SOFTWARE—PRACTICE AND EXPERIENCE
Softw. Pract. Exper. 2003; **33**:975–1001 (DOI: 10.1002/spe.537)

Explaining code for publication

Harold Thimbleby*,†

University College London Interaction Centre, London, U.K.

SUMMARY

The computer science literature discusses code and algorithms extensively, but not always reliably. Tool support can help ensure integrity between code and explanation so that published papers are more reliable.

A versatile, lightweight tool to support explaining code for publication is justified, described and compared with alternatives. The tool works with Java, C and similar languages, and provides support for publishing explanations of real code in LaTeX, XML, HTML, etc. Copyright © 2003 John Wiley & Sons, Ltd.

KEY WORDS: documentation; Java; Javadoc; LaTeX; literate programming; loom; publishing code; warp; XML

1. INTRODUCTION

A basic view of how science is done is that scientists do research; they keep records in laboratory books; they then submit selected material and reasoning to journals, which brings their key discoveries and ideas to the attention of the rest of the scientific community. With computers, of course, keeping and managing laboratory books and publications can be partly automated: word processing and other tools are widely used.

Despite the possibilities of computer support, the quality of published program code in scientific publications is often low. Numerous algorithms and code extracts published in refereed journals and books contain errors, even to the point of not being compilable as published. Some algorithms are published in pseudo-code, but the informality of pseudo-code risks errors being overlooked or new errors being introduced when the pseudo-code is translated by the readers of the papers to real programming languages.

This paper surveys the relevant issues in scientific dissemination as they apply to computer science, and the role and relevance of tools to support the explanatory process. Section 2 of this paper, then, establishes the rationale for tool requirements and, as a side effect, should also encourage computer scientists to take code explanation tools seriously.

*Correspondence to: H. Thimbleby, University College London Interaction Centre, 26 Bedford Way, London, WC1H 0AP, U.K.
†E-mail: h.thimbleby@ucl.ac.uk

Received 27 April 2001
Revised 11 March 2003
Accepted 11 March 2003
Published online 25 June 2003

Just as there is a continuum from laboratory notebooks to polished published results, there is a continuum from explaining entire programs, as required for instance for internal documentation, to explaining specific code and algorithms for journal publication. Literate programming and related tools address the laboratory notebook end of code explanation, though they can still be useful for publication. Section 3 of this paper reviews such tools.

The main practical contribution of this paper then follows. Section 4 is a discussion of the requirements for publication tools and a comparison of two such tools. A new tool, warp, is introduced and explained, and compared with an existing (but little known) alternative, loom.

This paper is followed by a complete stand-alone appendix, 'An improved insert sort algorithm' (by Olli Nevalainen, Timo Raita and Harold Thimbleby), presenting a new variant of insert sort, which used the tool warp to ensure the published code is accurate and reliable. The example is presented separately from this paper so that the value of warp can be assessed in context without reference to the discussion of this paper interfering—publication tools like warp must be transparent to readers who do not know or have in mind the details or the rationale for using them. (One of the main reasons why literate programming is unsuitable for general publication purposes is that its conventions need explaining to readers.) The appendix following this therefore looks quite unexceptional: it looks like an ordinary *Software—Practice and Experience* paper. That is the point: it demands nothing new of the reader—but using the tools described in *this* paper made the quality control process for its authors considerably easier and more reliable; in turn, its readers may have more confidence in it.

The ideas expressed in this paper are not limited to explaining code reliably, though this is the main area of concern: the approach has applications in any areas where publications include formal text or pictures, in fields as diverse as chemistry to theorem proving, and even in papers where the formal text is as simple as just requiring correct arithmetic. Further applications include generating user manuals or diagrams automatically from data or processes defined in programs.

2. PUBLISHING COMPUTER SCIENCE

Science, understood most generally, is concerned with finding explanations of the universe and its phenomena [1], and to do so it is guided by principles, such as Occam's Razor, and a conviction that reality is fixed. As a matter of fact, we cannot ensure reality conforms to our explanations [2]; but in computer science, at least, we have an advantage. In computer science we design and build the objects we explain, and we can change them—and perhaps we explain them again. As computer scientists, we can in principle ensure that *our* explanations are correct. We can do so by changing the explanations, changing the objects, or by establishing and maintaining relations between objects and explanations so they remain consistent.

As we know, however, there is both good and bad computer science; much computer science suffers from poor explanation. Indeed, there is good and bad science in the real world beyond computer science. Good science, specifically, is distinguished by being reproducible or, as John Ziman puts it, *consensible* [3]: its claims should be expressed clearly and precisely so that others can either give their assent or find well-founded objections. Others should be able to build directly on published work, with a minimum of guesswork or interpretation.

Computer scientists write programs and explain programs—whether to document them for other programmers, to explain them in the computer science literature, or to write manuals, or to provide help for users. As Abelson and the Sussmans put it in their classic *Structure and Interpretation of Computer Programs* [4], 'programs must be written for people to read, and only incidentally for machines to execute'. Alan Perlis (who wrote the preface for their book) wrote much earlier in 1966 of a 'firm conviction that fluency in writing algorithms for one another and reading those written is a fundamental property of a professional...'[5].

Publishing code reliably means that people can use the code directly and benefit from its correctness; it also means that people can independently check its correctness, efficiency, portability and so on. Informal means of publishing code, particularly using pseudo-code, are inadequate; errors are spread in the literature, work must be unnecessarily duplicated, and when an error is detected one does not know whether this is caused by a fault in the original work, a fault in authoring the paper, a fault in its printing, or a fault in understanding and implementing its ideas. Fortunately, as this paper shows, better approaches are possible and the process can be automated, and hence ensure to much higher standards the reliability of published code.

Dunham [6], writing before the World Wide Web provided alternatives to conventional journal publication, argued that programs should be published in full: full publication of program code enables reproducible experiments; and the exposure of full publication encourages people to write better programs. Conversely, without publication there can be little or no use of the algorithms, because nobody knows exactly what they are. Furthermore, when algorithms are stated vaguely or in informal language it disguises the difficulty of their real implementation. We agree with his sentiments, but not that programs should necessarily be published in full: part of the contribution of a scientific paper is the selectivity of the author in chosing the key pieces of code to publish. Indeed, a clear fragment of code, as might usefully be presented in a paper, is not necessarily one that is optimal or works nicely in a properly self-contained way under a particular operating system, with error recovery, user interface, on-line help and so on.

Surprisingly the whole enterprise of reliably publishing code in the computer science literature has received scant attention, with perhaps only two notable exceptions. Attention to the issues was focused during the early debates on whether the programming language Algol 60 was appropriate to publish algorithms [5]; and a renaissance of attention arose with literate programming, which was popular over a decade ago. We will look more closely at literate programming later in this paper.

2.1. Publishing code

Programming and explaining are activities that involve human intervention, typically using word processors. Since explanations of algorithms or programs often refer to code, parts of the documentation or explanation will be similar if not exactly the same as parts of the program. Thus there are opportunities for computer support to help explain code more reliably. Even when paper is used (which, in fact, the current *Software—Practice & Experience* author guidelines require) best practice, namely appropriate tools and procedures, can still be used to ensure that what is printed is correct and accurately reflects the original code.

Programs are usually written in plain ASCII text, but documentation usually has a special form. If a WYSIWYG word processor is used, program code needs editing to fix the font, alignment, point size and so on; if a mark-up language (like XML, HTML or LaTeX) is used, then various program symbols need converting to the mark-up language's conventions so that they can appear properly in the published document. For example, the C programming language symbol '&' has to be edited to

'`&amp;`' for HTML or to '`\&`' for LaTeX—and of course this mark-up is no longer valid in compilable programs, so accurate conversion to documentation cannot readily be confirmed with the direct help of a compiler. Even JavaScript, a programming language *designed* to be embedded in HTML files, does not help: JavaScript cannot be printed as HTML, nor is HTML-formatted JavaScript code valid JavaScript that can be run.

