The Genre of Mathematics Writing and its Implications for Digital Documents

Eleanor G. Rieffel
FX Palo Alto Laboratory, Inc.
3400 Hillview Ave., Bldg. 4
Palo Alto, CA 94304
rieffel@pal.xerox.com

Abstract

The genre of mathematics writing has several distinctive features that point to some of the weaknesses of current digital documents. Some of these weaknesses are surprising. While it might be expected that the importance of formatting and special symbols in mathematics writing would pose challenges for digital documents, the linked, chunked style of mathematics writing, with its Theorems, Lemmas, Corollaries and Remarks explicitly referring to each other, resembles standard hypertext so closely that one would expect that mathematics writing would take well to online hypertext form. It does not. This failure points to deficiencies in our understanding of the true strengths and weaknesses of digital documents.

This paper describes mathematics writing, with particular emphasis on features of interest with respect to digital documents. The difficulties in producing effective digital mathematics documents are then examined and used as a basis for talking about general challenges for digital documents. The paper then discusses strengths of digital documents and some of the problems that need to be overcome before digital documents can live up to claims made for them. It also examines some of the misguided claims, such as superior support for non-linearity, that are commonly made for digital documents, explains why these claims are unwarranted, and speculates on why the claims have been made anyway.

Suggestions are then given as to what the true benefits of digitization are, including performing computations on text, flexible control of time, and better support for hiding information. The paper concludes with a list of questions whose answers are critical to understanding the capabilities, and therefore the future, of digital documents.

1. Introduction.

For many years, work on digital documents was dominated by hypertext. Recently hypertext has become less fashionable, partly because it failed to live up to many of the promises made for it, but partly also because equating hypertext with digital documents was misguided and harmful to both. The claims that the chunked and linked style of hypertext is new and revolutionary, and that it is the computer that makes non-sequential text possible, or at least removes severe limits imposed by paper, are both patently false. Furthermore, the obsession with non-sequential text and with links has distracted from the true benefits of digital texts.

I should state at the outset that there are a number of successful digital documents that go under the name of hypertext that have broken with traditional forms and do require the computer for effective implementation. However, the success of these documents has nothing to do with the chunked and linked style of static hypertext, or with the non-linearity of the text, but rather with computer-enabled dynamic properties of these texts.

I am not the first author who has discussed the red herring of non-sequentiality. [15] However the misconception continues to be widespread and persistent. I hope that the explicit chunked and linked style of mathematics texts, dating back as far as Euclid's Elements of 300BC, and the depth of their hypertext structure will provide a particularly clear and convincing example that this style of writing is far from new and has nothing to do with computers.

Furthermore the successes and frustrations of current digital mathematics documents can give insight into the true strengths and weaknesses of digital documents in general, which will in turn point towards fruitful directions for the advancement of digital documents.

The paper begins with a discussion of the structure of traditional mathematics documents, common definitions of hypertext, and the wrong-headedness of viewing hypertext structure as the principal advantage of digital documents.

0-7695-0001-3/99 $10.00 (c) 1999 IEEE

Current mathematics documents are then discussed with an eye to both the benefits and shortcomings of these documents. When appropriate, non-mathematical digital documents will also be discussed to further illustrate these points. Fruitful directions for the advancement of digital documents are identified, and the short-term feasibility of major advances in these areas is estimated. The paper concludes with a list of questions for future research, whose answers would provide insight into the future of digital documents.

2. The structure of mathematical texts and its implications for digital documents.

The claim is often made that digital documents, and hypertexts in particular, provide superior support for non-sequential reading, though few go so far as to say as does Bolter that "in the computer, writing in layers is quite natural, and reading the layers is effortless."[1] People tend to jump around while reading mathematics even more frequently than when reading other texts. Mathematical proofs fall naturally into a complex structure, and readers jump around following different strands in the argument. Also, since mathematics is precise and human memories tend to be vague, people jump around a lot in the text to refresh their memories. Mathematical reading is notoriously difficult. Perhaps it was waiting for the supposedly superior support of digital documents?

Unfortunately digitalization, while helping the communication of mathematics substantially by enabling easy and fast transmission of mathematics documents, has done remarkably little to aid the reading of mathematics. To begin with it is difficult to put mathematics, with its special symbols and careful formatting, into many existing digital systems. But even when the formatting problems are overcome, the results are disappointing.

The truth is that traditional printed mathematics texts provides substantial support for non-sequential reading. Mathematical notation and writing conventions have had a few thousand years to develop. Given that non-sequential reading of mathematics is so common, it is not surprising that text structures assisting such reading have evolved.

2.1. The structure of mathematical texts

Scholarly mathematics papers and mathematics textbooks are made up of chunks of text, usually between one line to one page long. Most chunks are labeled and many are numbered. Common labels include "Theorem," "proof," "Definition," "Lemma," "Corollary," "Proposition," "Example," "Remark." The numbering conventions vary. See Figure 1 for a typical example [9] of labeled chunks in mathematics text. There are also unlabelled chunks. These fall into two categories: short connectors between labeled chunks and texts that serve as motivation or an overview. Proof chunks occasionally run for more than a page, and motivation or overview chunks often do.

264 CHAPTER V FIELDS AND GALOIS THEORY

(ii′) Since every intermediate field E is algebraic over K, the first paragraph of the proof of Theorem 2.5(ii) carries over to show that E is Galois over K if and only if E' is normal in $\mathrm{Aut}_K F$.

