Macro-level Traceability Via Media Transformations

Orlena C. Z. Gotel[1] and Stephen J. Morris[2]

[1] Department of Computer Science, Pace University, New York, USA
ogotel@pace.edu
[2] Department of Computing, City University, London, UK
sjm@soi.city.ac.uk

Abstract. This paper proposes an alternative approach to the examination of artifacts whose contents must be traceable to promote software quality. The approach places emphasis on media use and media transformations. We suggest that one cannot begin to assign and sustain traceability relations at a micro-level between units of content if the sign systems that have been created and transformed to represent this content are not considered at a more macro-level. Our approach is anticipatory, feasible to automate and exemplified.

Keywords: Media Transformation, Multimedia, Requirements Traceability.

1 Introduction

Recent traceability research has focused on establishing links between semantically similar terms to identify automatically content-based traceability relations between the artifacts of software development [2]. These approaches account for textual artifacts and, to a limited extent, the textual characteristics of structured diagrams. While they address some of the problems associated with traceability [5], they ultimately lend themselves to natural language ambiguity and many artifact types are precluded. The premise of our work is that artifacts relevant to the trace record will be held in multiple media in the future, especially those generated during upstream requirements-related activities, including the results of observational studies or sketches drawn by stakeholders. Video fragments from elicitation sessions are already used to provide supporting rationale for requirements in some contexts [3] and a vision of video-based requirements engineering continues to gain clarity [7].

Presuming a media-rich software development environment, we suggest that you first have to be precise about the nature of the relation between the underlying media types before you can say what the implications are for content change and any particular traceability relation between artifacts. The underlying assumption is that there is no such thing as a pure element of content, only some representation of it as an artifact. It is therefore essential to understand the process whereby representations come into being and are transformed because this is the only way to understand what happens to content. We extend previous research by marrying traceability with multimedia production to propose an approach through which to make decisions about media choices, combinations and transformations when seeking to create or recover a representative trace record [6]. The concept for the approach is exemplified to highlight its potential value in framing a familiar topic from a new perspective.

B. Paech et al. (Eds.): REFSQ 2008, LNCS 5025, pp. 129–134, 2008.

2 Media Use in Requirements Engineering

Any subject matter being communicated has an associated medium which is its carrier, or vector, in the physical sense. In a second sense media are abstract; they are agencies for the communication of subject matter. As such they are separate sign systems, the most common and most significant being natural language text and speech, graphics, still and moving images, and specialised systems such as numbers, mathematical and computer languages [9].

Fig. 1 shows how media may be involved in a simple requirements engineering scenario. During an elicitation interview, the respondent can indicate other relevant people, activities or documents. The range of potential responses and their referents suggests the need for text, video and sound recording. A minimal set of other material (e.g. an operations manual and a client's briefing document) represents sources in conventional print media. If the interviewer asks questions from a pre-written questionnaire, the abstract medium is written natural language and the physical medium is ink on paper. The spoken questions and answers comprise spoken natural language carried by sound waves. If the interview is recorded, moving pictures, natural language speech and sounds proper to the domain (e.g. traffic) are captured. Responses and supplementary detail may be recorded as written natural language text and images on paper. The primary source material is thus rich in media.

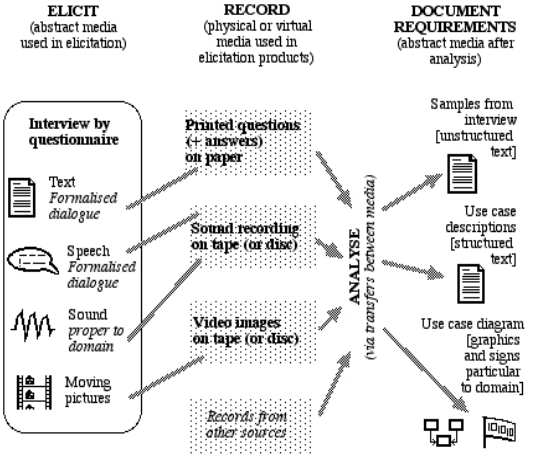

Fig. 1. The media in a simple requirements engineering scenario

During what is loosely defined as 'analysis', a series of media transfers take place. An audio segment may be transcribed and augmented with contextual information from the video to gain gesture and gaze detail. Progressively, the abstract media will be reduced in variety and from amongst the abstract media initially involved only text survives when requirements are first formally documented. Moreover, only the textual samples from the original interview can be unequivocally traced back to an original source. Although every output is now within the single digital medium, the complexities and implications of media transfers remain hidden.