The program Expect [8] can be used to check whether examples generated by running programs remain correct while the programs being explained are modified. Don Libes, the author of Expect, notes that writing about a program forces the author to reexamine implementation decisions: and he rewrote parts of Expect when he realized that parts of his explanation were overly complex for no good reason [8]. There are even examples where 'programs for humans' have also been improved by trying to explain them clearly: for example, shop signs explaining discount rules were tested and improved, but greater improvements were made to readability by changing the rules so they could be more easily explained [9]. With proper tool support, improvements to the code can be reflected in the documentation without further effort on behalf of the author. However, without tool support, an author might be tempted to avoid improving code because to do so would mean not just improving the code but also revising the documentation further—a tedious process.

For writing journal papers, which typically only use very small fragments of code, the overhead of learning or building tools may seem out of proportion to the advantages they might provide. When one starts to write a paper—say, for a conference deadline—the salient goal is usually to submit on schedule: thus building tools to automate a comparatively small part of the process is an unwelcome diversion. Typically, then, a fragment of program code is cut-and-pasted from a working program into the journal paper (if the code fragment is small, it may even be retyped *in situ*). The code is then edited carefully to conform to the documentation system's requirements (and to the typographical requirements, particularly line length and indentation of the journal or conference proceedings). This is rarely the end of the matter, as the resulting explanation will be read and revised, no doubt repeatedly.

Inevitably some changes will be made, say, to improve the style of the program for the explanation. For example, the documentation might read better if a name is changed, or some comment is added or altered ... this is the thin end of the wedge, and a subtle threat to integrity. The first may be a trivial change, so it perhaps isn't worth going to the trouble of making the corresponding change in the actual program. As time goes by, some changes will not be made consistently and some not at all. Eventually the explanation and original code will diverge to a point where there is no simple way to reconcile the differences.

The papers about Quine are an illustration of this problem. Quine is a system to generate accurate documentation, and the papers were written in Quine itself to guarantee their accuracy. Unfortunately the publishers edited the original paper [10] to make it look better for their journal, but their edits made it incorrect. The errors introduced were sufficient to justify republication [11]. This unfortunate outcome, as the republished paper noted, illustrates how the *entire* process should be automated, not just the author's part!

In comparison, using pseudo-code seems like an attractive option for explaining code, but even that approach is not without problems. Take the Chinese Postman Problem [12], which is closely related to the Travelling Salesman Problem, and has many applications. Many references to it describe the algorithm in a mixture of English and mathematical steps: for instance, the algorithm sketched in reference [13] cannot be made to work, if at all, without very close reading of the rest of the paper, plus expertise in network flow algorithms. Some publications (e.g., [14,15]) describe the algorithm entirely in English prose, so at least there are no unexpected sources of confusion, but it is unhelpful for people who want correct, executable algorithms. Some (e.g., [16]) just sketch the main steps; indeed its presentation is, in its own words, 'rather informal'. Some (e.g., [17]) provide a mixture of mathematics and English, and provide code for most of the relevant routines, but the code is difficult to translate into complete and correct algorithms (e.g., because what may look like a simple variable is in fact a non-trivial dynamically bound expression). The definitive review of the state of the art in the field [18], which has 19 leading contributors, gives no formal algorithms, and in one place says of certain efficient algorithms, 'According to our knowledge these very refined implementations were never materialized into machine-executable codes. . . '. One wonders whether they would still consider them 'very refined implementations' if they tried running them!

In short, pseudo-code has become the established mode of publication for the Chinese Postman Problem, and so even the very extensive literature on it is not an effective starting point for getting a reliable algorithm. The algorithms that have been published can at best be called *quasi-algorithms*, for they clearly fail the test of definiteness required in computing science [19]. Skiena's comprehensive *Algorithm Design Manual* [15] concludes that you have to do it yourself.

Another example of the problem of pseudo-code is the case of Porter's stemming algorithm. Stemming is finding canonical forms of words; for instance, programmable→program, programming→program. Porter's algorithm was originally published using a non-programmable notation [20], and this led to a proliferation of incorrect implementations. As Porter himself says, 'Three problems seem to compound: one is a misunderstanding of the meaning of the original algorithm, another is bugs in the encodings, and a third is the almost irresistible urge of programmers to add improvements. [. . .] Researchers frequently pick up faulty versions of the stemmer and report that they have applied "Porter stemming", with the result that their experiments are not quite repeatable. Researchers who work on stemming will sometimes give incorrect examples of the behaviour of the Porter stemmer in their published works' [21].

Rather than present clearly explained algorithms, it is easier to say that a program is available (e.g., on the Web); yet as Porter says, to extract an algorithmic description from source code of a complete program is hard. The original author could save themselves much effort and work by not preparing the algorithm for clear explanation (and hence publishes faster). But which details of their program are essential? Which are accidental bits imposed, say, by their implementation environment? Indeed, this was the problem with Porter's classic paper: providing complete source code was not an adequate explanation of the algorithm. It takes considerable work to polish an algorithm for clear explanation. If it takes considerable work to polish an algorithm, we need tools to make that work as easy and as reliable as possible.

I myself have published papers in this journal and elsewhere that fail to reach the standards I would now advocate. Taking just two recent publications of my own in this journal: I published an algorithm in an unimplemented pseudo-code [22], and, in another paper, I used manual cut-and-paste plus reformatting to present algorithms in Pascal [23]. My experience (publishing 92 algorithm-related papers to date) is probably representative of the standards to which many authors work. In almost every case my approach to preparing papers was relaxed—but it was not required to be otherwise. Referees for ACM, BCS, IEEE and other peer reviewed journals were happy with my sloppy standards. In my case the few exceptions were papers either written in systems I myself wrote (e.g., Quine [11]) or were papers (e.g., [24]) written in *Mathematica*, a system that combines word processing and programming, and hence permits the paper and the program to be more-or-less the same thing (*Mathematica* is reviewed below).

Even in a recent paper in *ACM Transactions on Computer Human Interaction* [25], which presents the entire code relevant to the paper in an appendix, we formatted the code by hand (manually marking it up in LaTeX) to make it conform to the journal's requirements. There is no guarantee that the appendix is correct, other than that we proof-read it conscientiously! As an author I made the code available on a Web page, which at least avoids the risks of rekeying for readers who wish to recover the code. The ACM journal did not require this.

2.2. Comparisons with other fields

Is this relaxed attitude to publishing programs actually a problem?

Because scientists are human and fallible, science has developed a collection of ethical principles, including the requirement that data reported should be authentic and reliable, that it should be obtained in ways that are described fully enough to be replicatable, and that claims made should be clear enough to be testable by others. Different disciplines apply these ideas in different ways: for example, the *Proceedings of the National Academy of Sciences* requires authors to make unique material (specifically including computer programs) available to non-commercial researchers.

In the field of cognitive psychology, John Anderson and Christian Lebiere promote a 'no-magic doctrine'. Their view is that the dissemination of theories in psychology traditionally relied on a sympathetic audience to understand how the theories should be applied in practice [26]. They go on to cite some of their own past work that exploited this sympathetic culture that accepted loose descriptions of research in publications. But their current programme of research now adheres to the no-magic doctrine. The doctrine is broken down into six tenets, not all of which concern us here, but which for instance require that there should be detailed and precise accounting of data. It is no coincidence that their no-magic doctrine is enabled by the use of computer simulation in psychology: the formal nature of computers allows experimental situations to be precisely specified and shared with the research community. They take advantage of that fact to improve the quality of their published work.

The sciences that have reacted to the problems of reliability of published explanations are older and more established than computer science. Of all fields, mathematics has the longest history of publishing precise reasoning, and its notion of proof lies close to our notion of algorithm. As in computer science, where there are some programs that are only described by their results, and others that can be concisely explained in a short paper, mathematics has some proofs that are too long to be understood and others where proofs are short and elegant. Fermat's Last Theorem and the Four Colour Theorem, to say nothing of proofs of computer programs themselves, are examples that highlight the differing opinions on what constitutes a reliable explanation of a proof for mathematicians [27].

As in mathematics, then, and in other areas (such as weather forecasting), there are complex computer systems that in principle cannot be explained concisely but which are, nevertheless, useful contributions to science: we do not need to—and should not—impose the requirements for explaining algorithms reliably on any and every publication. However, when an algorithm *seems* to be presented as explicit code, every step should have been taken to ensure the explanation is as reliable as it appears. Publication in any discipline ideally exposes to the community an explanation that should be able to be reliably assessed for what it apparently stands for.