If $E = E''$ is Galois over K, so that E' is normal in $G = \mathrm{Aut}_K F$, then E is a stable intermediate field by Lemma 2.11. Therefore, Lemma 2.14 implies that $G/E' = \mathrm{Aut}_K F/\mathrm{Aut}_E F$ is isomorphic to the subgroup of $\mathrm{Aut}_K E$ consisting of those automorphisms that are extendible to F. But F is a splitting field over K (Theorem 3.11) and hence over E also (Exercise 2). Therefore, every K-automorphism in $\mathrm{Aut}_K E$ extends to F by Theorem 3.8 and $G/E' \cong \mathrm{Aut}_K E$. ■

We return now to splitting fields and characterize them in terms of a property that has already been used on several occasions.

Definition 3.13. *An algebraic extension field* F *of* K *is* **normal** *over* K (*or a* **normal extension**) *if every irreducible polynomial in* K[x] *that has a root in* F *actually splits in* F[x].

Theorem 3.14. *If* F *is an algebraic extension field of* K, *then the following statements are equivalent.*

(i) F *is normal over* K;
(ii) F *is a splitting field over* K *of some set of polynomials in* K[x];
(iii) *if* $\overline{\mathrm{K}}$ *is any algebraic closure of* K *containing* F, *then for any* K*-monomorphism of fields* $\sigma : \mathrm{F} \to \overline{\mathrm{K}}$, $Im\ \sigma = \mathrm{F}$ *so that* σ *is actually a* K*-automorphism of* F.

REMARKS. The theorem remains true if the algebraic closure $\overline{K}$ in (iii) is replaced by any normal extension of K containing F (Exercise 21). See Exercise 25 for a direct proof of (ii) $\Rightarrow$ (i) in the finite dimensional case.

PROOF OF 3.14. (i) $\Rightarrow$ (ii) F is a splitting field over K of $\{ f_i \in K[x] \mid i \in I \}$, where $\{u_i \mid i \in I\}$ is a basis of F over K and f_i is the irreducible polynomial of u_i.

(ii) $\Rightarrow$ (iii) Let F be a splitting field of $\{ f_i \mid i \in I \}$ over K and $\sigma : F \to \overline{K}$ a K-monomorphism of fields. If $u \in F$ is a root of f_j, then so is $\sigma(u)$ (same proof as Theorem 2.2). By hypothesis f_j splits in F, say $f_j = c(x - u_1)\cdots(x - u_n)$ $(u_i \in F; c \in K)$. Since $\overline{K}[x]$ is a unique factorization domain (Corollary III.6.4), $\sigma(u_i)$ must be one of $u_1, \ldots, u_n$ for every i (see Theorem III.6.6). Since σ is injective, it must simply permute the u_i. But F is generated over K by all the roots of all the f_i. It follows from Theorem 1.3 that $\sigma(F) = F$ and hence that $\sigma \in \mathrm{Aut}_K F$.

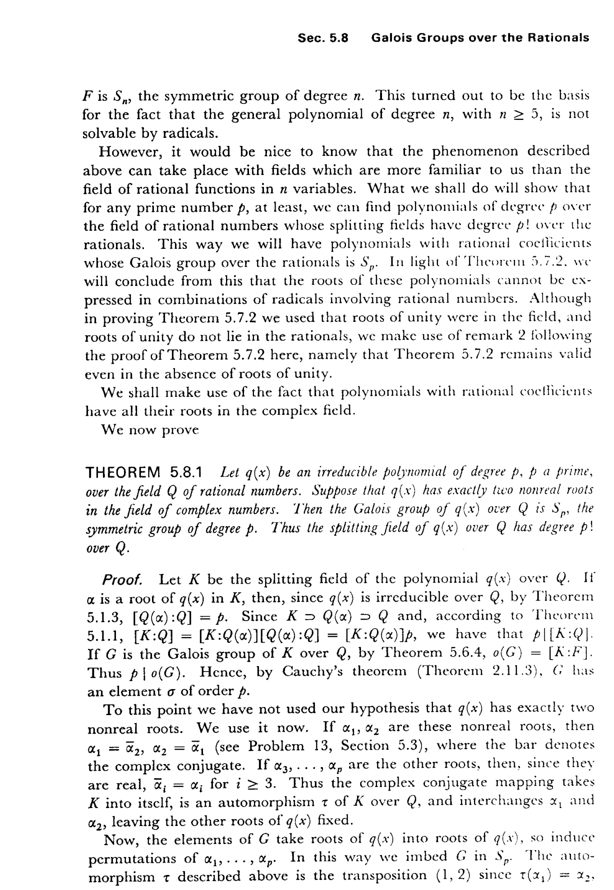
(iii) $\Rightarrow$ (i) Let $\overline{K}$ be an algebraic closure of F (Theorem 3.6). Then $\overline{K}$ is algebraic over K (Theorem 1.13). Therefore $\overline{K}$ is an algebraic closure of K containing F (Theorem 3.4). Let $f \in K[x]$ be irreducible with a root $u \in F$. By construction $\overline{K}$ contains all the roots of f. If $v \in \overline{K}$ is any root of f then there is a K-isomorphism of fields $\sigma : K(u) \cong K(v)$ with $\sigma(u) = v$ (Corollary 1.9), which extends to a K-automorphism of $\overline{K}$ by Theorems 3.4 and 3.8 and Exercise 2. $\sigma \mid F$ is a monomorphism $F \to \overline{K}$ and by hypothesis $\sigma(F) = F$. Therefore, $v = \sigma(u) \in F$, which implies that f splits in F. Hence F is normal over K. ■

