IDEA GROUP PUBLISHING

701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com

ITP4898

Evaluating the Evaluator: Modeling Systematic Data Analysis Strategies for Software Selection

Joanna Lumsden
NRC IIT eBusiness, 46 Dineen Drive, Fredericton, Canada, E3C 2P2, jo.lumsden@nrc.gc.ca

INTRODUCTION

Developers of interactive software are confronted by an increasing variety of software tools to help engineer the interactive aspects of software applications. Not only do these tools fall into different categories in terms of functionality, but within each category there is a growing number of competing tools with similar, although not identical, features. Choice of user interface development tool (UIDT) is therefore becoming increasingly complex.

Evidence suggests that, when selecting UIDTs, industrial software developers rely heavily upon *right-brain* (intuitive) decision-making [13], based on little more than an *ad hoc* inspection of marketing material, journal reviews, and recommendations from colleagues [9]. Although potentially adequate in some cases, this can result in poor choices, especially when the aforementioned information is not relevant to the context in which the tool is to be used. Acknowledging the existence of, and place for, intuition or non-rational decision-making in the overall process of UIDT evaluation and selection, this research facilitates *integrated* style [13] UIDT selection; it provides a mechanism which allows evaluators to approach UIDT selection from an analytical perspective but which does not suppress intuitive decision-making so that, where necessary, it can be used to handle areas of uncertainty, such as trade-offs. By promoting an analytical approach to UIDT selection, it is possible to guard against inappropriate use of intuition and thereby increase the effectiveness of data analysis leading to greater likelihood of tool acceptance.

Central to this research is SUIT - a framework, method, and innovative visualization environment (VE) for evaluating UIDT suitability for software development products such that the selected tools are the most appropriate for their anticipated context of use [6]. Adopting a reference model-based approach to UIDT selection, SUIT can be used in several different ways [6] that reflect a pattern of maturation analogous with the Capability Maturity Model (CMM) of software development [4, 12]; pertinent to the discussion in this paper is SUIT's use to select the 'best-fit' UIDT for a project based on the project's specific context and requirements.

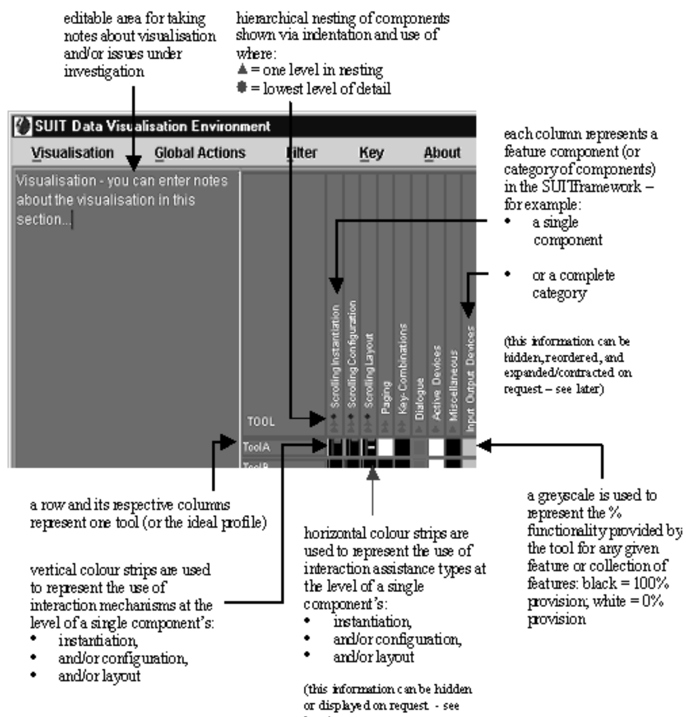
Previous publications have detailed various elements of the framework and method and have introduced the SUIT VE [6-8, 10]. Such publications have also discussed related preceding research and the background and motivation to this research, including decision-making and problem solving as it relates to UIDT evaluation [6, 7]; given space limitations imposed on this paper, we would refer readers to these publications for detailed information of this nature. This paper reports on the findings of an evaluative study designed to model the data exploration/analysis strategies adopted by evaluators when asked to assume an organisational position to select, using SUIT, a tool for a group of developers with experience and expectations orthogonal to their own. To situate discussion throughout the remainder of this paper, Section 2 provides a brief overview of the salient points of the SUIT data VE. Section 3 describes the evaluative study of the environment and discusses the results. Section 4 concludes with a discussion of further work.

