Integrating Claim Quality and Risk into a Usability Engineering Tool

Mathew Davis, Kai Tang, Jamie L. Smith, and D. Scott McCrickard

The Center for Human Computer Interaction and Department of Computer Science
Virginia Polytechnic Institute and State University
Blacksburg, VA 24060-0106 USA
{madavis6, ktang, jls05, mccricks}@vt.edu

Abstract

This paper described the instantiation of claim quality to allow a repository system to rate usability claims, helping designers establish up front the credibility of a claim in a library. In so doing, we introduce a web-based algorithm and interface that enhances interface development by providing quality measurement and risk assessment to designers of a proposed system. Integrated as part of LINK-UP, a usability engineering process and system, our combined approach for this tool focuses on the analytical evaluation process to help notification system designers see potential design problems, called risks, that reside within their systems. This work describes the motivation, the methodology, the implementation, and directions for future improvement of our development.

1 Introduction

1.1 What is LINK-UP

LINK-UP, or Leveraging Integrated Notification Knowledge with Usability Parameters, has been developed and validated as an integrated environment to facilitate notification system designs [1]. Fundamentally built on a scenario-based design approach [2], LINK-UP supports a reusable library to give notification system designers the advantage of sharing other designers' knowledge on a particular system. As shown (Figure 1), LINK-UP captures design knowledge in the form of *claims* to assist in reuse in interface development. For more information on LINK-UP, please refer to [1] and [4]; for an overview of risk management and how it relates to our approach, see [5].

1.2 What is a Claim

As Rosson and Carroll defined in their book, a claim is a statement about possible effects (benefit and weakness, or upside and downside) of a designed feature on users within a usage scenario [2]. In general, claims consist of a title, description, scenarios, theories, and artefacts that support the description [3]. For example, a claim may look like:

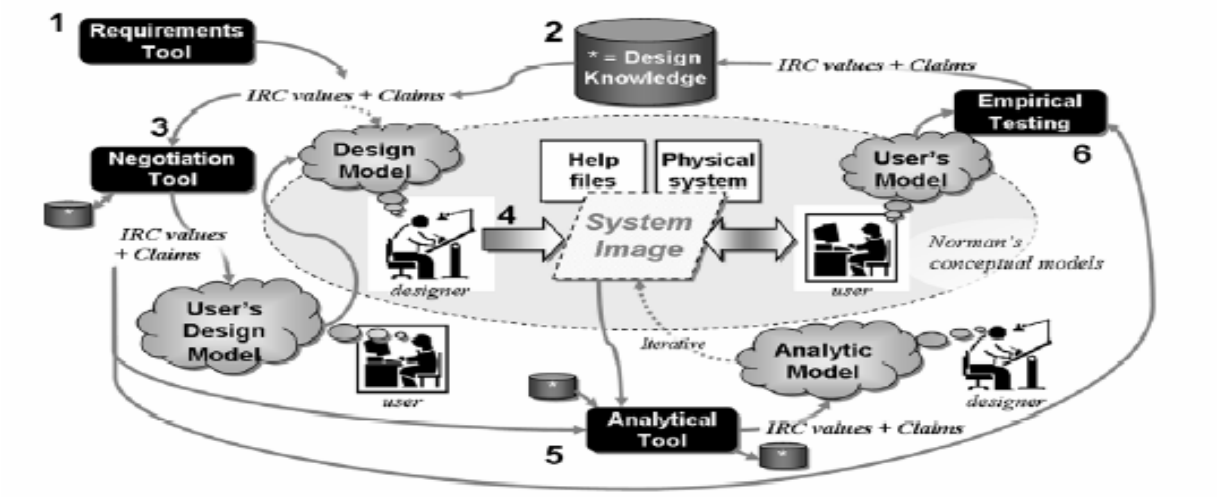

Figure 1: Overview of LINK-UP system as claim-based repository for reusable engineering.