Unfortunately, in computer science itself, as we saw above, it has become pretty routine to describe programming ideas without publishing, let alone depositing, the relevant code, programs or underlying algorithms for community access.

We must now ask why, and then see what can be done about it.

Partly, low publishing standards have become accepted because of the commercial value of programs—it has become inappropriate for commercial reasons to publish code [28]. Partly, because programs are large. (Neither of these issues inhibits biologists from making data available to the research community.) Partly, because it is easier to describe in unqualified terms a program that, perhaps, does not *quite* work in the general ways implied. Partly, because debugging is practically a permanent state of affairs, and adequate version control is a nightmare—particularly since the relation between code and its explanation in a publication is implicit and not addressed by automatic version control tools[‡].

It has become routine not to publish accurate code partly because computer science has developed a culture where radical honesty [29] is neither valued nor expected. It is actually *extremely* difficult to get programs working correctly, and why bother when the essential idea seems perfectly clear? The existing literature is now training researchers as to the minimal levels expected; it is not helping to improve standards, and it is not helping to introduce new approaches or new tools to increase the reliability of the future literature.

Partly it is because computer science is in a hurry. In the 1960s, George Forsythe argued that the publication of algorithms constitutes an important part of scholarship, but he warned that the process of writing, polishing, refereeing, editing and publishing can hardly be undertaken in less than two years [30]. Even if we can work a little faster today, we still require the programming notations we use to have a stable life of around five or so years for the final publication to have much value—but, with few exceptions, the only stable programming languages are the ones that are not widely implemented or widely used!

(One of the referees of this paper, perhaps noticing the opening of the previous paragraph, said that this paper's intended reader was 'the author in a hurry'. Although good authoring tools speed authors up, and therefore encourage authors to take less time preparing papers, this is not the purpose. The purpose of the tools described later in this paper is to make otherwise hard work routine and easy, and hence reliable; the purpose of this part of the paper is to motivate authors to *do* the work, to use the proposed or other such tools to improve the quality of their published work.)

A great deal of the published literature in computer science is about the commercial applications of systems, software and hardware reviews and such like. Computer science and business are so closely connected that the methods of science are often confused with the methods of business, where disclosure without limitation would be unprofitable. For example, a review of commercial Chinese Postman solutions [31] says that vendors are 'tight lipped' about their algorithms. From the commercial point of view, then, there is no interest in and even a resistance to clear exposition of algorithms.

Computer science must be unique in the continual triumph of hope over reality. Software is unreliable. We tend to emphasize hope in tomorrow's solutions, as exemplified by the exponential growth of almost every performance figure, yet we forget that every new solution bought represents

[‡]Les Carr has an interesting experimental system for converting change files into unfolding explanations (which he uses for lectures); see www.ecs.soton.ac.uk/~lac/annann.html.

a failed, obsolete solution that is unfixable because its workings are unknown and hidden: Moore's Law is a business, not a technical law [32]. As commercial software warranties show, customers are made responsible for fixing problems that should never occur, or which—in any other field—should be and would be the manufacturers' responsibility [33]. Undoubtedly, concealing and obscuring program code, regardless of its scientific hazards, is good business: it is more profitable to sell upgrades than to admit or fix problems.

As professionals we are so inured to this continual progress that it seems perfectly reasonable to publish code and shortly afterwards say that it does not work because the compilers have changed, or because we have upgraded the operating system, or because the work was on a computer that has now been replaced. Or perhaps we didn't make a backup before we lost it? There is a very, very fine line between publishing stuff that does not work or that cannot be got to work because it is described vaguely, and publishing stuff that *never* worked, or never worked quite so well as it has been written up. This is the computer science equivalent of removing data points that do not agree with predictions, or worse. In the established scientific disciplines any of this sort of behaviour would lead to outrage [34,35].

If program code is not made available, how can it be tested? How can it be refereed? How can the community develop the ideas without starting again? How can the discipline advance with confidence? In particular, if the publication itself does not provide an adequate description of the algorithm, where are future workers to find it?

These problems are ironic because the computer itself can provide a stronger test of validity than any that is available to other scientists in their work [36]; furthermore the testing can be automated, for instance using framework testing [37]—so there is no long-term burden on the author. And there is everything to gain.

2.3. Positive aspects of computer science publishing

Such a negative list of problems needs balancing with some undoubted successes. Certainly there are some specialist areas, such as the ACM's publication of numerical algorithms, where these problems have been recognized and addressed, but the sentiment is by no means universal.

Open source software (distributing program source code openly, allowing the code to be criticized and fixed by others) has resulted in some unusually reliable programs [38]; the Internet is encouraging authors to make their programs downloadable; Java applets allow authors to publish documents that *include* working programs (though the source code still may not be available).

In computer science we can make the tools to help make the necessary scientific assurances easier to achieve. We review suitable tools later, but there are many simple and effective approaches that are often overlooked: for example, using archives. There is a Web site of collected algorithms (http://www.acm.org/calgo/) so the source code of some algorithms associated with refereed publications in ACM journals can be downloaded. CALGO specializes rather on mathematical algorithms, and one wonders why algorithms more generally are not made available, for instance for the numerous programs discussed in the *Communications of the ACM* or in *Software—Practice & Experience*.

As noted above, being able to access complete source code, useful as it is, solves a different problem than assuring that the published explanation of code (and examples of runs, if any) is accurate.

2.4. Not just programs, but data too

One reason incorrect implementations of the Porter stemming algorithm proliferated is that the original paper on the algorithm [20] did not provide test data. People read the pseudo-code description of the algorithm in the paper, and they implemented what they thought was meant. Neither they nor the people they distributed their implementations to had any way of checking their work against what Porter intended. Porter's current work is now supported with a database of test cases to avoid this problem.

Even when a paper explains an algorithm lucidly, there is a danger that a reader's transcription error will lead to mistakes. In particular, even though a well-defined programming language is used in an explanation, some readers of the paper may need to translate the program into a language that is accessible to them. If it is plausible that errors will creep in, then there need to be mechanisms to reduce the likelihood that errors of any sort remain undetected. Test criteria and, typically, test data should be made available (or programs that generate test data should be made available).

It is of course possible that the original author of a program has made a mistake in the choice of test data, and thus their claims about their algorithm are faulty. Distributing the test data is one way to help detect this sort of mistake.

Some fields of computer science, notably information retrieval and machine learning, have made substantial test databases available. There are also Web sites of XML data, and programs to check XML standards conformance. There are many other examples, but it is by no means standard practice. One wonders why the need for publishing test data (whether for pure science or for profit) is not more widely recognized. For example, Java compilers would be more reliable—and Java programs more portable (cf. this paper's Acknowledgements)—if they could be and were routinely validated against standard tests [39].

2.5. The ethics of publication in computing science

When the community relies on unreliable publications, problems arise. In computer science this is at least tedious; in medicine, for instance, lives are obviously put at risk. Medicine has therefore developed a strong sense of appropriate ethical behaviour to be applied throughout the publication process. The medical community has indeed detected numerous cases of fraud and unethical publishing behaviour. It would be very strange if computer scientists were exempt from the sorts of temptation to which medical researchers succumb—only recently, even physicists had a chastening experience [35]. Yet the ACM Code of Ethics and Professional Conduct [40] makes no mention of ethical behaviour directly relevant to maintaining high standards in scientific publication, though 'the honest computer professional will not make false or deceptive claims about a system' (Code §1.3). Few computer professionals, however, will be conscious of the connection between this exhortation and their own standards of scientific publication.

There are many refereed computing science papers that 'envision' work that is clearly not yet working. Some imaginative papers do not even make their envisionment status clear. Such papers make it hard for subsequent work to progress, as referees tend to reject later papers that appear to repeat, test or develop prior work, even if that earlier work was purely envisionment, as if the scientific copyright to the idea was held by prior imagination rather than by actual fact. Taking conceptual ideas to something that *really* works uncovers important additional details and qualifications that would not have been known but for the concrete effort [41]. Yet to work to these standards takes time, and risks others jumping in with outline papers that take precedence—and then the mature work cannot (under current practice) readily be published.

