Data Quality Simulator: Uma Proposta de um Jogo Para Apoiar o Aprendizado em Qualidade de Dados

Bárbara F. Marques, Melise M. V. de Paula

Instituto de Matemática e Computação – Universidade Federal de Itajubá (UNIFEI)
Caixa Postal 50 – 37.500-903 – Itajubá – MG – Brazil

{barbarafmarques,melisepaula}@gmail.com

Abstract. *There is a growing concern in the systems context about the quality of information products that are generated. However, the lack of effectively pratical explicit knowledge in this subject can difficult the use of the approaches proposed in the literature. In this paper, a proposal will be presented to facilitate the learning of methodologies applied to management of quality data that was based on the development of a digital game, Data Quality Simulator, to facilitate learning strategies and technical proposals. In addition to the methodology and the game, this paper will discuss the approach to be used for evaluation of the proposal.*

Resumo. *Atualmente, há uma crescente preocupação com a qualidade dos produtos de informação que são gerados. Contudo, a falta de conhecimento explícito que seja efetivamente prático nesta área pode dificultar a utilização das abordagens propostas na literatura. Neste trabalho, será apresentada uma proposta para facilitar o aprendizado em metodologias para a gestão da qualidade de dados que foi fundamentada no desenvolvimento de um jogo digital, Data Quality Simulator, de maneira a facilitar o aprendizado das estratégias e técnicas propostas. Além da metodologia e do jogo, será discutida a abordagem a ser utilizada para avaliação da proposta.*

1. Introdução

Os sistemas que manipulam bases de dados são responsáveis por gerar o chamado produto da informação que, quando incorretos ou inconsistentes, distorcem análises de resultados podendo causar prejuízos e comprometer a corretude das ações. A qualidade deste tipo de produto pode ser comprometida resultando em retrabalho e perda de recursos. Já em 1999, Ballou et al. (1999) afirmavam que alguns sistemas de suporte a tomada de decisão não utilizam todo o conteúdo de informação disponível por não terem garantia da validade e adequação destes dados. Porém, a qualidade de dados não é apenas uma questão tecnológica, mas também de gerenciamento [Batini e Scannapieca 2006]. A definição de metodologias ou estratégias pode auxiliar os usuários a identificar problemas e gerenciar a qualidade dos dados com menos custos e de forma mais rápida [Barros 2009].

Para a implantação de uma metodologia é necessário que as pessoas envolvidas tenham um bom entendimento de como deve ser sua aplicação. Na literatura, é possível encontrar diferentes propostas de metodologia que, em sua maioria, apresentam muitas

DOI: 10.5753/cbie.sbie.2016.1326

discussões subjetivas e com roteiros pouco práticos com nenhuma ou apenas algumas ferramentas de suporte para que as organizações consigam se basear [Batini et al 2009].

A motivação para este trabalho foi a dificuldade em encontrar metodologias de análise e gestão de qualidade de dados que pudessem ser consideradas de forma mais prática durante a execução de um projeto real. O primeiro passo foi estabelecer este fluxo de trabalho para gestão da qualidade de dados, que foi esquematizado em etapas com ferramentas e técnicas práticas para sua execução. As atividades descritas neste fluxo de trabalho foram definidas com base, principalmente, na metodologia TDQM (Total Data Quality Management) desenvolvida pelo MIT [Wang 1998].

Contudo, o aprendizado deste tipo de conteúdo requer prática. Como a prática através de casos reais nem sempre é possível, a proposta foi facilitar a absorção de conhecimento através de simulações, mais especificamente, um jogo. O uso de jogos e técnicas de gamificação para facilitar o aprendizado vem sendo utilizado em diferentes cenários [Rocha et al. 2015] [Buchinguer e Hounsell 2015]. Desta forma, o objetivo é aumentar as possibilidades de criação do conhecimento nesta área considerando tanto cenários acadêmicos quanto profissionais.

O artigo está estruturado da seguinte maneira: a seção 2 descreve o fluxo da metodologia para gestão de qualidade de dados considerada e suas etapas; a seção 3 apresenta o jogo em desenvolvimento e o método de validação que será realizado e, por fim, a seção 4 expõe as considerações finais.

2. Metodologia

A metodologia descrita para executar um projeto de qualidade é composta por cinco etapas que não devem ser entendidas de forma estritamente sequencial: preparação, definição, avaliação, análise e ação. A cada iteração de um ciclo de execução do projeto de qualidade, o foco pode variar entre as etapas. O ciclo, apresentado na figura 1, deve ser executado até que a qualidade identificada nos produtos de informação esteja compatível com as expectativas dos atores envolvidos.