3 Media Transformations

In earlier work, the concept of media transformations for describing and prescribing changes to abstract media elements was developed [8]. We give examples below as they apply to the scenario above and explain the implication for traceability.

Origination: Examples would be text of the questionnaire responses, recorded video images, speech and other sounds from the interview environment, text and graphics in the operations manual, or text and images in the brief. This transformation should guarantee the identification of all primary source material and provide starting points for forwards traceability and all end-points for backwards traceability.

Amplification: An example would be the elaboration of the interview text transcribed from the audio recording with the text notes of the visual indicators in the video. Use should provide the opportunity to identify partial content changes, not involving the complete merger or amalgamation of media that might otherwise be without an identifiable source, along with the ability to retrace provenance in the original context. This may be important for forwards and backwards traceability where clarity of change is significant and wider history is relevant to understanding.

Revision: Examples would be alterations to the text of the questionnaire responses to ensure they reproduce the recorded speech, or the structuring of pre-existing text to form use case descriptions. Differentiating from amplification, elements of an artifact are completely replaced by elements in the same medium as opposed to extended. This transformation demands identification of the basis for any revision, even if this requires the origination of a primary source. The implication is the possible need to incorporate an element of rationale into the trace record.

Translation: Examples would be the translation of interview speech to written text, the video images to text of the content, or the use case descriptions to the use case diagram. Switching between abstract media involves representation in an alternative sign system and there are implicit losses involved [1]. This is problematic, and irreversible for backwards traceability, where signs are wholly undifferentiable one from another either syntactically or semantically [4]. The translation from the video of the interview into text would be subject to such restriction. Even recorded speech to text, although guided by transcription conventions, comes with some losses.

Outline: An example would be the list of use cases from the text in the operations manual. With neither abstract medium nor domain of content changing, this transformation should not cause problems for traceability, but is subject to the restriction that detail is lost. Where the media vary and are reduced in number, say from a video, speech and sound recording to text, it is important to know for traceability whether the outline represents indexical properties.

Merger: An example would be text from the interview answers combined with text derived from extraneous dialogue. This transformation indicates the fusion of pre-existing paths, henceforth treated as a single path. The complete merger of elements in the same abstract medium should not be problematic unless it is important to differentiate between the contributions of sources, to propagate forward changes from things prior. Traceability implications depend on the volatility of contributing paths.

Amalgamation: An example would be the fusion of some of the text from the use case description with graphics to form the use case diagram. The process of specification is one in which the number of abstract media used is steadily reduced,

often to text and notational sign systems such as UML. Only if the separate abstract media remain differentiable is the traceability in either direction unimpaired.

Proxy creation/use: An example of creation would be the questionnaire text to be spoken; use would be the questions spoken. Tracing depends on an accurate mapping between elements in the different abstract media (e.g. direct mappings between text and the spoken version). Proxy transformations clarify the role of artifacts, such as a storyboard standing in place of a design, but present issues for any trace.

Substitute creation/use: An example of creation would be any UML representation that is an intermediate artifact between requirements and implementation; use would be transitional UML models prior to implementation. These transformations play an important role in the exploration of concepts and promote trace continuity.

Comparate use: When one artifact influences another, but does not participate in a media transformation, it acts as a basis for comparison. This transformation contributes to the accuracy of others, thus to the completeness of the trace record, and so to the effectiveness of traceability. An example would be the recorded speech used for comparison with the written answers to the questionnaire to check for accuracy.