AN OVERVIEW OF THE SUIT VISUALIATION ENVIRONMENT

The importance of context is recognised in software design. Clarke comments that: "Designers often neglect to take account of contextual factors due to their focus on the artefact itself" ([1], pg. 10).

Making the relationships between context and design explicit enables accurate judgement about the use of contextual information in design [1]. In the case of UIDT selection, evaluators often fail to adequately consider contextual information with the result that inappropriate UIDT selection is made [5, 9]. Often the choice that is obvious to an evaluator may not function in the context in which it is used due to cost, time and lack of acceptance. Problem solving and decision-making changes when an individual evaluator is asked to assume an organisational position to select a UIDT not for himself, but for members of a group. In these circumstances, evaluators are required to adapt their goals and values to their responsibility [15] – namely, the context of use of the selected UIDT. Without an adequate model of the UIDT's context of use, evaluators are prone to reverting to their individual preferences and goals. When developing the VE, it was therefore considered imperative to include a mechanism by which to record and reflect anticipated context of use during the selection

Figure 1: An example of part of a SUIT data visualization.



process, and to explicitly represent its influence over the suitability of any given UIDT. To achieve this, the SUIT VE includes a representation of the *ideal tool* - that is, a model which combines the functional requirements and context of use identified for the selected tool – alongside the data for *actual* tools such that the relationship between the context and the artefacts is made explicit and analysable. Thus, an evaluator has a convenient point of reference in terms of the context to which he is required to give due consideration.

Figure 1 shows part of a visualization that has been created to compare two fictitious UIDTs, annotated to highlight the main features. For each tool, the number of data components can total thousands, most of which are explicitly represented in the visualization; the remainder of the information is available on demand. To simplify a visualization, an evaluator can hide information at varying levels of detail; similarly, he can selectively juxtapose components/columns to ease visual comparison. On a category-by-category basis, an evaluator can select the level of detail to view by expanding/contracting columns in the component/category hierarchy. An evaluator can query the data by applying predefined filters (identified during an evaluative assessment of SUIT [6-8]) to the active data set; a textual history of filter application is maintained so the semantics of the active data set can be viewed at any stage. At any point, an evaluator can chose to take a snapshot of the result of his data analysis to maintain the state of the visualization for reference during the remainder of the decision-making process.

MODELING DATA ANALYSIS STRATEGIES

Generic information exists regarding the manner in which people explore and compare data (e.g. [14, 16]). There is, however, no established corpus of knowledge about the strategies adopted by evaluators during the visual exploration/analysis of UIDT data according to SUIT's project-specific evaluation concepts. In particular, there is no model of the manner in which an evaluator structures data exploration/analysis to make an evaluative decision *on behalf of* a client where the requirements of that client are orthogonal to those of the evaluator. The studies described here were therefore performed to establish an initial knowledge base – in particular, a model – of the strategies for data exploration/analysis adopted by evaluators asked to assume an organizational role when using the SUIT VE to make project-specific UIDT selection decisions guided by the SUIT method. It is anticipated that, not only will this knowledge inform future development of SUIT, but it will also contribute to an improved understanding of the manner in which data exploration/analysis might be structured for systematic third-party context-sensitive decision-making.

Performing the Studies

The investigative studies, which were designed to provide *qualitative* data about UIDT data exploration/analysis strategies, used a combination of direct (video-taped) observation, 'think aloud' and 'question asking' protocols, and an end of session interview. Five participants were selected on the basis of their abilities – all were software developers, of which one was a CASE tool expert, and one was an expert in the development and evaluation of a range of software tools. Prior to their evaluation session, each participant was given an outline [8] of the SUIT method and framework with which to familiarize themselves with the related concepts. They were each given a short tutorial on the VE and, provided with a small data set, given as much time as required to investigate/familiarize themselves with the environment before answering a series of questions designed to ensure they had all acquired the same basic understanding of the environment. After this preliminary familiarization stage, each participant was asked to return after 1.5 hours (to avoid user fatigue which a pilot study identified as a likely hurdle to the evaluation); on their return, participants were given 5 minutes to re-familiarize themselves with the environment. They were then asked to read an evaluation scenario which outlined their task; given a project profile, they were asked to assume the role of a third-party evaluator to select one of two commercially available, large scale UIDTs such that the selected UIDT best fit the identified needs of the target user group. The requirements identified in the profile were intentionally orthogonal to the known preferences of the participant group to allow for observation of the relative importance of the ideal tool profile within the VE and its role in the participants' data analysis strategies. After starting the video camera, the researcher asked each participant to outline his[1] proposed strategy for tackling the data exploration/analysis task. Participants were reminded to think aloud as they performed their task and advised that the researcher might ask questions to clarify their comments if necessary. At the end of the interactive session, each participant took part in a short interview/discussion session. Together, the interactive session and interview lasted between 2 and 3 hours, dictated by the participants' chosen strategies.