There is certainly scope for further ethical research to develop and apply the traditions of reliable science and reliable publication to computer science—for instance to be clear about the dangers of fudging, trimming and cooking published code. See Resnik for a general review [42]—especially if these terms are unfamiliar! However, this paper is not the place to pursue in any greater depth a discussion of the processes and ethics of publishing; instead we turn to pragmatic ideas to help avoid errors in the first place.

3. 'LABORATORY BOOK' TOOLS FOR EXPLAINING CODE, LITERATE PROGRAMMING AND RELATED WORK

Literate programming was introduced by Knuth [41,43–46] to combine programs and their documentation to create readable programs. Numerous programs have now been prepared using literate programming: sizable examples include Knuth's own outstanding books (e.g., on graph theory algorithms [47]). In literate programming, there is no duplication of either code or documentation: there is only one shared copy. Code and documentation are interleaved in a file, and as they are adjacent in the same file it is *very* much easier to keep them consistent. Reducing the obstacles for editing both together, and increasing pride in the polished results, has an invigorating effect on programming, as well as on dissemination.

Conventional literate programming systems support the internal documentation for entire programs. When literate programs are processed, cross referencing, tables of contents, and indexes are generated automatically. Literate programming breaks code up into separate 'modules'; this makes the code easier to structure, but introduces various conventions the documentation reader must understand. The overall result, as well as a compilable program, is essentially a book (with automatically generated cross references, contents, indices, etc.) but one providing unusually good internal documentation.

In the past, few people wrote literate programs using tools they had not written themselves, because if you built your own tool, you understood its behaviour and building your own system was easier than fathoming out the workings of someone else's tools. Furthermore, most tools make assumptions, and circumventing the imposed approach to fit in with a particular project may be harder than starting from scratch. Today, only a few literate programming tools have survived the test of time.

Despite its advantages, after almost 20 years the use of literate programming for publishing code in the mainstream literature is now negligible. In whatever ways people may be using literate programming internally in software development projects (in lab notebook type uses), it is evidently not addressing the needs of the broader research community for publication. Probably the main reason for literate programming failing to survive in the literature is that it imposes its own styles and conventions (e.g., modules), which adds significantly to the 'noise' of a paper and makes it hard to conform to journal style requirements. Ramsey's paper in this journal [48], for example, required a section to explain the notation being used.

Nevertheless literate programming is certainly a 'good thing' and numerous variations and alternative approaches have been developed to achieve some or all of its advantages. The following are examples, and illustrate the diversity of useful approaches:

Autoduck [49] generates documentation from program source code, generating Microsoft Rich Text Format (RTF) text, which is presentable in WYSIWYG applications such as Microsoft Word. If programs are to be released to customers, technical authors can add examples and other documentation to the source code in order to generate programmer documentation. Autoduck can be used to generate online hypertext help for programs. Autoduck supports much more extensive documentation than Javadoc (*q.v.*) does, but it is much more complex than Javadoc; its limitations for writing about programs are similar.

Doc [50] allows TEX code to be documented in such a way that the code can be used directly (because documentation is simply written in standard comments), but if the documentation is wanted a 'driver' file is used to obtain it. Since documentation makes files larger, and hence slower in an interpreted system like TEX, the partner system **docstrip** can be used to remove documentation.

Haskell [51] is a programming language with two styles: in the standard one, comments are conventional (i.e., text on lines following a `--` code, corresponding to Java's `//`, is ignored); in the converse literate style, comments and program are 'swapped', so the documentation itself is the default, and lines of program are 'uncommented' (by `>` codes).

Javadoc [52,53] generates program interface documentation for Java: it is therefore more specialized than literate programming, though the result is intended to be browsable hypertext rather than printed. Javadoc uses an extended Java comment and is invisible to Java programming tools. It is not sensible to modify the generated documentation, and the approach is not suitable for writing *about* programs.

LiSP2TEX [54] is one of several systems that places program code in the documentation. A tool reads the code, evaluates it, and merges the results into the documentation. This approach assumes that the source documentation fully defines the program.

Loom [55,56] inserts named sections of program code into documentation. It is described more fully later in this paper, in Section 4.

***Mathematica* [57]** is a combined word processor and powerful symbolic mathematics package. There are *Mathematica* journals that take *Mathematica* articles and publish them directly. Although documentation and code can be interleaved, it must be in the right order for *Mathematica*, and it is not easy to omit code (such as initialization) that one may not want to write about. There are literate papers written in *Mathematica* [58] and examples using concealed code, just publishing the explanation and output [59].

Noweb [60,61] is a simplified literate programming tool, following closely in the style of Knuth's approach, but which aims to reduce the learning and effort hurdles to using literate programming. Noweb can convert a literate program into a conventional compilable program, by putting the noweb explanation into ordinary comments: this makes the resulting program more accessible, but unconstructively encourages programmers to edit a copy of it, not the actual explanation. Noweb has been used for publishing programming books, such as Dave Hanson's book on C programming [62], and it has been used (by its author) to publish papers in *Software—Practice and Experience* [48].

Quine [10,11] is a logic language designed to generate manuals. The language includes an explanation mode, where text between the symbols { and } is copied directly and logic expressions and their evaluations are formatted in HTML unambiguously. The published papers were generated in HTML using the system to explain itself.

Soap [63] is mentioned here because it is a *really* old system: some people have been concerned with these issues since at least 1977! Soap provides a converse to literate programming: it could read documentation to extract compilable source code—and therefore helps check that what documentation says is still accurate.

Warp is the main practical subject of this paper, and is described more fully below, in Section 4.

A wider review and details of the current state-of-the-art in literate programming, as well as tools to download, can be found at www.literateprogramming.com.

4. PUBLICATION TOOLS FOR EXPLAINING CODE WARP AND LOOM

Published papers are edited by hand, and often go through a long and arduous life-cycle; papers are often revised and, in some cases, have to conform to changing typographical requirements as they are resubmitted to different journals. Equally, the program code on which a paper is based has to conform to specific programming language, compiler and system requirements in order to work at all.

We now summarize, after the wide-ranging discussion in Sections 2 and 3, the now self-evident requirements for tools for supporting the reliable explanation of code for publication. Each requirement is followed by **LP**, **L** or **W**, depending on whether the requirement is met well by literate programming in general, or more specifically by the tools loom or warp (which will be described in more detail later).

1. As a paper is revised, it is very easy for the text of the actual program to diverge from the text of the paper: program compilers and word processors have quite different criteria, and the author has quite different goals when working in each context. Under these circumstances, to ensure published papers are reliable, it makes sense to share common text automatically as much as possible. The key idea is to make it very easy to keep together and maintain what are normally separate—and sometimes independent!—documents: shared text should only be edited in one place. This is basically standard good software engineering practice: don't duplicate information unnecessarily, otherwise discrepancies soon accumulate. The requirement is to establish and automatically maintain a reliable relationship between parts of two or more files: the explanation will include or refer to fragments of program text, and the program will include (as substrings or files, or in some other way) fragments that are to be included in the documentation. **LP**, **L**, **W**.
2. The *shared* texts (i.e., the original program code eventually appearing in the explanation) could in principle reside anywhere: in separate files, in program files, or in documentation files. Management will be easier if the number of files that require human attention is reduced. Since program development tools are usually sophisticated and interactive, and encourage the author to edit program text, all shared text should originate and reside in the program source file, where it can be edited and used without restriction to support all normal programming activities (e.g., testing), with no overhead. **LP**, **L**, **W**.

