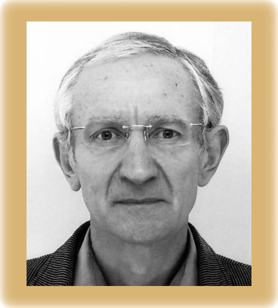
More than Just "Lost in Translation"

Olly Gotel, *Pace University*

Stephen Morris, *City University London*

Furnished with digital cameras and other handheld devices, many people now find media-rich content creation to be an inexpensive, everyday activity. In addition, ever-greater Internet accessibility and bandwidth let us post digital records of events on YouTube or turn them into podcasts in seconds. These technologies inevitably play a role in requirements engineering. As a consequence, agreed requirements are likely to come from information content transformed across and among myriad originating and intermediary representations. Not all such transformations preserve content and meaning, and some are irreversible. For example, translation transformations don't operate satisfactorily in a backward direction, which presents problems for subsequent traceability.



Traceability in Jeopardy

Software engineering needs traceability to aid requirements understanding, help manage requirements change, and demonstrate requirements satisfaction. A traceability process usually involves an engineer creating links between fragments of natural-language text, supposedly interrelated, to show each requirement's development path, from original source material to eventual system components.

When digital multimedia artifacts enter the requirements record, we encounter new challenges with traceability. The basis on which we establish traceability links depends on how the media types involved interact, and as a consequence, the nature of any relationships between elements might not be obvious from a syntactic or semantic perspective. Furthermore, we might have trouble precisely pinning down the links' exact sources and sinks. So, problems already known to traceability practice become compounded as engineers begin to use the latest technologies to explore and communicate their understanding of a project's requirements.

Figure 1 illustrates a simple example of using media in requirements engineering. The figure shows media that might be featured during a workplace interaction in which engineers are in the early stages of determining requirements. Data captured for the record come from both digital video and audio recording and from traditional print media. The "Interview" sidebar shows part of the dialogue that might be exchanged during this interaction.

Understanding Media

In creating any representation with which to communicate, an engineer chooses—often unconsciously—some particular system of signs to work with. Sign systems are essentially abstract media made concrete or reified as some rendered representation or notation. Written and spoken natural language, sounds proper to their domain, and moving images are all examples used in requirements elicitation (Figure 1, first column).

Once, all these sign systems and their many sophisticated derivatives were associated with particular physical media—that is, their vectors or carriers. Written natural language and its structured forms, such as a novel or biography, became synonymous with "print" or "book"; animations and