Figure 1. Reprinted by permission of Springer-Verlag © 1974 [9]

Propositional Statements are generally followed by proofs. The proofs contain explicit links to previous Propositional statements. The density of these explicit links can be seen in Figure 2. The linked structure can be quite complex, as can be seen by Figure 3, a diagram of part of I.N.Herstein's Topics in Algebra [8], a commonly used upper division undergraduate text. Figure 3 was generated by starting with Theorem 5.7.3, finding all of the explicit links within its proof to other statements, and then finding all the explicit links within the proof of these statements. This process was continued until either there were no explicit links contained within the proof of a statement, or the explicit links were broad, for example "referring the reader to Chapter 2 for the proof." [8] Note that since certain propositional statements are used in the proof of multiple statements (for example, Theorem. 5.6.2), the structure is not a tree.

Chunk type:	**Common labels for chunks of that type**
Labeled chunks:	
Propositional Statements	Theorem, Corollary, Proposition, Lemma
Proofs	Proof
Non-derivable Statements	Definition, Remark, Example
Unlabelled chunks:	
connectors	
motivation, overview	

Table 1: A classification of chunk types in mathematical writing.



Sec. 5.8 Galois Groups over the Rationals

F is S_n, the symmetric group of degree n. This turned out to be the basis for the fact that the general polynomial of degree n, with $n \geq 5$, is not solvable by radicals.

However, it would be nice to know that the phenomenon described above can take place with fields which are more familiar to us than the field of rational functions in n variables. What we shall do will show that for any prime number p, at least, we can find polynomials of degree p over the field of rational numbers whose splitting fields have degree $p!$ over the rationals. This way we will have polynomials with rational coefficients whose Galois group over the rationals is S_p. In light of Theorem 5.7.2, we will conclude from this that the roots of these polynomials cannot be expressed in combinations of radicals involving rational numbers. Although in proving Theorem 5.7.2 we used that roots of unity were in the field, and roots of unity do not lie in the rationals, we make use of remark 2 following the proof of Theorem 5.7.2 here, namely that Theorem 5.7.2 remains valid even in the absence of roots of unity.

We shall make use of the fact that polynomials with rational coefficients have all their roots in the complex field.

We now prove

THEOREM 5.8.1 *Let $q(x)$ be an irreducible polynomial of degree p, p a prime, over the field Q of rational numbers. Suppose that $q(x)$ has exactly two nonreal roots in the field of complex numbers. Then the Galois group of $q(x)$ over Q is S_p, the symmetric group of degree p. Thus the splitting field of $q(x)$ over Q has degree $p!$ over Q.*

Proof. Let K be the splitting field of the polynomial $q(x)$ over Q. If α is a root of $q(x)$ in K, then, since $q(x)$ is irreducible over Q, by Theorem 5.1.3, $[Q(\alpha):Q] = p$. Since $K \supset Q(\alpha) \supset Q$ and, according to Theorem 5.1.1, $[K:Q] = [K:Q(\alpha)][Q(\alpha):Q] = [K:Q(\alpha)]p$, we have that $p \mid [K:Q]$. If G is the Galois group of K over Q, by Theorem 5.6.4, $o(G) = [K:F]$. Thus $p \mid o(G)$. Hence, by Cauchy's theorem (Theorem 2.11.3), G has an element σ of order p.

To this point we have not used our hypothesis that $q(x)$ has exactly two nonreal roots. We use it now. If α_1, α_2 are these nonreal roots, then $\alpha_1 = \bar{\alpha}_2$, $\alpha_2 = \bar{\alpha}_1$ (see Problem 13, Section 5.3), where the bar denotes the complex conjugate. If $\alpha_3, \ldots, \alpha_p$ are the other roots, then, since they are real, $\bar{\alpha}_i = \alpha_i$ for $i \geq 3$. Thus the complex conjugate mapping takes K into itself, is an automorphism τ of K over Q, and interchanges α_1 and α_2, leaving the other roots of $q(x)$ fixed.

Now, the elements of G take roots of $q(x)$ into roots of $q(x)$, so induce permutations of $\alpha_1, \ldots, \alpha_p$. In this way we imbed G in S_p. The automorphism τ described above is the transposition $(1, 2)$ since $\tau(\alpha_1) = \alpha_2$,

Figure 2. Reprinted by permission of John Wiley & Sons, Inc. Copyright © 1964 [8]

The actual link structure stemming from Theorem 5.7.3 is significantly more complex than the graph shows for a