Figura 1: Sequência das etapas da metodologia

O objetivo da etapa de Preparação é definir os aspectos gerenciais do projeto de qualidade. Dentre os artefatos a serem elaborados pode-se citar: a estrutura analítica e de gerenciamento do projeto, o cronograma de entrega das atividades e um plano de gerenciamento de riscos. Em sequência, a etapa de Definição tem como objetivo especificar quais características dos produtos de informação serão analisadas. A execução desta etapa é dada pelas atividades: obter informação relevante e identificar as fontes de conhecimento; definir temas estratégicos e estabelecer critérios para priorização; analisar o ciclo de vida dos produtos de informação utilizando a Matriz do POSMAD [McGilvray 2008]; identificar as dimensões relevantes, especificando as

características que deverão ser consideradas para examinar quantitativamente a qualidade dos dados [Batini e Scannapieca 2006]; e, definir métricas que deverão estabelecer os valores esperados, aceitáveis e as condições de erro relacionadas a cada dimensão [Batini et al, 2009]. Na etapa de Avaliação ocorre a execução de todas as métricas definidas e é seguida pela etapa de Análise. Nesta etapa as atividades envolvidas consistem em identificar as dimensões que não atenderam o padrão estabelecido, relacionar as informações coletadas buscando as causas e os efeitos dos problemas e, por fim, definir ações factíveis que neutralizem as causas encontradas. A próxima etapa, denominada Ação, inclui a realização das ações de melhorias estabelecidas na etapa anterior. Com as ações em prática é necessário retornar ao ciclo da metodologia para que uma nova avaliação possa validar se as ações tomadas foram de fato efetivas para o aprimoramento da qualidade dos dados.

3. Data Quality Simulator

O jogo está sendo desenvolvido utilizando a ferramenta RPGMaker MV, disponível pela Degica (2016). Devido à extensão de tópicos e definições que a metodologia abrange, o jogo mantém o foco nos aspectos técnicos da metodologia; logo, a etapa de preparação, por considerar questões gerenciais, não será considerada. Esta primeira versão do jogo aborda as seguintes atividades: estudo do ciclo de vida com o uso do POSMAD, identificação das dimensões e definição das métricas.

Inicialmente o jogador é estimulado a explorar um ambiente virtual que simula um ambiente de trabalho de uma organização atuante no setor da informação. No cenário, além do jogador, são representados NPC (Non-Player Character) e objetivos característicos. O jogador começa com uma quantia em uma moeda fictícia. O primeiro objetivo é preencher a matriz do POSMAD de maneira que seja possível identificar um problema de qualidade. Depois de identificado o problema, o jogador deverá definir as dimensões e métricas apropriadas. A cada problema, é usado um valor de moeda. A quantia gasta depende da quantidade de acertos/erros cometidos pelo jogador.

O jogador pode interagir com o NPC quando quer obter informações sobre o problema. Para cada interação, são disponibilizadas opções de diálogos e o jogador deve selecionar a mais adequada em relação ao aspecto do POSMAD que está sendo analisado. Interagir com objetos do cenário, como computadores e livros, retorna conhecimentos sobre a metodologia que auxiliam na escolha das opções de diálogo. Caso seja selecionada a opção correta, a matriz do POSMAD é preenchida de forma automática conforme ilustra a Figura 2, enquanto opções incorretas impedem que o jogador obtenha a informação deixando lacunas na matriz. Cada nova lacuna preenchida reduz o valor de custos necessários para solucionar o problema a qual ela se refere.

Figura 2: Matriz do POSMAD no canto superior esquerdo da tela

Para cada fase do POSMAD, referentes ao eixo superior da matriz, é possível encontrar informações relacionadas aos componentes chaves que se referem ao eixo lateral. Os problemas são detectados somente quando o jogador identifica as informações mínimas que caracterizam o problema. Neste momento do jogo, o objeto que representa o problema aparece no ambiente virtual permitindo interação (figura 3).

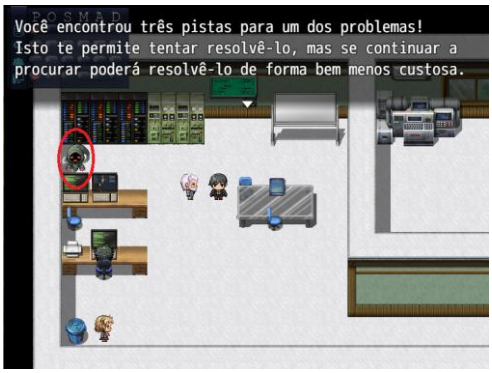

Figura 3: Objeto que representa o problema

Ao clicar na figura que representa um problema, o jogador deve selecionar, dentre as dimensões e métricas apresentadas, aquelas que são mais apropriadas. Quando são selecionadas as dimensões e métricas corretas, o problema é solucionado. O jogo encerra quando o jogador encontra todos os problemas ou o dinheiro fictício acaba, a relação de problemas solucionados e a quantia final de moeda representa a pontuação do jogador. Este fluxo de ações é observado na figura 4.