3. More specifically, the *original* program source files and the documentation source files must be or remain in their standard formats; that is, they must be able to be edited and manipulated directly using existing development/word processing tools without affecting the integrity of the shared material. **W**.
4. The tool must be lightweight and so easy to use that it does not encourage any manual touching-up—even when starting to write a paper, when there is 'hardly anything to do' and it looks seductively easy to do it all by hand! In contrast, a heavyweight system would tempt authors to do it all by hand because to do so would seem easy in comparison to getting up to speed with the tool. **L**, **W**.
5. The tool must scale, and not introduce 15 or more of its own problems with larger projects. It should work equally well with small programs and small papers, as with large programs and long papers. In particular, the tool should be useful throughout the lifecycle of the document, as it goes from the earliest stages of drafting to final publication. **L**, **W**.
6. Whereas literate programming supports the documentation of an entire program, publication requires the author to be more selective (except for the very shortest programs). The tool must permit *any* fragments of code to be explained, and in any order that suits the explanation. **L**, **W**.
7. The approach itself must not need any special explanation when used in papers or other documents—it should simply embed fragments of code as required without imposing any notation or conventions on the reader of the paper. (Literate programming systems are excluded by this requirement.) **L**, **W**.
8. The tool must work with real code written in real programming languages. Ideally the approach should be language independent. **LP**, **L**, **W**.
9. The tool must make all necessary translations between the programming language and the text processing system (e.g., if the programming language is Java and the text processing system is LaTeX, then '{' has to be converted to '\{', etc.). **LP**, **W**.
10. Given that the World Wide Web is such an important tool for dissemination, the tool should provide support for publishing on the Web. Obviously most authoring systems can generate material that can be rendered for Web pages (e.g., by post-processing to PDF or other formats [64]), but the tool should ideally support more flexible hypertext authoring directly, at least in HTML and XML. **W**.
11. Since there are numerous programming languages and numerous text processing systems, the tool must be extensible—and it must be extensible (e.g., by using post-processors) in a well-defined way. **L**, **W**.
12. The tool must have a 'commensurate level of complexity': sufficient to solve the core problems, but not having so much generality that the complexity of parameterizing and using it is a hurdle in itself. (The complexity of literate programming systems is a significant hurdle limiting their widespread use [65].) Commensurate complexity has been widely promoted through the slogan, 'make the simple easy and the complex possible'. **L**, **W**.

Such a list of requirements can be met in many ways, especially as compromises have to be made. What is easy and effective to use depends, for the author, on their experience, the platform they work on, and the community they work within.

4.1. Warp: an approach to explaining code reliably

Warp is a tool for supporting the reliable publication of code that satisfies the requirements listed above[§]. Warp works with C, C++, Java, *etc*, but we will explain it here using Java. Warp is itself written in Java. Warp would need trivial modifications to work with some other programming languages, say, Prolog or LISP: the current version of warp does not permit programming language comment conventions to be changed.

Although warp has various features and options, we will introduce each main feature in turn, gradually building up the complete picture. First, at its simplest, warp is run from the command line, operating on (as it may be) a Java file:

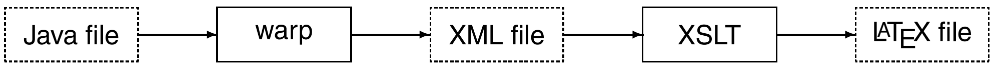
```
warp prog.java
```

This basic use of warp outputs the program `prog.java` but marked up in XML, which is a widely used, general-purpose international standard markup language [66]. Although users of warp would probably not look at the XML, it helps understand warp to see how it translates typical parts of Java programs:

Java fragments	Generated XML
`"A string"`	`<string>A string</string>`
`/* comment & stuff */`	`<block-comment> comment &amp; stuff </block-comment>`
`identifier`	`<name>identifier</name>`

Thus warp allows any program to be considered directly as XML.

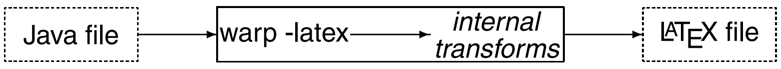
The complete XML translation of the program generated by warp can now be processed with any of the wide range of the available XML technologies. Once warp processes Java to XML, it is then routine for tools like XSLT [67] to generate explanations. The schematic below shows how warp and XSLT together might generate a LaTeX file corresponding to a source Java file.



An advantage of this approach is that XSLT, which is programmable, can be used to generate any format, not just LaTeX as shown here. However, since standard XML tools (e.g., XSLT, XPath) are complex and have a daunting learning curve, warp has been designed so it can do all the necessary processing itself without using them. In fact, warp can be used without generating any XML at all. Typically, for most purposes, then, warp will be used to extract the relevant code fragments directly, which it converts to any of various documentation formats as required, including LaTeX, ASCII

[§]Warp can be obtained from http://www.uclic.ucl.ac.uk/harold/warp.

and HTML. No intermediate XML files are generated if warp is used in this way, and the author does not need to use any XML tools; the small disadvantage is that the author has to use warp's fixed internal transforms (in contrast, XSLT is fully programmable and can be set up to conform to any specific conventions the author requires).



As far as warp has been explained so far, it has no feature to help isolate the specific parts of a program that may be needed in an explanation. Warp therefore additionally allows a programmer to introduce arbitrary XML into programs, and then the relevant parts of the program can be more easily picked out by warp, XSLT or other tools. Here's how it is done...

Java allows comments:

```
// ordinary Java comment on one line

/* or block comments
   over several lines
 */
```

which simply get ignored by the Java compiler (and sometimes by the programmer). Warp marks these up in XML as described above, *except* when the comment text is itself XML it gets copied directly. So a Java comment containing XML like `/* <highlight> */` is translated to just `<highlight>`—the comment markers, `/*` and `*/`, themselves disappear and no `<block-comment>` tags are generated by warp.

XML in a comment like this could be used to highlight part of a Java program up to a following comment like `/* </highlight> */` which, matching the earlier `<highlight>` tag, would end the highlighting.

Since the extra XML is introduced into programs is standard Java comment it is completely ignored by the Java compiler and has no effect on the meaning of the program. With this approach, warp ensures programs correspond directly to XML in a straightforward way *and* that programmers can very easily introduce additional XML for any purpose, and all while using standard Java.

Typically XML is written into comments in programs like this:

```
// <explain id="helloworld">
System.out.println("Hello World!");
// </explain>
```

The tag name (here, `explain`) is arbitrary and can be chosen by the author for any purpose. The XML tags and attributes can be used by XML tools to extract and translate the corresponding Java from the entire source file; in this simple case just one line of Java has been defined. However, it is generally easier to use warp for the whole process without involving any XML tools: thus to pick out the Java code above and convert it to LaTeX directly, warp would be used as follows:

```
warp id=helloworld -latex prog.java > helloworld.tex
```

Or to convert it to HTML:

```
warp id=helloworld -html prog.java > helloworld.html
```

In both cases, the warp parameter `id=helloworld` is a simple pattern (in these examples) that matches the attribute of the `<explain id="helloworld">` tag, and causes warp to extract all text which is that tag's content, that is, up to the matching `</explain>` end tag. The `-latex` flag makes warp mark up that code to LaTeX; in turn, the `-html` flag makes warp mark up that text to HTML. In both examples above, the formatted output of warp is saved in a file, `helloworld.tex` or `helloworld.html` respectively, for further processing.

This trivial example could be made more interesting by including further markup in the Java program, for instance:

```
// <explain id="helloworld"> <h1>Hello World Example</h1>
// <p>This is now a fully self-contained example!</p>
System.out.println("Hello World!");
// </explain>
```

The final feature of warp to describe here extracts XML attributes that have been introduced in comments, rather than the Java code. Warp outputs lines of text, and in this case it would be just `helloworld`. This mode of warp can be used to identify the code fragment or fragments the programmer intended to be explained: that is, those that have been marked-up with XML. (In the appendix following this paper, the Java source file has four such fragments, and they all have different attributes.) The attribute value (here, `helloworld`) can be chosen so it is also a suitable file name stem, so it is quite easy to write a script that extracts every code section and puts each in an appropriate file, or processes them in other ways.

In summary, warp is trivial to use for most of its intended applications, but the full power and generality of XML is available for authors who wish to do complex operations.

Although warp helps explain code, it has many other uses, since XML is so versatile. Here are some possibilities, all of which are easily supported:

- A Java program can be converted to XML using warp. Arbitrary transformations can be performed on the XML, for instance embedding data structures generated by other programs. XSLT can then translate the warped XML *back* to Java, where it can be edited and debugged as usual. If the data structures later require updating, warp can be used to translate the program back to XML, and then the XML tools can be used to update the data, and so on. Thus we have the advantage of normal Java development, and automatic insertion of correct data—a powerful generalization of techniques used in some Rapid Application Development tools.
- A Java program may have a complex data structure, such as a graph. Warp can extract the data structure to an XML file, and the Java data can then be processed to display it graphically or to check it has desired properties.
- Programmers are often unable to finish programs in one sitting, and often have to create stubs to be completed later. Often comments get littered with notes, to remind programmers

to finish a piece of code in a particular way. Warp can be used to identify stubs and other notes-to-programmers, which can then be extracted easily as reminders any time in the future. For example, writing a comment `/* <note>Must add error code here</note> */` can be picked out by warp's patterns (or by using XML tools) along with other notes, and hence help the programmers to work more reliably.