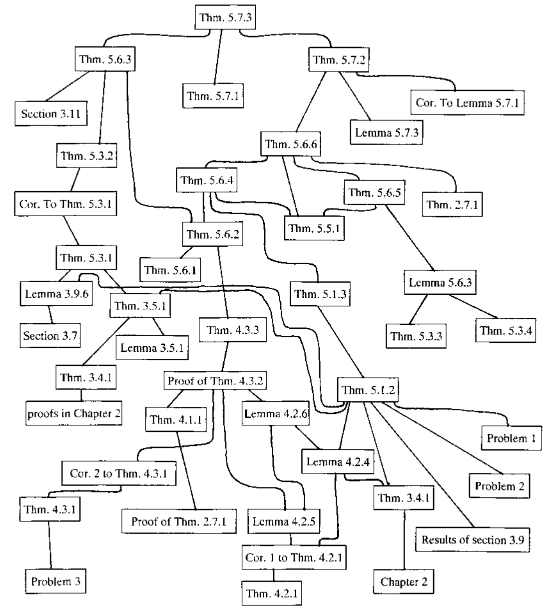


Figure 3.

couple of reasons. The first is that since the diagram already showed great complexity, I did not show the links stemming from references to larger sections of the text than the ones I identified in the chart above. From these lengthy sections, graphs of equal or greater complexity to the one shown would have been generated. Also I did not include any references to definitions since here we begin to get into the cloudy issue of what in a paper document is really analogous to a link in electronic hypertext, and the case was strong enough without them. But certainly some of the references to definitions are analogous, which would add significantly to the complexity of the structure here, especially since many definitions refer to previous definitions.

Some mathematics texts provide maps of the interrelation of the text, like those in Figure 4 [9], that prefigure those seen in some hypertexts. Formatting is used extensively in mathematics text to support non-sequential reading. Many of the chunks of text in Figure 1 are not only labeled but space is put around them. Some are indented, some are in other fonts. These chunks have been carefully formatted so that chunks can easily be skipped or located. For example, in the majority of cases a statement of a theorem is followed by a proof. But the two are carefully separated, and the end of the proof is marked, as in Figure 1. Thus it is easy for readers to locate the statement of a theorem, or to skip a proof. Words that are being defined are highlighted so that a reader can more easily locate the definition. The careful formatting indicates that the text expects readers to jump around.

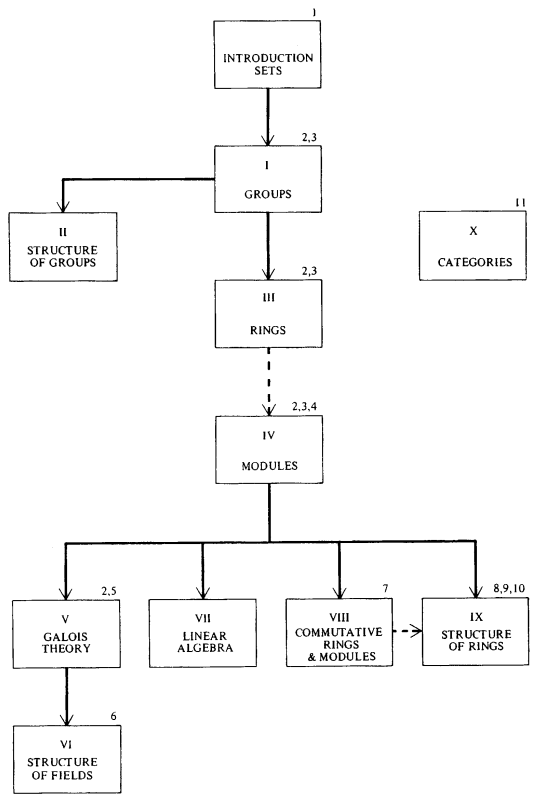


Figure 4. Reprinted by permission of Springer-Verlag © 1974 [9]

Bolter writes about hypertext that “In one sense this is simply the electronic equivalent of the footnote used in printed books for hundreds of years. ...The machine is merely handling the mechanics of reading footnotes. But there is this important difference: the second window can also contain boldface phrases that in turn lead the reader to other paragraphs. The process can continue indefinitely.... The second paragraph is not necessarily subordinate to the first. A phrase in boldface may lead the reader to a longer, more elaborate paragraph. One paragraph may be linked to many and serve in turn as the destination for links from many others. In a printed book, it would be intolerably pedantic to write footnotes to footnotes. But in the computer, writing in layers is quite natural, and reading the layers is effortless.”[1] It should be clear from the description of the structure of mathematics writing that the structure that Bolter describes is exactly the structure that has existed for years in printed books of mathematics. These “footnotes to footnotes” are not only tolerated by the mathematics community, but have been adopted as the standard by which mathematics communication takes place. There is nothing new to these structures.

2.2. How hypertext has been defined.

Bolter writes that “it is the ability to create and present hypertextual structures that makes the computer a revolution in writing.” [1] This emphasis on hypertext structure as the principal advantage of digital documents is a common one. To evaluate the merit of this claim, let us look at what is meant by “hypertext.”

“Hypertext Defined: By hypertext I simply mean non-sequential writing ... Computers are not intrinsically involved with the hypertext concept. But computers will be involved with hypertext in every way,....” [11]

“All traditional text, whether in printed form or in computer files, is sequential, meaning that there is a single linear sequence defining the order in which the text is to be read.... Hypertext is non-sequential .... hypertext consists of interlinked pieces of text “ [13]

“In hypertext, information is organized as a network in which nodes are text chunks ... and links are relationships between the nodes ...” [15]

“A hypertext is like a printed book that the author has attacked with a pair of scissors and cut into convenient verbal sizes. The difference is that the electronic hypertext does not simply dissolve into a disordered bundle of slips, as the printed book must. For the author also defines a scheme of electronic connections to indicate relationships among the slips.” [1]

These definitions all represent a standard view of what hypertext is, which I will call *static hypertext*. As we shall see, some documents are labelled hypertext documents even though they are far from fitting any of the above definitions. Nevertheless the definition in terms of non-sequential text and authored chunks of text connected by explicitly authored links prevails. To add to this confusion, in some cases authors make implicit, or even explicit claims, that certain texts fit the above definitions when they do not. We will see examples of such documents and such claims in the next section.