4 Media-Based Traceability

In software engineering, improving lifecycle-wide traceability involves defining an exact path for tracing and understanding the alteration of information content along this path over time. Where this involves a range of abstract media representations, carried by physical and digital media, there are many paths that might be followed. One set of paths for the scenario is given in Fig. 2. It shows how media transformations may provide a practical means of linking artifacts at a high level.

The media transformations in the scenario begin with origination, shown by the creation of primary sources via the interview or preparatory tasks (1.1-1.5). Given the varied abstract media, translation into a common abstract medium may be necessary to enable the construction of new artifacts. Switching between abstract media is a translation (2.1, 2.2) and involves content discontinuity. The trace path and record will be impaired where translation is undertaken, implying the potential need to revisit original sources. Whether an artifact is being used as a subsequent comparate in its original medium or in a translated form also has implications for what is re-examinable. Translation presents a problem for traceability since it is not bi-directional without effort. When there are choices between using media other than natural language text, we should be able to determine which media are going to be the most problematic and know what measures to take to preserve future traceability.

Four transformations affect changes within a single abstract medium: amplification, revision, outline and merger. The scenario shows an amplification (2.3) where an element in one medium elaborates another of the same medium and a revision (2.4 etc.) where an element replaces one of the same medium. One would expect to see many such transformations within software development since its essence is the distillation of content. Both these transformations demand versioning information for understanding. Amplification typically requires a subsidiary artifact to ensure backwards traceability. This is not so for revision, which may simply follow from changes to the prior artifact. However, if the basis for the revision is relegated to memory, rationale is not automatically retrievable and the path not easily reversible.

1 Primary source material [in multiple media and multimedia]
0.1 *proxy creation*: text of questionnaire to be spoken in interview
0.2 *proxy use*: questions spoken in interview
1.1 *origination*: text of responses to questionnaire
1.2 *origination*: recorded video images
1.3 *origination*: speech and other sounds from interview and work environment
1.4 *origination*: text and graphics in operations manual
1.5 *origination*: text and images in client brief
2 Conversion to text [transferring back to the 'primary modelling language' of natural language] – now within digital carrier media
2.1 *translation*: speech in recorded interview *1.3* to written text
2.2 *translation*: video images *1.2* to text notes of content
2.3 *amplification*: elaboration of interview text *1.1* with *comparate use* of text notes of video *2.2*
2.4 *revision*: alterations to text of responses *2.3* with *comparate use* of speech recorded *1.3*
2.5 *merger*: text from structured interview answers with text derived from any open-ended exchanges or extraneous dialogue *2.1* and *2.4*
2.6 *outline*: text list of possible use cases from interview text *2.5*
2.7 *outline*: text list of possible use cases from operations manual *1.4*
2.8 *outline*: text list of possible use cases from client brief *1.5*
3 Documentation [transferring to a 'secondary modelling language' of structured natural language]
3.1 *revision*: structuring of interview text *2.5* to form initial use case descriptions
3.2 *revision*: structuring of operations manual text *1.4* (and possible *translation* from graphics) to form initial use case descriptions
3.3 *revision*: structuring of client brief text *1.5* (and possible *translation* from images) to form initial use case descriptions
3.4 *merger* and *revision 2.6, 2.7, 2.8*: use case list
3.5 *merger* and *revision 3.1, 3.2, 3.3*: use case descriptions
3.6 *proxy creation*: use case descriptions *3.5* in part for later UML activity diagrams
3.7 *revision*: sample extractions from interview text *2.5* on basis of *comparate use* of use case list *3.4*
4 Modelling [transferring to a 'tertiary modelling language' using non-textual components as the foundation for representation]
4.1 *outline* and *translation*: use case list *3.4* to use case names and diagram elements; use case descriptions *3.5* to associations in diagram
4.2 *amalgamation* and *substitute creation*: elements from *4.1* brought together in a use case diagram (or model)

Fig. 2. Traceability from a media-based and transformational perspective

The outline transformation, used in the scenario to structure use case descriptions (2.6-2.8), provides a précis version without media change. This appears innocuous, but if the source artifact combines abstract media, the textual outline may be a result of implicit translations coupled with additional undocumented information derived from the juxtaposition of media. An outline in one medium derived from multiple media presents a potential break in continuity for the trace record and may need to be re-examinable in its wider derivation context.