The videotape of each session was analysed to create a content log describing the activities performed by the participants and précising discussion that took place with the researcher, and to create a transcript of the interview session [11]. The accuracy of each content log was verified by an independent assessor.

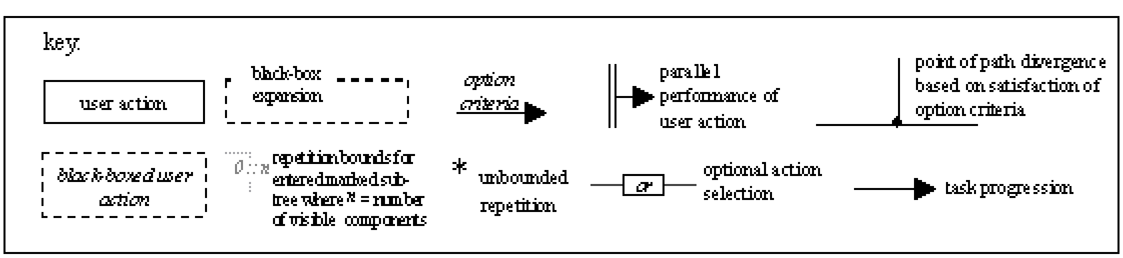
A Model of SUIT Data Exploration/Analysis

The content logs were examined to determine participants' action sequences in order to derive an observational model of the manner in which third-party evaluators used the SUIT VE to explore/analyse data.

The Strategy Model Notation

Initial attempts to model participants' strategies highlighted the complexity of these strategies and the need for a tailored notation to clearly present the activities, their sequence, and flexibility. Participants tended to adopt (often repetitive) patterns of activity; it was therefore important to represent their strategies in such a way as to make these patterns identifiable. There are limited task notations available that could have been used to model the identified strategies [2, 3]. However, their structure did not manageably accommodate the observed complexity of task activity (e.g., task tables became too large for practical use [2, 3]) and they did not allow activity sequencing to be represented so as to make the patterns and flexibility of activity salient [2, 3]. It was, therefore, necessary to devise a suitable notation specifically for the representation of the data analysis strategies (see Figure 2). Notational constructs were required to represent: low-level atomic user actions; black-boxed[2] combinations of actions - and a means to identify the corresponding expansion of the activities internal to the black-box; both bounded and unbounded repetition of collections of one or more activities; choice of action based on satisfaction of given criteria; and parallelism of task performance. In particular, it became necessary to black-box user actions to simplify the diagrams and therefore increase their readability.

Figure 2: Strategy model notation.