0740-7459/09/$25.00 © 2009 IEEE

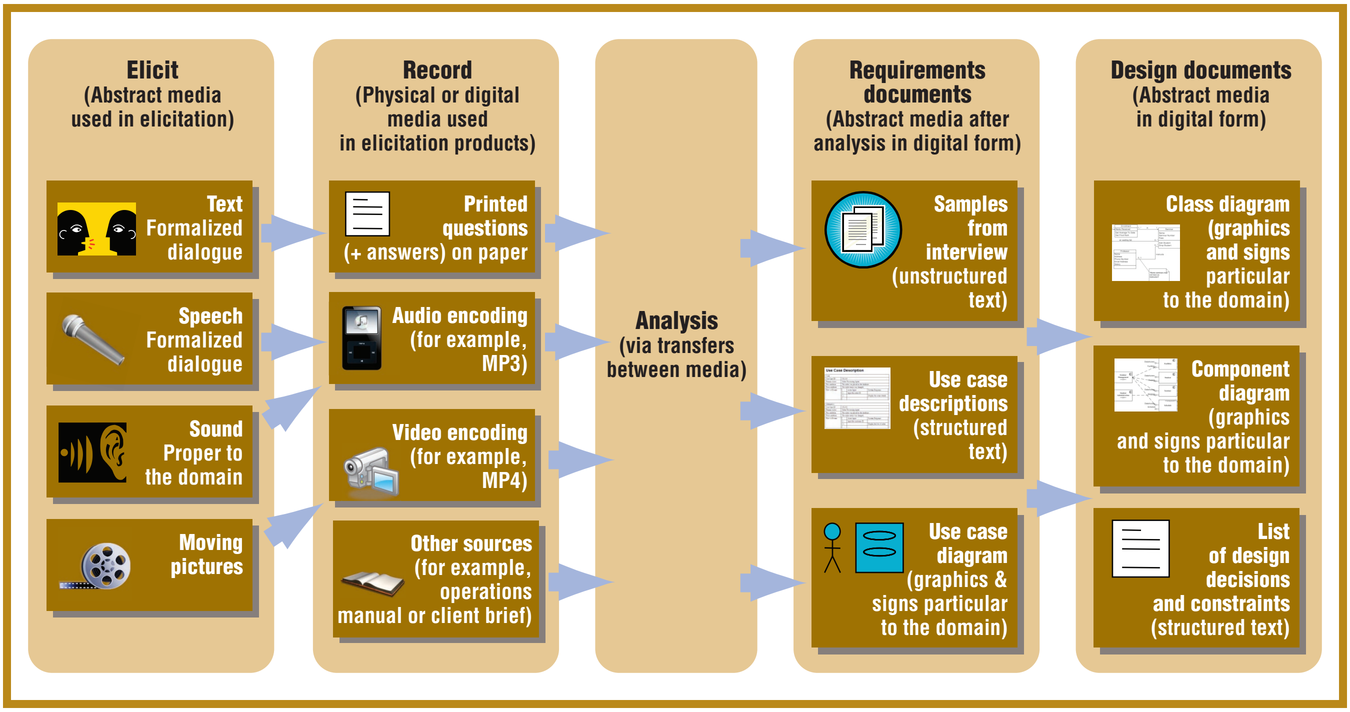

Elicit
(Abstract media used in elicitation)
Text
Formalized dialogue
Speech
Formalized dialogue
Sound
Proper to the domain
Moving pictures
Record
(Physical or digital media used in elicitation products)
Printed questions (+ answers) on paper
Audio encoding (for example, MP3)
Video encoding (for example, MP4)
Other sources (for example, operations manual or client brief)
Analysis
(via transfers between media)
Requirements documents
(Abstract media after analysis in digital form)
Samples from interview (unstructured text)
Use case descriptions (structured text)
Use case diagram (graphics & signs particular to the domain)
Design documents
(Abstract media in digital form)
Class diagram (graphics and signs particular to the domain)
Component diagram (graphics and signs particular to the domain)
List of design decisions and constraints (structured text)

Figure 1. Transfers between media in requirements. We can use multiple media, held in a variety of physical and digital forms, to elicit, record, and document requirements. The process involves many and varied transformations of these media, all of which can affect the final outcome and the quality of traceability.

Interview

In this example, an interviewer (Q) seeks information about how an everyday workplace activity occurs. A representative worker (A) answers questions while seated in a side office that opens onto the control room in which all activity occurs.

Q: And if you can't decide what to do?
A: I ask him, or her.
[A points at two colleagues.]
Q: What do they do?
A: He looks it up in this.
[A picks up a manual. Q writes down the title.]
A: She usually knows.
[A shrugs.]
Q: And if they're not here?
[Q and A both look around the control room.]
A: Jim always says [to] call that other guy.
[A points to a third colleague. Q writes down "Alternative: Call other."]

visual narrative were, and often remain, simply "cartoon," "film," or "video," each a term derived from a physical carrier. We can now find all these types of communication, including animated representations, in combination because their physical carriers are no longer disparate. A single, virtual digital medium can replace a multiplicity of simple and composite physical media. Whether a text with embedded pictures and speech is in a cloud of machines or on a simple memory stick is irrelevant, so long as it's available in known formats. This is increasingly the case with the requirements record (Figure 1, second column).

The apparent ease with which we can now create and manipulate multiple abstract sign systems opens the way for their more prevalent use in requirements engineering as old and new media are created, combined, and changed. This looks like a win-win situation until we examine its implications more closely.