- The opposite of explaining a program reliably might be setting an exam paper on programming. Warp can be used to generate reliable exam questions: first with the code, so they can be checked by examiners, then later omitting code to set the exam paper for students; or it could highlight code from working programs that students are expected to explain; and so on. In all cases, warp helps ensure the question is accurate, and based on a compiled, working program.

Finally, the current version of warp is intentionally simple; it does not do parsing, pretty printing and many other functions that might be desirable for other purposes than for which it is intended.

4.2. Using warp

This paper is followed by an appendix which consists of a complete example showing the results of using warp in a realistic context, written and formatted as a typical short paper to *Software—Practice & Experience*. The example was processed directly by warp (and nothing else) without using any XML tools at all, and it was submitted to the publishers in LaTeX, so they have had very little cause to modify it whether deliberately or accidentally.

One should note that the code published in the short appendix does not constitute the entire code of a complete program: for example, Java's use of `import` statements is irrelevant for the purposes of the example, and therefore none appear in it; likewise, the actual program provides a test harness (that can, amongst other things choose between two different implementations) but none of that appears in the example. Nevertheless, the code shown in the appendix is the exact code and it has (along with its `import` and other necessary support statements) been checked by a compiler, and has been tested and run successfully.

Warp is used from the command line, and a file or files (presumed to be C, C++ or Java, etc.) are converted to XML or other specified format (ASCII, LaTeX, . . .), as follows:

`warp -`*`format files`*. . .

Convert *`files`* to the specified format (currently implemented: `-ascii`, `-html`, `-javascript`, `-latex`, `-xhtml` and `-xml`); the default format is XML. Case is not significant, so purists can write `-XML` or `-LaTeX` if they wish.

Any XML comments in the files are copied directly by warp, but with the comment symbols stripped (unless the `-all` option is used: see below). An 'XML comment' is any comment text (within `/*...*/`, or after `//...`) starting with `<` (perhaps preceded by blanks), ending with `>` (perhaps followed by blanks), and containing lexically correct XML. This is a more relaxed restriction than being properly well formed XML, since it is unnecessary to match start and end tags within the comment (though if XML's `CDATA` or `<!--` tags are used these would have to be contained entirely within the program comment).

When generating LaTeX, warp 'left shifts' code to remove any uniform indentation, though the original indentation can be restored if desired (e.g., by using a LaTeX macro). Thus methods and nested blocks, and so on, can be explained without their original nesting looking inappropriate in the context of the final document.

`warp -`*`format pattern files`*`...`

When a pattern is specified, warp extracts, converts and concatenates all matching XML sections of *`files`*. This is warp's normal mode of use.

`warp` *`pattern files`*`...`

List all matches of the pattern in *`files`*. The output is formatted as lines of text in the same syntax as a warp pattern, and it can therefore be reused in another run of warp (e.g., in a shell script).

`warp -`*`format`* `-all` *`files`*`...`

Convert *`files`* to the specified format, ignoring the special status of XML comments, and hence *including* comments containing XML. (The flag `-all` simply stops warp recognizing XML in comments.) The main purpose of `-all` is to help authors review complete code within the relevant document processing system.

`warp -help`

Provide summary help of warp (including some details we do not cover in this paper, including the use of XML DTDs, and a description of patterns).

4.3. Comparing loom and warp

Janet Incerpi and Robert Sedgewick's Loom [55,56] was used originally for Sedgewick's book *Algorithms* [7]. Loom has no published description (there is only an example of its use [55]): a detailed comparison of loom and warp is therefore justified here.

Loom extends comments by matching forms `/* include` *name* `*/` to `/* end` *name* `*/`. This defines named sections of code, and the effect is similar to warp's XML approach, which would be to use `/* <include stuff="`*name*`"> */` to `/* </include> */`.

A loom documentation file then includes the named sections using `%include` commands:

`%include` *file section*

or

`%include` *file section* | *command*

which (when processed by loom) are replaced by the named section, which loom locates in the named file. If a command is specified, the named section of code is filtered through it (the command is a Unix filter, and could be used possibly even to compile and run the section of code to insert its output).

This is similar to warp, except warp writes the named section to a file and the documentation system reads the file. Loom can transform the program text inside the documentation, by using the command filter; in contrast, warp processes the file outside of the documentation—but this allows the documentation to use a standard include in warp, rather than a special syntax as loom requires. Warp can be used directly with proprietary tools such as Microsoft Word (which can include files), whereas loom would need to be modified to parse the Word format (and it would create two versions of the documentation file, one with and one without the code included). Just as warp has various built-in document formats, an obvious extension to loom would be to support various include formats.

Apart from the syntactical details, a main difference to using warp and loom is that loom must process all files together—both program and documentation—whereas warp only processes the program. If the author completely forgets to run loom, then the formatted documentation will contain no code; if the author forgets to run warp, however, the code files will not be created and there will be errors reported by the formatter (or an older version of the code will be used if the files are still available).

Warp and loom are compared in more detail in Table I.

Neither loom nor warp will be the last word in tools for explaining code. Fortunately both of them are small, simple programs: developing and evaluating new tools in this area will make a worthwhile and interesting project.

4.3.1. Confidence

The point of using an approach like loom or warp is to increase confidence in published results; moreover sufficient integrity has to be achieved with reasonable cost. We now discuss issues affecting confidence a little more closely.

The opposite approach to using markup would be to use a WYSIWYG mechanism, such as publish-and-subscribe [68]: but in all commercial implementations, the design goal of 'ease of use' makes it *far too* easy to make changes in one place and not the other. Small edits can lead to unnoticed and unknown consequences, particularly as a WYSIWYG system leaves no mark up trail to show explicitly that changes have been made.

When writing internal documentation, as opposed to explaining code, an issue is to ensure that *all* of the program is documented, or at least that nothing relevant is omitted from the documentation. On their own, loom and warp, but unlike a literate programming system, cannot make any such guarantees. However, for writing about programs, it does not matter—indeed it is helpful—when bits are concealed.

Any process that allows explanation to take full advantage of a typesetting system is open to abuse. One might insert hand-crafted code that looks like it has been warped, but has none of the warranties doing it automatically achieves. Warp itself permits anything that XML permits, which is a lot; however, when used with its internal filters, what the author can do is more restricted (e.g., code cannot be deleted without going to unreasonable effort). Nevertheless a high integrity version of loom or warp might closely restrict what is allowed so that code cannot be modified.

4.3.2. Design decisions with warp

The first version of warp used comment codes like `//>` *`file`*, which while neat, brief and Unix-like, were easy to mistype, hardish to search for when editing (all the symbols used were existing Java operators), incomprehensible to third parties, and lacking in redundancy—many errors in their use were undetectable (e.g., there would be no warning if you failed to shift the > key and accidentally typed a dot instead: `//.`—which would be just ignored comment). In short, they suffered from all of the problems of conventional literate programming codes, `@[`, `@;`, `@'` and so on (there are over 30 such codes, plus a collection of cryptic TEX macros, such as `\0`). Indeed when Knuth and Levy say of

Table I. Outline comparison of warp and loom.