When two or more elements in the same medium are combined to form another, the result is a merger. While not directly reversible, merger transformations within a single abstract medium minimize content loss. The first merger (2.5) combines text that has been translated from speech with text elaborated with textual information derived from video. This artifact could equally have been constructed as a result of revisions/amplifications, with comparate use of primary sources, but this would have had negative consequences for the trace record because the absence of discrete intermediary stages compounds impact analysis. Where a merger takes place with an accompanying revision/amplification (3.4, 3.5), the traceability path is likely to be jeopardized unless a supporting artifact is provided.

Amalgamation transformations (4.2) should be common in requirements engineering since they are used when constructing use case models. However, the combination of media elements of different types, whatever the medium of the result, may be even more unpredictable in outcome than translation. This is the case where it is difficult to untangle the contributing media elements and their individual paths for traceability. The choices made as to the types and ordering of transformations is of interest because amalgamation can be avoided, for example, by translation into a common medium (e.g. text in the scenario) followed by merger. Such ordering will alter the path through which requirements are engineered, in one case retaining some ability to retrace separate contributing paths. Amalgamations are worth examining if preserving the integrity of the trace record is crucial and effort should be made to retain separation potential if parts of the embedded content are likely to change.

5 Towards a Framework for Macro-level Traceability

There are implications for content loss or gain when different media types are used in software development, impacting traceability. Decisions need to be made as to the use of media, combinations of media and the ordering of transformations between media as content from those artifacts constituting primary source material is created and transformed into specification and code. Our research seeks to provide a framework and develop guidelines to help engineers take these decisions in support of their anticipated traceability needs. We suggest that an understanding of the artifact collection, at a foundational and representational level, is critical for contextualizing more discerning forms of traceability, irrespective of manual provision or automated recovery. The routine generation of macro-level traceability between media-rich artifacts should not be an insurmountable task for future requirements management environments and integrated guidelines would alert to and help mitigate critical traceability issues, focusing effort only when and where most needed given the potential costs incurred. An exploration of these important topics, accompanied by validation of the underlying approach, forms our on-going research.

References

1. Anderson, P.B.: A Theory of Computer Semiotics. 2nd edn. CUP, Cambridge (1997)
2. Cleland-Huang, J., Settimi, R., Romanova, E., Berenbach, B., Clark, S.: Best Practices for Automated Traceability. IEEE Computer 40(6), 27–35 (2007)
3. Gall, M., Bruegge, B., Berenbach, B.: Towards a Framework for Real Time Requirements Elicitation. In: 1st Intl. Workshop on Multimedia Requirements Engineering (with 14th IEEE Intl. Requirements Engineering Conference), Minneapolis, MN (2006)
4. Goodman, N.: Languages of art: An approach to a theory of symbols, 2nd edn. Hackett, Indianapolis, IN (1976)
5. Gotel, O.C.Z., Finkelstein, A.C.W.: An Analysis of the Requirements Traceability Problem. In: 1st IEEE Intl. Conference on Requirements Engineering, Colorado Springs, CO, pp. 94–101 (1994)
6. Gotel, O.C.Z., Morris, S.J.: Crafting the Requirements Record with the Informed Use of Media. In: 1st Intl. Workshop on Multimedia Requirements Engineering (with 14th IEEE Intl. Requirements Engineering Conference), Minneapolis, MN (2006)
7. Jirotka, M., Luff, P.: Supporting Requirements with Video-Based Analysis. IEEE Software 23(3), 42–44 (2006)
8. Morris, S.J.: Media transformations for the representation and communication of multimedia production activities. In: Sutcliffe, A., et al. (eds.) Designing Effective and Usable Multimedia Systems, pp. 72–85. Kluwer Academic Publishers, Norwell Mass (1998)
9. Morris, S.J., Finkelstein, A.C.W.: Engineering via discourse: Content structure as an essential component for multimedia documents. International Journal of Software Engineering and Knowledge Engineering, World Scientific Pub. Co. 9(6), 691–724 (1999)