Let us look at some of the unwarranted assumptions in the statements given above:

1. Traditional texts are authored to be read sequntially.
2. The computer supports non-sequential reading better than paper.
3. Support for non-sequential text structures, consisting static text chunks connected by author-defined links, is the principal advantage of digital documents over paper ones.

Assumption 1 is patently false, as the discussion above of the structure of mathematics texts shows with particular clarity. Mathematics texts call into question even lesser assumptions like "computers make possible more complex forms of hypertext than are possible on paper." As the discussion of the structure of traditional mathematics texts makes clear, sopisticated hypertext structures have been the norm for scholarly mathematics writing for centuries.

Other people have argued this point before, but without the clear-cut example of mathematics texts. Andrew Dillon argues this point particularly eloquently in his paper "Myths, Misconceptions, and an Alternative Perspective on Information Usage and the Electronic Medium"[15] in which he cites others that have made the same point. Many more people have pointed to the chunked and linked style of encyclopedias, newspapers, and commentaries. But these examples tend to have a less explicit, less dense, or less deep linked structure than mathematics texts, or don't form a coherent whole in the same way most texts do. For example, the main text of commentaries does not reference the commentary, so there is a strict limitation on the kind of linked structure that can develop. A good case can be made that the majority of books have implicit links and hypertext structure, but such arguments lead into the confused ground as to what in a standard text actually constitutes a link or non-sequential structure. For example, does the name of a person link back to the initial description given of them in a novel? Do two descriptions of the same event given by two different characters in a mystery story link to each other? The explicitness of the chunks and links in mathematics text allows us to avoid this confusion in that case, while lending support to the general argument.

Moving on, the main support for assumption 2, that the computer supports non-sequential reading better than paper, has been the misconception that paper cannot support sufficiently rich static hypertext structures. The structure of traditional mathematics texts provides strong evidence against the limitations of paper in this regard. Furthermore, any static hypertext document can be converted to a paper one which maintains the links by numbering all lines of the document in any convenient order, putting a box around the link anchor, and putting a margin note on the side giving the lines the link goes to. Other conventions are possible. The interface may not be as convenient to use as the online one, but it is still easy to use for reasonably sized documents.

This conversion exercise leads to a couple of interesting points about the advantages of digital form. One is the ability to have readily available documents of such size that physical manipulation or storage of paper versions of such documents would be difficult or impossible. But this is a difference of scale, not structure, between digital and paper documents. Also, there is much greater difference between the ability to create a highly linked document with or without the use of a computer, than there is between the convenience of using the paper or the online linking mechanism. As will be seen in the next section, relatively few mathematics documents are be read online, but most mathematicians use the computer to aid in the creation of their documents. Thus, if the assumption had been that the computer supports the writing of non-sequential texts better than paper, I would have had little quarrel with it.

While the online linking mechanism may be more convenient to use than the paper one, paper has many other advantages which include better facilities for other aspects of non-sequential reading, like scanning and skimming. Furthermore, digital static hypertext is more sequentially restricted that a paper version because it is much harder than on paper to do anything other than follow author-provided links. A key question for digital documents is how to provide mechanisms for sequences that have not been put in place by the author. There are many theories as to under what circumstances people prefer to read from paper and why, but the reasons are complex and, as of yet, poorly understood. [5] But there are compelling reasons for continued widespread use of paper. How people actually read in general is poorly understood. As a result it is hard to evaluate the effectiveness of any system claiming to support reading. Andrew Dillon makes these points eloquently and at length in his book <u>Designing Usable Electronic Text</u> [5]. I will not repeat his exposition here.

Altogether there is little evidence to support any claim stronger than that digital documents provide minimally better support for a certain type of non-sequential reading than do paper documents, and even this statement is debatable. Given the feebleness of this claim, we must hope that assumption 3 is also false, and that there are much more compelling advantages of digital documents than their support for non-sequential text structures. In the next section, we will look at more and less successful digital documents, particularly digital mathematics documents with an eye to what the true advantages must be. Before doing that, however, I will briefly speculate on the reason for the misplaced emphasis on non-sequentiality and the persistence of this belief.

Computers can store huge amounts of information, and networked computers can store even larger amounts of information. But in its raw form all of this information is hidden. Mechanisms for accessing it must be built in. Contrary to popular belief computers are naturally better at hiding information than they are at providing access to it. As people began to write documents on the computer, they struggled with means for enabling readers to move from one part of the document to another. Because scanning and skimming are so much more difficult on the computer than on paper, the fact that people want to jump around within a text came to people's attention. And thus began the obsession with non-sequentiality. It began, not because computers supported non-sequentially well, but because they supported it so badly it got noticed!