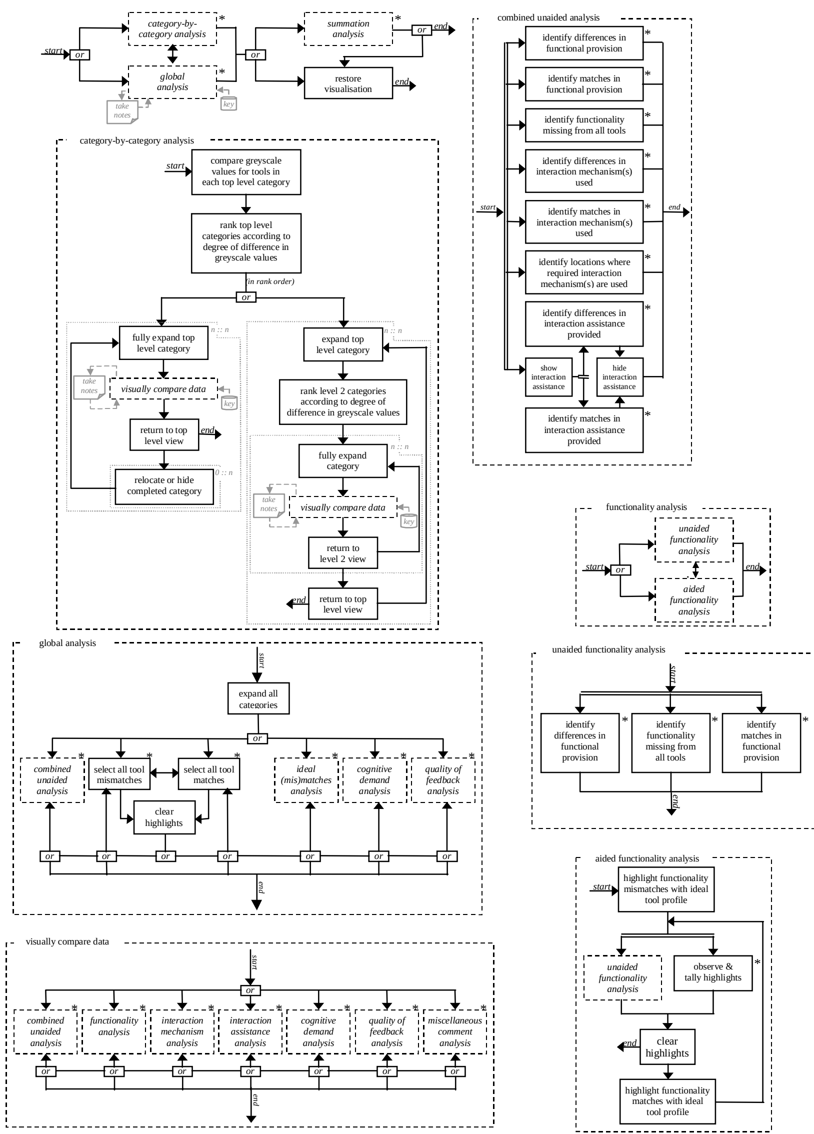
Each participant spent an initial period investigating the data to ‘get a feel for’ the most appropriate strategy; it was normally during this period that any *predetermined* strategies were confirmed/altered. Participants typically sub-divided their analysis into a period of *global* analysis and a period of *category-by-category* analysis. Although participants’ strategies for each type of analysis differed in terms of detail, there were several common points.

During their category-by-category analysis, participants generally ranked their top level categories according to the degree of difference in the greyscale values (functionality provision) across the different tools. Thereafter, some started with categories with the least difference in the hope to be able to eliminate categories from their analysis (i.e. the decision-making process), and others focussed primarily on the categories with the greatest degree of difference considering them to be the categories in which the decision criteria would be found. Irrespective of whether the category ranking was performed, participants selected each category in turn and (usually) expanded the category to view the lowest level of detail. After examining the data at that level, they all returned the visualization to the top-level category view. Little use was made of the intermediate levels of detail in the visualization.

Without exception, the participants focussed their analysis on the differences between tools; given that the tools being considered were functionally and stylistically very similar, this proved an effective strategy. During one interview session, a participant questioned whether this approach would be as effective if the tools being considered were radically different in most respects. Although this would require further investigation, it is likely that evaluators would still adopt the ‘negative stance’ since it is the differences rather than the similarities that set tools apart and enable the selection of one over another. In general, when analysing data at the lowest level of detail, participants relied most heavily on ‘manual’ comparison of the data components – a process commonly referred to by participants as ‘eyeballing’ the data – as opposed to using the provided filter facilities. Even when filters were applied to the data, participants still relied on eyeballing the data for close scrutiny.

