technology hard at work in a wide range of important applications, it has also been proven in independent testing. In reliability testing by the Ohio University Center for Automatic Identification, 32 million PDF417 characters were read without an error. Tests at the University of Pittsburgh proved PDF417 compatibility in high-speed overhead scanning and its readability even when symbols were damaged.

Technology that's Proven in Diverse Applications
Just about everyone in the private and public sectors is in the business of information, communication, identification and data management. PDF417 symbology serves all of these needs and has been proven in a highly diverse range of end user applications.

PDF: Expanding on Your Bar Code Capabilities
The two-dimensional benefits of PDF417 symbols are ideal for applications that are limited by the natural constraints of 1-D bar codes. The amounts of data needed as well as the environment it exists in are key deciding factors when choosing to implement a system using 2-D symbols. A single scan of a PDF417 symbol can easily replace multiple 1-D bar codes while delivering more information more quickly.