3. Digital mathematics documents and their implications.

In a weak sense, digital mathematics documents are a common form of communication within the mathematics community. Many mathematics papers are posted to preprint servers, like the Los Alamos preprint server at http://xxx.lanl.gov/ which has thousands of preprints with thousands of people connecting everyday. Similarly, many mathematicians post copies of their papers on their web pages. And the vast majority of mathematicians have become fluent with some variant of the TEX typesetting program like LATEX, which numbers propositional statements and updates the references to these statements automatically, as well as formatting equations properly. Thus most mathematics papers are prepared online. However, as these papers will be downloaded and printed before they are read, it is debatable as to whether they truly constitute digital documents.

A small number of mathematics papers can be viewed on the web. However, the limitations of HTML make the inclusion of mathematical symbols and formulas awkward, time-consuming, and difficult to do well. Problems include slow delivery, sizing, spacing, and centering difficulties, and the inability to print, depending on the method used. For a discussion of these problems, see http://forum.swarthmore.edu/typesetting/index.html. To see some examples of the problems, with varying severity depending on your browser and its settings, see the papers at http://www.geom.umn.edu/docs/research/. Scholarly mathematics documents that can be viewed on the web remain rare. Even the Geometry Center of the University of Minnesota, which has been at the forefront of advances in technology supporting mathematics on the web, has a preprint server of papers that can be viewed only after being downloaded.

CD-ROM versions of books suffer from similar problems, and thus, while they have been tried, they remain rare. See, for example, Jet Wimp's review of a CD-ROM version of Gradshteyn, et al's Table of Integrals, Series and Products, a book the reviewer says he "adores." [16] He describes the formatting difficulties that mean "You can't see the formulas!" in spite of the fact that it was produced using "the opulent and flexible text display software called DynaText 2.3." [16] But the formatting difficulties weren't the only problems: "What else could go wrong? Something else did. You can't *find* things. ...You have to know where something is to find it. ... the honest to god paper book, has to be visually scanned too, but it is all in front of you: nothing is buried under something else." [16]

More successful are web sites that contain tidbits of mathematics for which multimedia illustrations are helpful. See, for example, the web site http://java.sun.com/applets/archive/beta/Pythagoras/index.html which gives a proof of the Pythagorean theorem through animation. Similarly, the site http://aleph0.clarku.edu/~djoyce/java/elements/elements.html contains interactive figures for Euclid's Elements which are particularly helpful for visualizing the proofs in three dimensions.

Most common are hybrid documents, printed books or papers that have an attached CD-ROM or an affiliated web site that contains animations or interactive components that help illustrate or promote fluency with concepts in the text. For example, the CD-ROMs may enable the user to run an algorithm on their favorite numbers, or to graph their favorite function. However the interface between the text and the relevant component of the CD-ROM is often clumsy. Those of us interested in digital documents should take a broad view and look at ways of improving these hybrid documents as well as purely digital documents. As paper as well as digital form has its advantages, we may be dealing with hybrid documents for a long time. We need better mechanisms for linking the digital and paper worlds together.

We now discuss some benefits of digital documents and the possibility of further improvements in the future. Other than availability, these benefits all have to do with computer enabled dynamic properties of the text, from the inclusion of dynamic components, to dynamic changes to the text itself.

3.1. Ready availability and ease of distribution.

Ready availability of documents has been the greatest benefit of digitization for the mathematics community.

Ready availability of huge documents, impossible on paper, should not be confused, however, with more complex document structures. This confusion is common in the hypertext literature. For example, as his first example of hypertext in Hypertext and Hypermedia, Nielsen describes the MIT Media Lab's NewsPeek system, a system that "watches the nightly television news for you and records those parts that it knows are of interest to you." [13]. Two things are striking about this system, neither one having anything to do with the chunked and linked structure or Nielsen's definition of hypertext. One is that it can gather information quickly from a vast quantity of news programs and newspaper stories. The digital information is readily available. The other wonderful thing about this system is that, through the use of automated algorithms, the computer can filter the input to present information tailored to the reader's desires. We will discuss this benefit of digital documents in the subsection on Tailoring.

As digital documents tend to mean ones that are read online, not just available online, we will not discuss availability further.

3.2. Inclusion of multimedia and interactive components

Digital documents enable the integration of video, audio, graphics, interactive elements, and text into a tightly interconnected whole in a way that was not possible before computers. More generally, digital documents provide effective ways for authors and readers to go between precisely timed media, like video and audio, and media, like text, that are not. Furthermore, for computer simulations, digital documents can integrate explanations of the model with these models themselves. Hybrid digital/paper documents will continue to be common as both have different strengths. Better interconnections between the digital and paper components are needed.

However integrating multimedia can have adverse effects on digital documents as well. There is a proliferation of color, graphics, and animation in web pages and CD-ROMs that add nothing to the content. This trend towards glitz becomes increasingly hard to reverse as more and more people develop an idea of what a web page or CD-ROM is "supposed to look like." One effect of this trend is that would-be authors need expertise in graphic design skills as well as writing. In many cases this overburdens the author, resulting in less efficient information flow. For these reasons, authors who see a possibility for greater expressiveness using digital form, may nevertheless restrict themselves to the medium of print. This trend can hardly help but have an effect on digital documents in general.