Warp	Loom
Standard program code with comments.	Standard program code with comments.
Standard documentation file.	Special `%include` command in documentation.
Requires no external programs for basic use.	Requires special filters for use with most text formatters (such as LaTeX).
Code can specify typographical or other features for documentation.	Code cannot specify any typographical features, but may be arbitrarily transformed (by filters) in documentation.
Designed to concatenate code together from several parts of a program.	Does not concatenate code.
Must be run before documentation is formatted if code has been updated.	Must be run when documentation is formatted.
Main advantages	
Documentation can be given to publishers without any special instructions.	
Both code and documentation files remain standard source files that need no special processing.	Only code remains standard.
Documentation can make full use of any multi-file structuring facilities of formatter.	
Uses XML. Code can be processed in any way if desired. The XML is embedded in the program code.	Ability to filter code using Unix tools. The filter command is embedded in the documentation.
Forgetting to generate files will get error messages from the documentation system.	
Main disadvantages	
	Forgetting to run loom will generate no diagnostics (and in LaTeX, loom's `%include` syntax is a comment, so generates nothing).
	Documentation cannot use structuring (e.g., be a nested multi-file document).
Documentation is not a single file (but the generated files never need editing or looking at).	Original documentation cannot be distributed without complete program and the program loom.
Main technical differences	
Basic use with HTML, LaTeX etc. built-in.	Basic use limited to ASCII.
Optionally uses XML.	Optionally uses Unix filters.
Relies on document processor being able to include document files. (Since HTML does not strictly permit file inclusion, warp can generate equivalent JavaScript that can be included.)	Relies on loom parsing documentation file format (and being restricted effectively to LaTeX without reprogramming).
Generates intermediate files.	Uses Unix pipes.

the `@'` code that 'this code is dangerous' [43] you know something is wrong, and to be avoided for a reliable system!

The next version of warp used mnemonic codes (such as `// $save$ filename`), more like loom, but these were still arbitrary and difficult to make powerful enough without inventing a whole new language. The insight, suggested by George Buchanan, was to replace warp's codes with XML. At a stroke, one was using a standard notation, and warp changed from being a special purpose tool into a general purpose tool.

XML is verbose for different reasons than warp; nevertheless the overhead in typing the extra characters should be seen in perspective. XML text can be shared, reused and checked easily, and the savings in errors avoided can be considerable. Perhaps programmers' habitual preference for special characters and short cuts is, more, a symptom of poor design and the fact that, normally, we have to repeatedly edit and re-edit things for multiple purposes—we naturally wish to reduce the effort of repeated editing. A better way is to increase the reuse of shared texts, as here, to multiply the impact of our work rather than to make it quicker to edit *per se*: easier editing is not the goal, writing reliably is, and therefore editing less often. With longer keywords and a stricter syntax, ultimately we do less work and we achieve our real goals faster—which is much the same rationale for using high level languages instead of assembler.

4.4. Generalized explanations of code

Literate programming and the other approaches described here explain programs to people who are interested in their operation. Future research should find better and more reliable ways of explaining programs to their non-specialist users as well, perhaps by automatically taking advantage of some comment scheme, and perhaps by generating reliable interactive help. Programmers would then become a bit closer to the explanatory needs of their programs' end users. This might encourage programmers to make their programs easier to understand.

Elsewhere we have discussed writing better manuals for users [11,69,70] (user manuals being explanations of programs for users, rather than for computer scientists); we have also discussed the useful impact of quality explanation on programming language design [71] and on physical device design [72].

Of course, programming, writing and explaining are vast areas in their own right; see [73] for explanation by visualization, [74] for a review of annotation, [75] for a proposed approach to multiple-use documents, and [68] for a summary of commercial linking, embedding, and publish-and-subscribe technologies.

5. CONCLUSIONS

Programming can be undertaken for many reasons, but if the purpose is to advance computer science by publishing, then we should work at least to the same scientific standards that have been found appropriate for other disciplines. In the conventional sciences there are many forms of documentation, from laboratory books through to refereed publications. In computer science, a range of tools support managing the internal documentation of the lab book, through to supporting the reliable publication of code. Although there are overlaps, literate programming mainly supports the laboratory book end of the spectrum, and tools like loom and warp mainly support the journal publication end.

Compared with literate programming both loom and warp are much simpler for the author, for the reader, and for the publisher. Their simplified approaches avoid the intellectual and typographical hurdles conventional literate programming approaches impose. The approaches are ideal for explaining algorithms, which typically requires code extracts, rather than for writing *complete* readable programs, which remains conventional literate programming's forte. Both loom and warp are extensible; both can be used simply for basic authoring, and yet can accommodate sophisticated requirements.

Unlike loom, warp treats programs as XML documents, and it does so with such a light touch that the programs being explained can still be edited and developed as normal, using existing editors and programming tools. Although XML can be processed in many ways, warp also provides a simple scheme for extracting sections of code and for transforming them into LaTeX, XML or other formats for processing in other documents, in particular for 'stand alone' documents such as journal publications. Warp is probably the simplest possible scheme, yet remains enormously versatile thanks to its optional use of XML.

In an ideal world, it might not be necessary to have any separation between code and published explanation. Even though some tools, such as *Mathematica*, make them almost the same, this is the exception—and the awkwardness of working with 'almost the same' rather than 'exactly the same' is a great impediment for reliable writing, because the author is continually either compromising or using work arounds—for example *Mathematica* does not allow an author to change the order of code or to quote code fragments to explain them better. Unless we want to publish code in 'toy' languages designed specifically for the purpose, we will probably achieve no fundamentally better solutions than those discussed in this paper.

Knuth writes that 'Science is what we understand well enough to explain to a computer' [76]. This paper argues, further, that science—including computer science—will progress better when those programs are explained more reliably to the scientific community, and that tools such as warp are a very good way to do this.

ACKNOWLEDGEMENTS

Ann Blandford, Roland Backhouse, George Buchanan, Paul Cairns, Anthony Finkelstein, Michael Harrison, Matt Jones, Peter Ladkin, Gary Marsden, Norman Ramsey and Russel Winder made constructive comments on this work. Graham White solved a Java system bug by compiling warp on another system that behaved differently (and hence revealed the cause of the bug)—so much for Java being portable! The author would also like to thank the anonymous referees for their contributions to this paper. Harold Thimbleby is a Royal Society-Wolfson Research Merit Award holder.

Since this paper is intended to stimulate further development, the author also acknowledges future contributors to warp-like tool technologies.

REFERENCES

1. Deutsch D. *The Fabric of Reality*. Penguin: London, 1997.
2. Cartwright N. *How the Laws of Physics Lie*. Oxford University Press: Oxford, 1983.
3. Ziman J. *Reliable Knowledge*, Canto (ed.). Cambridge University Press: Cambridge, 1991.
4. Abelson H, Sussman GJ, Sussman J. *Structure and Interpretation of Computer Programs* (1st edn preface). MIT Press: Cambridge, MA, 1985.
5. Perlis AJ. A new policy for algorithms? *Communications of the ACM* 1966; **9**(4):255.