Use of tickering text-based animation to display news headlines in a small desktop window:

+ Preserves user focus on a primary task, while allowing long-term awareness
- BUT is not suitable for rapid recognition of and reaction to urgent information. (as described in [1])

As mentioned earlier, LINK-UP treats claims as reusable design knowledge. However, similar to many knowledge repositories, it is difficult to determine if a claim in the LINK-UP reuse library is sufficiently more credible than any other claim. Thus, LINK-UP needs a model to assess and manage the risks brought by various claims designers create or reuse from the library. Claim quality and risk establishes this credibility and risk management for claims.

2 Claim Quality

We implement within LINK-UP a claim quality algorithm and interface using a quality rating system. There are five components that come together to make up this rating: author experience, quality of rationale, quality of artifact, degree of reuse, and average user rating. Once quality values are established for each component, the overall claim quality can be calculated using a claim quality formula. These ratings allow users to assess the credibility of a claim before incorporating that claim into a design.

2.1 Motivation

When notification systems designers know the credibility of a claim before including it in a design, there is improved design. A core idea of LINK-UP is that a large growing library of claims is available for a designer to use making improvements on a design. When a designer is considering including a claim in a design, if there is little to no meta-data included about that claim then the designer is at a disadvantage and likely to make a poor decision as to which claim to include. The designer does not know if the claim author is an experienced usability engineer or a student who is just learning the basics of HCI. There could be many other questions that the designer has concerning the credibility of the claim.

2.2 Methodology

Quality of claim is based on a claim rating. This rating is not meant to be a static number that is assigned to a claim at creation and kept throughout the lifetime of the claim. The rating is dynamic and adjusts and shifts itself as the claim is evaluated, updated, and used in designs. The current rating system assigns a claim a number between 1.0 and 4.0 with 1.0 being of the lowest quality and 4.0 being the highest. The rating is calculated using five factors concerning the claim.

2.2.1 Rationale Quality

Each claim in the library has trade-offs to including the claim in a design. Rationales are associated with each trade-off that gives reason for why the trade-off is positive or negative. These rationales may be established rationales that are published. A rating is given to a rationale for each level of publish including 4.0 for a journal, 3.0 for a full conference paper, 2.0 for a short conference paper, and 1.0 if it is not published. In this way, credibility is given to the claim based on the merits of published rationales. The rationales could also report results of a study. Another rating is associated with a rationale for each level of results including 4.0 for a large-scale empirical evaluation, 3.0 for a small-scale empirical evaluation, 2.0 for analytical evaluation, and 1.0 for no results. This should increase the quality of the claim based on having established results reported. The rationale quality factor is established by taking the rating for these two factors and averaging them (Equation 1). The averages of all rationales in a claim are then averaged over the number of rationales and the result is the rationale quality (RQ) factor (Equation 2).

Equation 1

$$avg = \frac{venue + results}{2}$$

Equation 2

$$RQ = \frac{sum(avg)}{num\_tradeoffs}$$

2.2.2 Author Experience

Each claim has an author. The author's prior experience in the field of HCI and in designing notification systems needs to be reflected in the claim quality. A system administrator will oversee the ratings of each author and update them as their experience increases. An author's experience is broken down into two factors. The first factor is knowledge of HCI and design theory. The author will be given 1.0 for no knowledge, 2.0 for novice level, 3.0 for intermediate level, and 4.0 for expert level. The second factor is experience designing notification systems applications. The author will be given 1.0 for no experience, 2.0 for novice experience, 3.0 for moderate experience, and 4.0 for advanced experience. The author experience (AE) factor is then calculated by averaging the two sub-factors.

Equation 3

$$AE = \frac{theory + application}{2}$$

2.2.3 Artifact Quality

Each claim can have an artifact associated with that claim showing the aspects of the claim through multimedia. This adds depth to the claim and should increase the quality in accordance with the quality of the artifact. The artifact quality (AQ) is assigned according to the level of detail in the artifact. In no artifact is present, artifact quality will be 1.0. An artifact will be given 2.0 for one or more screen shots, 3.0 for an animated movie depicting functionality, and 4.0 for an interactive prototype.