A Lot Said, a Lot Captured, a Lot Lost

The fourth column in Figure 1 shows the artifacts constituting the requirements documentation, all now finalized normally in specific digital formats. The two uppermost documents use natural language in different ways—first, in its original "primary" and unstructured form as interview excerpts and then as use case descriptions structured via a "secondary" modeling system. The bottom document, the use case diagram, employs what we might call a "tertiary" modeling system using graphical forms to represent aspects of the secondary system via a defined set of notational signs. In this case, we give natural language this primary status because it's the basis for the others and provides a way to interpret the data as a last resort. This view is based on a theoretical position from the Moscow Tartu school of semiotics.[1] In cases where UML class diagrams depict initial design decisions, as per the fifth column in the figure, the primary modeling system itself changes from natural language to some object-oriented programming language—

Traceability and Multimedia

Media incur costs. If we insist on attempting to capture absolutely everything in our projects, we should establish an approach that distinguishes between those media fragments that are worth keeping ready-to-hand and those that we can archive just for a full record. We need to be selective when creating a media-enhanced record for traceability purposes because it can be challenging enough to get the simplest of text-based traceability in place. Here are a couple of places to start.

Focus on the Media Translations

Translations between media aren't necessarily bidirectional and, as such, can cause permanent discontinuities in the traceability record. For our example interview (see the "Interview" sidebar), if we create a textual description of the background control-room activities on the video, then a significant difference will likely exist between the information content in these two artifacts. A media translation from video to text involves unavoidable information loss and is therefore one-directional. Recreating the particulars and atmosphere in the control room would stretch even the most accomplished writers' abilities.

Focus on Media Mergers and Amalgamations

When separate traceability paths merge as we form a new artifact, the ability to untangle the original sources' contributions can be challenging. This is complicated further when the new artifact has amalgamated previously distinct and different media. In our example, if we recorded on video and noted the exchange, a storyboard of the issue-resolution process might initially feature a telephone call. Learning later that the word "call" in fact just means to appeal in some way to yet another colleague would require correction. Accounting for changes such as this, which we must propagate from an original source as a new understanding emerges, can be problematic. Attempting to trace back the provenance of this putative telephone call would demand juxtaposing the gesture in the video, the recording of the spoken word, and the simultaneously written interview note.

an even more fundamental change but one outside the normal scope of requirements engineering concerns.

The combination of media used to carry the information content—and the various media transformations that occur as elicitation, recording, analysis, and documentation proceed—influence the degree and type of traceability that's possible. Moving between sign systems is always problematic. Transformations from one general abstract sign system to another come in many forms, which all have ramifications for traceability.[2] The "Traceability and Multimedia" sidebar examines a couple of these in more detail.

Although it might be expedient when engineering requirements to "Just get it all on video!" such decisions can ultimately compound traceability problems. Storing, using, and maintaining extensive media-rich materials is far more costly than creating them in the first place. You can easily underestimate the added time necessary for analysis or ignore new media types' long-term accessibility. In an era in which we elicit, record, analyze, and document requirements in a potentially limitless variety of media, engineers must appreciate the media transformations taking place if both the preservation of full information content and the potential for rich traceability are priorities.

References

1. M. Groden and M. Kreiswirth, eds., *The Johns Hopkins Guide to Literary Theory and Criticism*, Johns Hopkins Univ. Press, 1997.
2. O.C.Z. Gotel and S.J. Morris, "Macro-level Traceability via Media Transformations," *Proc. 14th Int'l Working Conf. Requirements Eng.: Foundation for Software Quality* (REFSQ 08), LNCS 5025, Springer, 2008, pp. 129–134.

Olly Gotel is an assistant professor of computer science at Pace University in New York City. Contact her at ogotel@pace.edu.

Stephen Morris is an honorary senior visiting research fellow in the Department of Computing at City University London. Contact him at sjm@soi.city.ac.uk.