During the course of the interactive sessions, it was interesting to note that, had the ideal tool profile not been present to provide the information about the context of use, participants would have intuitively evaluated the tools on the basis of their personal preferences rather than those of the intended user group. This is best illustrated in the following quotation taken from one participant’s transcript (the participant is responding to a question from the researcher as to whether he thought that, in the absence of the ideal tool profile, he felt he might

Figure 3: Selection of initial levels of the strategy model for systematic third-party context-sensitive decision-making using SUIT.



have been inclined to compare the tools on the basis of personal preferences):

yes…definitely, because I think everybody, although they would like to take into account how other people see things and how other people like to work, I think initially – more than initially actually – if you're looking at some project, some problem, and you…you find you approach it from what you know and if what you know is programming language [one mechanism for achieving functionality] rather than some of the other interaction mechanisms [alternative mechanisms for achieving functionality], then that would be your baseline – that would be where you were judging them from…

The ideal profile therefore successfully maintained participants' focus on the requirements of the group for which they had assumed an evaluative role; it served as a visual reminder that their personal preferences were not only orthogonal to those for whom the participants were acting, but that their preferences were essentially irrelevant.

Although most participants compared tool data according to a common prioritization sequence (functionality, then interaction mechanisms, then help facilities), some performed all comparisons at once (i.e. all data types were compared during a single full expansion of the top-level category) whilst others repeated the category-by-category expansion for each of the data dimensions in turn. For those participants who performed global analysis *after* category-by-category analysis, the former appeared to constitute a confirmation exercise; they had, in general, identified the points upon which to base their selection decision but used the global analysis to 'double check' their observations. In some cases, it was at this point that participants compared the data set on the basis of the ancillary data dimensions to determine whether there was any other dimension, over and above functionality and interaction mechanism observations, upon which to base their recommendation. Very rarely did participants combine the effects of filters over their data set; instead, they tended to work methodically through their desired filters, clearing the effects of one before applying the next.

Based on these studies, a strategy map has been developed that, via flexibility of choice, captures the majority of the observed approaches. Due to the comprehensive yet complex nature of the strategy model it is infeasible to include it in its entirety in this paper (for the complete model, see [6]). Instead, some of the initial breakdowns of the model are included to provide an illustration of the strategy model as a whole - see Figure 3.

CONCLUSIONS AND FURTHER WORK

The strategy model presented here is a comprehensive, combined summary of the data exploration/analysis strategies adopted by the different participants. It is a first attempt to model how evaluators structure data exploration/analysis when asked to assume an organizational role to make a context-sensitive decision on behalf of a third-party. The observational studies illustrated the importance of visualizing context-sensitive goals/preferences to correctly direct/maintain evaluator focus when an evaluator is asked to assume such a role. Evidence suggests that the exclusion of such a component in SUIT data visualizations would lead to evaluators abandoning their assumed role and selecting UIDTs on the basis of their individual preferences. This would, in turn, result in poorer quality decisions wherein the selected UIDT would be less likely to be accepted by the intended user group.

Interestingly, all bar one participant recommended the same UIDT; those who agreed on their recommendation adopted an often similar, obviously systematic strategy, whereas the one subject who differed on his recommendation was less obviously systematic in his approach. This implies that, for the data being considered in this research, inclusion of the ideal tool profile within the VE not only focuses the attention of the evaluator to the goals of the assumed role, but it also generally leads to observably systematic analysis strategies and to consistency of decision-making across individual evaluators acting for a third-party. Although it is not the intention of this research to impose data analysis strategies upon evaluators using SUIT, the strategy model can potentially be used as an optional and flexible guide for future evaluators using the SUIT VE; if they choose to follow this model, they may benefit from the strategies attempted by previous evaluators.

At this point in time, we make no claim as to the effectiveness of the strategies modeled other than that which is demonstrated through the participants' responses. We intend to complete longitudinal observational studies of the use of the strategy model in combination with the SUIT VE to select UIDTs for real-world projects and to monitor the effectiveness of the selection decision. A seminal component of this future research will indefinitely be to address the question of *how to measure effectiveness* within this context.

REFERENCES

[1] S. Clarke, "Encouraging the effective use of contextual information in design", Ph.D. Thesis, Department of Computing Science, University of Glasgow, 1997.

[2] P. Gray, D. England, and S. McGowan, "XUAN: Enhancing UAN to Capture Temporal Relationships Among Actions," in People and Computers IX: HCI'94, Glasgow, Scotland, pp. 301 - 312, 1994.

[3] H. R. Harston and P. D. Gray, "Temporal Aspects of Tasks in the User Action Notation," *Human-Computer Interaction*, **7** (1), pp. 1 - 45, 1992.

[4] W. S. Humphrey, *Managing the Software Process*. Reading MA: Addison-Wesley, 1989.