2.2.4 Degree of Reuse

Each claim can be reused in a number of projects over the lifetime of the claim. If a claim is used over and over again in notification system designs, this should be a tip to other designers that the claim is popular because it is useful in design. A claim is assigned a degree of reuse (DR) rating factor according to how many times the claim has been reused. For reuse in 0 to 1 projects a rating factor of 1.0 is given, 2 to 5 projects a rating of 2.0, 6 to 9 projects a rating of 3.0, and 10+ projects a rating of 4.0. In this way a claim draws credibility for having a long lifetime of use in the library.

2.2.5 Average User Ratings

Each designer or user in LINK-UP has the capability to give a rating for a claim. The claim can then reflect the reliability as decided by the designers using the claim. User ratings (UR) are calculated by combining the number designers who rated the claim with the credibility of those designers. When evaluating ratings given by designers, author experience (AE) needs to be evaluated in order to assess credibility of the designer giving his opinion. Thus, a system has been developed to calculate the number of user ratings based on author experience. If a user has an AE of 1.0 then his or her rating only counts as 25% of a rating, 2.0 AE counts as 50%, 3.0 as 75%, and 4.0 as 100%. By using this system a number of rating variable (num_ratings) is established. Then the actual average rating for all users is established (avg_rating) and the two are combined to get the degree of the reuse (DR) (Equation 4).

Equation 4

$$UR = num\_ratings^{\frac{avg\_rating}{5}}$$

2.2.6 Claim Quality Rating

The claim quality is then calculated by combining the factors. Each factor is weighted according to what we felt is the most important in assessing claim quality. The weights (Figure 2) and overall claim equation (CQ in Equation 5) are as shown.

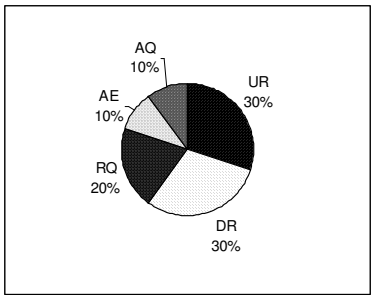

Figure 2: Weights of quality factors when calculating overall claim quality

Equation 5

$$CQ = \frac{(2*RQ)+(AE)+(AQ)+(3*DR)+(3*UR)}{10}$$

3 Risk Analysis Module

To manage the risk levels designers face with reusable claims from the repository, we develop the risk-driven, claims-based management model to prioritize claims according to their degree of risk exposure, with a high exposure value resulting in a higher priority rating. Numerous factors that contribute to a claim's risk exposure include: claim quality (discussed in previous section), IRC difference factor, stakeholder concern ratings, external rationale factor, internal testing factor, and mitigation factor. Necessary information for the calculation is collected from claim quality computation, from the claim author during claim creation, from other designers who reuse the claim in their own projects, from project stakeholders during participatory design, and from system evaluators during either an analytic of empirical evaluation of the interface.

We calculate the overall claim risk exposure (CRE) (equation 6) as a combination of base claim exposure value (BCE) and the maximum downside exposure value (DSE) for all downsides included in the claim.

Base claim exposure (BCE) (equation 7) accounts 20% of the total risk level of a claim, and it considers the claim quality rating (CQ) and the IRC difference factor (IRCD) between designed system IRC and each individual claim IRC. Claim quality and the IRC difference factor together represent an initial measurement of the probability that a given claim will cause a problem in system design.