6. Dunham CB. The necessity of publishing programs. *Computer Journal* 1982; **25**(1):61–62.
7. Sedgewick R. *Algorithms*. Addison-Wesley: Reading MA, 1983.
8. Libes D. Writing a Tcl extension in only 7 years. *Proceedings of the Fifth Annual Tcl/Tk Workshop'97*, 1997; 14–17.
9. Underhill P. *Why We Buy: The Science of Shopping*. Texere: New York, 2000.
10. Ladkin PB, Thimbleby H. From logic to manuals. *Software Engineering Journal* 1997; **11**(6):347–354.
11. Ladkin PB, Thimbleby H. From logic to manuals again. *IEE Proceedings Software Engineering* 1997; **144**(3):185–192.
12. Kuan (Kwan or Guăn) M-K Graphic programming using odd or even points. *Chinese Mathematics* 1962; **1**:273–277.
13. Lin Y, Zhao Y. A new algorithm for the directed chinese postman problem. *Computers and Operations Research* 1988; **15**(6):577–584.
14. Chartrand G, Oellermann OR. *Applied and Algorithmic Graph Theory*. McGraw-Hill: New York, 1993.
15. Skiena S. *The Algorithm Design Manual*. Springer: Berlin, 1998.
16. Jungnickel D. *Graphs, Networks and Algorithms* (*Algorithms and Computation in Mathematics*, vol. 5). Springer: Berlin, 1999.
17. Ahuja RK, Magnanti TL, Orlin JB. *Network Flows: Theory, Algorithms, and Applications*. Prentice-Hall (Simon and Schuster): Upper Saddle River, NJ, 1993.
18. Dror M (ed.). *ARC ROUTING: Theory, Solutions and Applications*. Kluwer Academic Publishers: Dordrecht, 2000.
19. Knuth DE. *The Art of Computer Programming* (3rd edn). Addison-Wesley: Reading, MA, 1997.
20. Porter MF. An algorithm for suffix stripping. *Program* 1980; **13**(3):130–137.
21. Porter MF. Snowball: A language for stemming algorithms. http://snowball.sourceforge.net/texts/introduction.html[2001].
22. Thimbleby HW. An efficient equivalence class algorithm with applications to autostereograms. *Software—Practice & Experience* 1996; **26**(3):309–325.
23. Thimbleby HW. Using sentinels in insert sort. *Software—Practice and Experience* 1989; **19**(3):303–307.
24. Thimbleby HW. Specification-led design for interface simulation, collecting use-data, interactive help, writing manuals, analysis, comparing alternative designs, etc. *Personal Technologies* 1999; **4**(2):241–254.
25. Thimbleby HW, Cairns P, Jones M. Usability analysis with Markov models. *ACM Transactions on Computer Human Interaction* 2001; **8**(2):99–132.
26. Anderson JR, Lebiere C. *The Atomic Components of Thought*. Lawrence Erlbaum Associates: Hillsdale, NJ, 1998.
27. MacKenzie D. *Mechanizing Proof.* MIT Press: Cambridge, MA, 2001.
28. Sterling TD, Weinkam JJ. Sharing scientific data. *Communications of the ACM* 1990; **33**(8):112–119.
29. Feynman RP. Cargo cult science. *Surely You're Joking Mr. Feynman!*, Hutchings R (ed.). Vintage, 1992.
30. Forsythe GE. Algorithms for scientific computation. *Communications of the ACM* 1966; **9**(4):255–256.
31. Hall RW, Partyka JG. On the road to efficiency. *OR/MS Today* 1997; **24**(3):38–47. http://lionhrtpub.com/orms/orms-6-97/Vehicle-Routing.html.
32. Brown JS, Duguid P. *The Social Life of Information*. Harvard Business School Press: Boston, MA, 2000.
33. Thimbleby HW. You're right about the cure: Don't do that. *Interacting with Computers* 1990; **2**(1):8–25.
34. Park R. *Voodoo Science*. Oxford University Press: Oxford, 2000.
35. Adam D, Knight J. Journals under pressure: Publish, and be damned.... *Nature* 2002; **419**(6909):772–776.
36. Simon HA. *The Sciences of the Artificial* (3rd edn). MIT Press: Cambridge, MA, 1996.
37. Beck K, Gamma E. JUnit. http://www.junit.org/ [2001].
38. Raymond ES. *The Cathedral & The Bazaar* (Revised edn). O'Reilly: Sebastopol, CA, 2001.
39. Thimbleby H. A critique of Java. *Software—Practice & Experience* 1999; **29**(6):457–478.
40. *ACM Code of Ethics and Professional Conduct*. http://www.acm.org/constitution/code.html [1992].
41. Thimbleby H. Experiences with literate programming using CWEB (A variant of Knuth's WEB). *Computer Journal* 1986; **29**(3):201–211.
42. Resnik DB. *The Ethics of Science*. Routledge: London, 1998.
43. Knuth DE, Levy S. *The CWEB System of Structured Documentation*, Version 3.0. Addison-Wesley: Reading, MA, 1994.
44. Mall D. http://www.literateprogramming.com [2001].
45. Ramsey N, Marceau C. Literate programming on a team project. *Software—Practice and Experience* 1991; **21**(7): 677–683.
46. Knuth DE. Literate programming. *Computer Journal* 1984; **27**(2):97–111.
47. Knuth DE. *The Stanford GraphBase*. Addison-Wesley: Reading, MA, 1993.
48. Ramsey N. A simple solver for linear equations containing nonlinear operators. *Software—Practice & Experience* 1996; **26**(4):467–487.
49. Artzt E. Autoduck. http://www.literateprogramming.com/autoduck.pdf [2001].
50. Mittelbach F. The `doc` and `shortvrb` Packages. *Technical Report*, Gutenberg Universität Mainz, 1997.
51. Peyton Jones S, Hughes J (eds.). Augustsson L, Barton D, Boutel B, Burton W, Fasel J, Hammond K, Hinze R, Hudak P, Johnsson T, Jones M, Launchbury J, Meijer E, Peterson J, Reid A, Runciman C, Wadler P. *Haskell 98: A Non-strict, Purely Functional Language*. http://haskell.org/onlinereport [1999].

52. Arnold K, Gosling J, Holmes D. *The Java^{TM} Programming Language Second Edition* (3rd edn). Addison-Wesley: Reading, MA, 2000.
53. Friendly L. The design of distributed hyperlinked programming documentation. *Hypermedia Design*, *Proceedings of the International Workshop on Hypermedia Design (IWHD'95)*, Fraïssé S, Garzotto F, Isakowitz T, Nanard J, Nanard M (eds.). Springer: Berlin, 1996; 151–173.
54. Queinnec C. *Literate Programming from Scheme to TEX*. Université Paris 6 & INRIA-Rocquencourt, 2000.
55. Hanson DR, Van Wyk CJ (moderator), Gilbert J (reviewer). Literate programming. *Communications of the ACM* 1987; **30**(7):594–599.
56. Hanson DR. Loom—weave fragments together. `loom.1` in ftp://ftp.cs.princeton.edu/pub/people/drh/loom.tar.gz [1987].
57. Wolfram S. *The Mathematica Book* (4th edn). Addison-Wesley: Reading, MA, 1999.
58. Thimbleby HW. Specification-led design for interface simulation, collecting use-data, interactive help, writing manuals, analysis, comparing alternative designs, etc. *Personal Technologies* 1999; **4**(2):241–254.
59. Thimbleby HW. Analysis and simulation of user interfaces. *BCS Conference on Human-Computer Interaction*, vol. XIV, McDonald S, Waern Y, Cockton G (eds.). 2000; 221–237.
60. Ramsey N. Literate programming simplified. *IEEE Software* 1994; **11**(5):97–105.
61. Ramsey N. http://www.eecs.harvard.edu/~nr/noweb/ [2002].
62. Hanson DR. *C Interfaces and Implementations*. Addison-Wesley: Reading, MA, 1997.
63. Scowen RS. Some aids for program documentation. *Software—Practice & Experience* 1977; **7**(6):779–792.
64. Goossens M, Rahtz S. *The LATEX Web Companion*. Addison Wesley: Reading, MA, 1999.
65. Van Wyk CJ. Literate programming: An assessment. *Communications of the ACM* 1990; **33**(3):361, 365.
66. Harold R, Means WS. *XML In a Nutshell*. O'Reilly: Sebastopol, CA, 2001.
67. Tidwell D. *Mastering XML Transformations: XSLT*. O'Reilly: Sebastopol, CA, 2001.
68. Microsoft Corporation. *Getting Started: Microsoft Office*, Document No. OF64076-0295, 1992–1994.
69. Addison MA, Thimbleby H. Intelligent adaptive assistance and its automatic generation. *Interacting with Computers* 1996; **8**(1):51–68.
70. Ladkin PB, Thimbleby H. A proper explanation when you need one. *BCS Conference HCI'95*, *People and Computers*, vol. X, Kirby MAR, Dix AJ, Finlay JE (eds.). Cambridge University Press: Cambridge, 1995; 107–118.
71. Thimbleby HW. Java: A critique. *Software—Practice & Experience* 1999; **29**(5):457–478.
72. Thimbleby HW. Calculators are needlessly bad. *International Journal of Human-Computer Studies* 2000; **52**(6):1031–1069.
73. Braune B, Wilhelm R. Focusing in algorithm explanation. *IEEE Transactions on Visualization and Computer Graphics* 2000; **6**(1):1–7.
74. Ovsiannikov IA, Arbib MA, Mcneill TH. Annotation technology. *International Journal of Human-Computer Studies* 2000; **50**(4):329–362.
75. Phelps TA, Wilensky R. Multivalent documents. *Communications of the ACM* 2000; **43**(6):83–90.
76. Knuth DE. Foreword. $A = B$, Petkowvšek M, Wilf HS, Zeilberger D (eds.). A K Peters: Wellesley, MA, 1996.