[5] C. F. Kemerer, "How the Learning Curve Affects CASE Tool Adoption," *IEEE Software*, **9** (3), pp. 23 - 28, 1992.

[6] J. Lumsden, "SUIT - A Methodology and Framework for Selection of User Interface Development Tools", Ph.D. Thesis, Department of Computing Science, University of Glasgow, 2001.

[7] J. Lumsden, "Selecting the 'Invisible' User Interface Development Tool," in People and Computers XVI: Human Computer Interaction 2002 (HCI'2002), London, UK, pp. 365 - 380, 2002.

[8] J. Lumsden and P. Gray, "SUIT - Context Sensitive Evaluation of User Interface Development Tools," in Interactive Systems Design, Specification and Verification Revised Papers, 7th International Workshop, DSV-IS 2000, Limerick, Ireland, pp. 91 - 108, 2000.

[9] J. McKirdy, "An Empirical Study of the Relationships Between User Interface Development Tools & User Interface Development," University of Glasgow, Glasgow TR-1998-06, March 1998.

[10] J. McKirdy, "Choosing the UI Tool Which Best Suits Your Needs," in Human Computer Interaction INTERACT'99, Edinburgh, pp. 49 - 50, 1999.

[11] E. J. O'Neill, "User-developer cooperation in software development: building common ground and usable systems", PhD Thesis, Department of Computing Science, Queen Mary and Westfield College University of London, 1998.

[12] M. C. Paulk, B. Curtis, M. B. Chrissis, and C. V. Weber, "The Capability Maturity Model for Software," *IEEE Software*, **10** (4), pp. 18-27, 1993.

[13] V. L. Sauter, "Intuitive Decision-Making," *Communications of the ACM*, **42** (6), pp. 109 - 115, 1999.

[14] B. Shneiderman, "Dynamic Queries for Visual Information Seeking," *IEEE Software*(November), pp. 70 - 77, 1994.

[15] H. Simon, "Research Briefings 1986: Report of the Research Briefing Panel on Decision Making and Problem Solving," National Academy of Sciences, Washington 1986.

[16] M. Spenke, C. Beilken, and T. Berlage, "FOCUS: The Interactive Table for Product Comparison and Selection," in ACM Symposium on User Interface Software Technology (UIST), Seattle Washington, pp. 41- 50, 1996.

ENDNOTES

[1] 'He'/'his' is used irrespective of participant gender to prevent participant identification.

[2] A 'black-box' being a high-level abstraction of, or placeholder for, a complex process, the detail of which is shown in an associated 'black-box expansion'.

Related Content

Research on Irregular Flight Recovery Strategy Under Different Flight Route Types With Big Data Computing
Wei Fan, Yanfei Xu, Liang Lu, Honghai Zhang, Xuecheng Wu, Yan Jiangand Yingfeng Zhang (2024). *International Journal of Information Technologies and Systems Approach (pp. 1-20).*
www.irma-international.org/article/research-on-irregular-flight-recovery-strategy-under-different-flight-route-types-with-big-data-computing/349135

High-Performance Reconfigurable Computing
Mário Pereira Vestias (2018). *Encyclopedia of Information Science and Technology, Fourth Edition (pp. 4018-4029).*
www.irma-international.org/chapter/high-performance-reconfigurable-computing/184109

Forecasting Model of Electricity Sales Market Indicators With Distributed New Energy Access
Tao Yao, Xiaolong Yang, Chenjun Sun, Peng Wuand Shuqian Xue (2023). *International Journal of Information Technologies and Systems Approach (pp. 1-16).*
www.irma-international.org/article/forecasting-model-of-electricity-sales-market-indicators-with-distributed-new-energy-access/326757

Fuzzy Rough Set Based Technique for User Specific Information Retrieval: A Case Study on Wikipedia Data
Nidhika Yadavand Niladri Chatterjee (2018). *International Journal of Rough Sets and Data Analysis (pp. 32-47).*
www.irma-international.org/article/fuzzy-rough-set-based-technique-for-user-specific-information-retrieval/214967

Accumulation and Erosion of User Representations or How is Situated Design Interaction Situated
Sampsa Hyysalo (2012). *Phenomenology, Organizational Politics, and IT Design: The Social Study of Information Systems (pp. 196-220).*
www.irma-international.org/chapter/accumulation-erosion-user-representations-situated/64685