Equation 6

$$CRE = BCE + \max_{claim}\{DSE\}$$

Equation 7

$$BCE = IRCD^{(\frac{CQ}{4})}$$

The other 80% of the total risk level comes from the maximum downside risk exposure form all the downsides reside within a particular claim. Since a downside represents the weakness of a designed feature within a usage scenario [2], each downside of a claim is assigned a downside exposure value (DSE). Instead of a summing all the downside exposure values in a given claim, we use one maximum downside exposure value, to exemplify the concept that claims with one critical downside are more likely to possess higher risk values than claims with multiple non-critical downsides. Thus, the most critical downside of a given claim will primarily determine the priority of that claim. This method of calculation maintains the simplicity of prioritizing claims while preserving the significance of prioritizing individual downside risks.

Each downside exposure value (DSE) (equation 8) accounts for four key factors:

- *Stakeholder concern ratings(SC):* this value gets assigned by system stakeholders based on end-users' goals of the system. A higher stakeholder rating would result in a higher risk exposure value and indicates a higher priority to mitigate the downside
- *External rationale factor(R):* this value is assigned as a combination of the relevance and the quality of each rationale of the downside. The more relevant and more confident of the rationales, the lower the downside risk exposure value.
- *Internal testing factor (T):* this value gets assigned during an analytical or an empirical evaluation. It measures the degree to which the downside affects current system and how confident such measurement represents. A higher internal testing factor would result a higher risk exposure value.

- *Mitigation factor (M):* this value depends on whether or not another claim in the system has a mitigation relationship with the claim in question. Downside mitigation reduces the risk exposure value by 50%.

Equation 8

$$DSE = \left[\frac{SC(\frac{1}{R} + 2T)}{3} \right] * M$$

4 Conclusion

4.1 Benefits

Once claim quality is established, it allows designers to work with the most highly credible claims in the library. It also allows good claims with credibility to increase its rating over the lifetime of the claim and become more visible to designers. Designers can now search for claims in the library and view those claims along with their quality ratings. Claim quality ratings allow users to be more informed when making the decision of which claims to include in a design.

Combined with claim quality and other designers' inputs, the web application for risk exposure analysis enables users to monitor their project risks, to analyze and to prioritize the risks, and eventually to mitigate such risks in the attempt for a better interface design.

4.2 Future Improvements

In the future claim quality could be extended to include different factors, adjusted weights of current factors, and completely different quality rating systems. In this way claim quality is quite dynamic. As claims develop more factors, these factors may need to be included in claim quality. There is possibility for future exploration into additional factors being added to claim quality. After some empirical evaluation of the claim quality system is completed, an adjustment in the weight factors could be possible to better reflect claim quality. Additional quality rating systems could also be added on top of the current system. Including more than one rating system could give the designers a choice as to which quality rating to use.

References

[1] Chewar, C. M., Bachetti, E., McCrickard, D. S., and Booker, J. (2004). "Automating a Design Reuse Facility with Critical Parameters: Lessons Learned in Developing the LINK-UP System." In *Proceedings of the 2004 International Conference on Computer-Aided Design of User Interfaces (CADUI '04)*, Island of Madeira, Portugal, pp 236-247.

[2] Carroll, J.M., Rosson, M.B. *Usability engineering: scenario-based development of human-computer interaction* (New York, NY: Morgan Kaufman, 2002).

[3] Carroll, J. M., Rosson, M. B., George, C Jr., & Koenemann, J. "Requirements Develpement in Scenario-Based Design." *IEEE Transactions on Software Engineering.* 24(12), 1998.

[4] Lee, Jason Chong., Lin, Sirong., Chewar, C. M., and McCrickard., D.S. "From Chaos to Cooperation: Teaching Analytic Evaluation with LINK-UP." In *Proceedings of the World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education (E-Learn '04)*, Washington DC, November 2004, pp. 2755-2762.

[5] Smith, J. L., Bohner, S. A., and McCrickard, D. S. (2005). "Project Management for the 21st Century: Supporting Collaborative Design through Risk Analysis." In *Proceedings of the ACM Southeast Conference (ACMSE '05)*, Kennesaw GA, pp. 2-300 - 2-305.