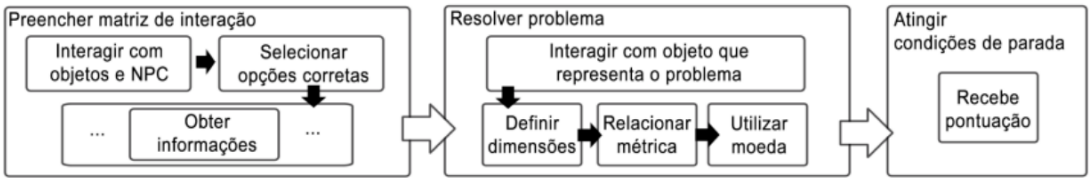

Figura 4: Fluxo de ações do jogador

4. Considerações finais

Neste trabalho, foi descrita uma proposta de um jogo cujo objetivo é facilitar o aprendizado dos conceitos abordados em metodologias para gestão de qualidade de dados. As atividades e etapas do jogo foram descritas.

Um dos desafios da aplicação de jogos considerados jogos sérios, que não são de intuito apenas de entretenimento, consiste na validação do jogo que pode apresentar diferentes propósitos. Logo, é necessário ter definido “o que” e “como” o jogo será validado para garantir a eficácia de seu uso [Rocha et al 2015]. Nesse trabalho, serão avaliados três aspectos: o modelo de simulação, o software (ou a interface) e o jogo como instrumento de aprendizagem. Especificamente para avaliação da aprendizagem será utilizada uma abordagem semelhante a descrita por Buchinger e Hounsell (2015). Será definido um grupo de controle e um grupo de testes que serão formados por alunos de cursos de computação de uma universidade federal. Para os dois grupos, será dada pelo instrutor em sala de aula uma explanação sobre os conceitos de qualidade de dados e a metodologia descrita. Inicialmente, os dois grupos serão avaliados em relação ao

conhecimento sobre o conteúdo abordado. Contudo, somente ao grupo de teste será dado o acesso ao jogo, sendo esse acesso realizado uma vez presencialmente de forma assistida pelo instrutor, que não deverá auxiliar em relação ao conteúdo ou usabilidade, e disponível também durante duas semanas para acesso fora da sala de aula. Apenas após o período estipulado de utilização os alunos dos dois grupos serão avaliados novamente de maneira que seja possível analisar o impacto do uso do jogo.

Os demais aspectos serão avaliados somente com o grupo de teste através da aplicação de questionários para que sejam coletadas as opiniões dos alunos quanto a facilidade de compreensão do conteúdo apresentado pela ferramenta, a escala de conhecimento acrescentada a respeito dos conceitos da metodologia durante a utilização da ferramenta, a relação da utilização da ferramenta com o nível de absorção dos conceitos e a colaboração da ferramenta quanto a motivação dos alunos a buscarem aprender sobre o conteúdo; seguido de um espaço para eventuais comentários dos alunos.

References

Ballou, D., Chengalur-Smith, I., E Pazer, H. (1999). "The Impact of data quality information on decision making: An exploratory analisys". In IEEE transactions on knowledge and data engineering (volume:11, issue:6) p.853-864.

Barros, R. (2009). "Qualidade de informação na web: um prognóstico fuzzy baseado em metadados". Tese (doutorado) – UFRJ/COPPE/Programa de engenharia de sistemas e computação.

Batini, C., Cappiello, C., Francalanci, C. E Maurino, A. (2009). "Methodologies for Data Quality Assessment and Improvement". ACM Computing Surveys, Vol. 41, No. 3, Article 16.

Batini, C. E Scannapieco, M. (2006). "Data Quality: concepts, methodologies and techniques". In Springer.

Buchinger, D., Hounsell, M. (2015). "O aprendizado através de um jogo colaborativocompetitivo contra Dengue". In Anais do XXVI Simpósio Brasileiro de Informática na Educação (SBIE).

Degica Company Limite (2016). RPGMaker Web.

Lee, Y., Pipino, L., Funk J., E Wang, R. (2006). Journey to Data Quality. Cambridge, MA: MIT Press.

McGilvray, D. (2008). Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information™, 1st Edition.

Rocha, R., Bittencourt, I. E Isotani, S. (2015) "Análise, projeto, desenvolvimento e avaliação de jogos sérios e afins: uma revisão de desafios e oportunidades". In Anais do XXVI Simpósio Brasileiro de Informática na Educação (SBIE).

Wang, R. (1998). "A product perspective on total data quality management". In Magazine Communications of the ACM. Volume 41, Issue 2, p.58-65.

Wang, R. E Strong, D. (1996). "Beyond accuracy: What Data quality means to data consumers". In: Journal of Management Information System, Vol.12, N.4, p.5-33.