3.3. Tailoring to the reader.

Almost everyone writing about the advantages of digital documents, particularly for education, talks about how digital documents can be tailored to a reader's goals, background, interests, style, etc. For example, DeVault and Kriewall write in the *Yearbook of the National Society for the Study of Education* that "it is easy to agree with Suppes when he says that computers offer the only real hope for providing learning experiences that are individually tailored to the unique needs of each pupil. The question seems to be mainly one of time."[6] Similarly Roger Schank suggests that for education "computers play an important role, for they make individualized attention a real possibility." [15] Jay Bolter states generally that "An electronic book can tailor itself to each reader's needs." [1] Given how much this advantage is talked about, it is surprising how few documents provide tailoring mechanisms, and how primitive these mechanisms are in the documents that do provide them.

Some effective customization exists. The NewsPeek system mentioned above is one such system. And many systems provide buttons that make the font larger on all the pages or change the background globally. But few mechanisms are provided that change the content. Here are a couple example applications that would greatly benefit mathematics texts.

Example 1. The examples and exercises related to certain concepts could match the reader's interests or goals. For example, the exercises and examples related to a certain aspect of linear algebra could be geared towards the application of interest to the reader, say computer graphics, quantum mechanics, or micro-economics. Along similar lines, students in a language class could all learn the same grammar while reading text that was geared to their own interests.

Example 2. Many mathematics books, and other technical books, begin with familiar objects and notation and move toward highly abbreviated notation which would have meant nothing to the reader if it had been presented first. Such texts provide difficulties for readers who want to dive into the middle of the text, or wish to skip certain sections. Digital documents could support such readers by keeping track of what a reader had read and inserting a full explanation of a term the first time the reader sees it and then gradually weaning them away from the explanation towards the standard jargon or abbreviation.

There are three main obstacles that prevent widespread customizable digital documents: poor natural language processing and production, difficulties determining a reader's preferences, and the tension between customization and ease of sharing. Some valiant attempts have been made to produce customizable documents. See for example [4], [10], [12]. But in each of these systems, achieving some level of customizability comes at the price of burdening both readers and authors. Authors must create heavily annotated and tagged text for the computer to process, and readers must manually input their preferences.

Many proponents of static hypertext claim that customization can be achieved by providing sufficiently many links. Such a solution is naive. Say we have a text chunk on the habits of rabbits and we are trying to anticipate where the reader might want to go next. They may want more details about something on the page, independent confirmation of what is said, more facts about mammals in general, or cuter pictures of rabbits. Perhaps the text reminded them of the white rabbit in Alice in Wonderland, or perhaps they just realized that “rabbits teeth” would make a great name for a rock band and are wondering if someone has already used it. The possibilities are endless. Even providing links to a large number of the most reasonable or likely of these possibilities will overwhelm the content.

A better partial solution is to dynamically customize the link structure through keeping track of what has been read and the interactions of the reader with the document. See for example [3] and [7]. Golovchinsky’s system is particularly attractive since it manages to provide relatively sophisticated tailored links with only minimal and natural input from the reader. Search mechanisms are another kind of dynamic linking mechanism, needing at the best of times only minimal input.

The overall goal should be to design digital documents in such a way that readers can quickly and accurately get at the information they want, including information they weren’t explicitly looking for. In many cases people are willing to spend longer scanning for information than formulating a search and waiting for it to be carried out. The reason is that they are more likely to find something of interest along the way when scanning than when formulating and waiting search results. For this reason (and others less admirable) people often flip through books searching for information instead of using a table of contents or index which might seem more efficient. Authors of digital documents must guard against enabling readers to find what they are explicitly looking for more quickly, but decreasing the amount they learn per unit time. When we pick up a book to read it in the simplest and most conventional sense of reading, we are making a leap of faith that the author will take us in reasonable manner towards our goal, or at least to something else of interest. The beauty of well-written text, whether on paper or on machine, is that the interest/time ratio is high.

3.4. Restricting access to parts of the document.

Digital documents differ from printed documents in the control of access. Many authors of digital documents struggle to give readers a sense of what a document contains and how to access it, and the difficulty of doing so is often discussed in the hypertext literature. But few people mention that this difficulty can be an advantage in that it allows the author to hide parts of the document from readers. Video games make great use of this ability, gaining many coins from people repeatedly playing the beginning levels because that is the only way to gain access to higher levels they are curious to see. But restricting access can be just as valuable in documents. People use documents as much to filter out information as gain information, and authors, as the experts on the subject, are often in a better position than readers to decide what is most important to know, and in what order to present it. The NewsPeek system discussed earlier hides more information than it provides access to.

The strength of Michael Joyce’s “Afternoon” is its adventure game like structure with its author-defined restrictions of access which dynamically change as the reader interacts with the document. The digital form enables authors to take more control over where ... readers go” [2] and close off or open up parts of the document depending on what has been read. Again, it is not the chunked and linked structure of “Afternoon” that makes it interesting, but rather its dynamic properties, although it is possible to read the many pages in which Bolter describes this work [1] without realizing its dynamic nature, or that it does not fit the picture of static hypertext given above. The non-sequentiality of “Afternoon” can be done in book form. The dynamically changing access cannot.

Choice can be bewildering. Hiding inappropriate or unlikely choices can be a benefit. While in many situations it is better to allow the reader to decide what they will read and in what order, there are a greater number of instances when the author should suggest a route. Andrew Dillon cites a study of Edwards and Hardman which showed that people often feel lost in hypermedia documents not because they know where they want to go but not how to get there, but because they do not know where to go next.[5] Giving the reader choices when appropriate has long been done in conventional text of all sorts. The computer does not make the difficulties of good writing any easier.

Simpler applications of hiding information exist. For example, mathematics books handle question and answer awkwardly; it is easy to see an answer to a question you have not yet gotten to, when looking up the answer to another question. Since hiding parts of the text is easy in digital documents, they can handle simple question and answer more elegantly than books.

3.5. Feedback

The superiority of digital documents in dealing with feedback is even more pronounced when the feedback required is more sophisticated than simple question and answer. Some books present problems, then hints, then solutions. Multiple levels of feedback can be handled more smoothly on a computer. For instance, the first level of feedback might merely let the reader know that the answer he or she provided was wrong, allowing the reader to go back and try to solve the problem more carefully. Then hints could be provided, and finally a solution, if necessary.

4. Conclusions.

Mathematics documents, both traditional and digital, provide insight into the potential benefits and the limitations of digital documents. Traditional mathematics texts illustrate exceptionally clearly that deep and dense static hypertext structures have existed in printed text for years. The recognition of this fact should encourage people interested in digital documents to avoid following the red herring of non-sequentiality, and to instead look at the potential for various types of dynamic behavior possible only in digital documents. Existing digital mathematics documents point to fruitful directions for further exploration including support for flexible control of time, tailoring, access restrictions, and more sophisticated feedback. They also point to the importance of examining ways of improving hybrid digital and paper documents.

5. Questions for future research.

I end with a list of those questions towards which the research community should concentrate more effort if it wishes to understand how best to advance the future of digital documents.

1. How can we increase the interest/time ratio for people accessing information, and not just increase the speed and accuracy with which they retrieve information specifically searched for?

2. How do people actually read, including skimming, scanning and information location? What support can be provided for effective reading? What sort of devices can we design that will aid reading, or support the authoring of documents that can be most effectively read?

3. Why are color, graphics, and animation viewed as necessary elements of web pages? Is digital pure text a culturally viable medium? What can we deduce about the future of digital documents and communication?

4. How can electronic documents better exploit human's ability to use information outside of his or her conscious focus? A better understanding of this process is needed.

5. What mechanisms can be provided for digital documents that enable readers to follow paths not explicitly put in place by authors.

6. Why aren't explicit static hypertext structures more common in writing other than mathematics? Are there reasons why this might change? What can we deduce about the structure of future documents?

6. References.

[1] Jay Bolter, *Writing Space: The Computer, Hypertext, and the History of Writing*, Lawrence Erlbaum Associates, Publ. Hillsdale ,1991.

[2] Jay Bolter, Michael Joyce, John B. Smith and Mark Bernstein *Getting Started with Storyspace for Windows*, Eastgate Systems, Inc, Watertown, MA, 1996.

[3] Licia Calvi and Paul DeBra, "Improving the usability of hypertext courseware through adaptive linking" in the *Proceedings of the Flexible Hypertext Workshop*, Macquarie Computing Reports, Macquarie University, August 1997.

[4] B. DECarolis and A. Martellotta, "GeNet: a system to generate adaptive hypermedia from discourse plans" in the *Proceedings of the Flexible Hypertext Workshop*, Macquarie Computing Reports, Macquarie University, August 1997.

[5] Andrew Dillon, Designing Usable Electronic Text: Ergonomic Aspects of Human Information Usage, Taylor and Francis, London, 1994.

[6] DeVault and Kriewall, *Yearbook of the National Society for the Study of Education* (1970) as quoted in Roland Haun, et al., *Computerized Instruction in Mathematics versus Other Methods of Mathematics Instruction Under ESEA Title I Programs in Kentucky*, Bureau of School Service Bulletin, College of Education, University of Kentucky, Lexington, Vol. XLV, No. 1, 1972, p.42.

[7] Gene Golovchinsky, "What the Query Told the Link: The Integration of Hypertext and Information Retrieval" in *Proceedings of Hypertext 97*, ACM Press, 1997, pp. 67-74.

[8] I. N. Herstein, *Topics in Algebra*, John Wiley and Sons, New York, 1964.

[9] Thomas Hungerford, *Algebra*, Springer-Verlag, New York, 1974.

[10] C. DiMarco, G. Hirst and E. Hovy, "Generation by selection and repair as a method for adapting text for the individual reader" in the *Proceedings of the Flexible Hypertext Workshop*, Macquarie Computing Reports, Macquarie University, August 1997.

[11] Ted Nelson, *Literary Machines*, published by Ted Nelson, 1981.

[12] Mick O'Donnell, "Variable length on-line document generation" in the *Proceedings of the Flexible Hypertext*

Workshop, Macquarie Computing Reports, Macquarie University, August 1997.

[13] Jakob Nielsen, *Hypertext and hypermedia*, Academic Press, San Diego, 1990.

[14] Roger Schank, *"Active Learning through Multimedia"* in *IEEE MultiMedia*, Vol. 1, No. 1, Spring 1994, p. 70.

[15] Jean-Francois Rouet, Jarmo Levonen, Andrew Dillon, and Rand Spiro, *Hypertext and Cognition*, Lawrence Erlbaum Associates, Publ., Mahwah, MJ, 1996.

[16] Jet Wimp in Reviews of *The American Mathematical Monthly*, Vol. 104, #4, April 1997, pp.373-376.