



APLICAÇÃO DE TÉCNICAS DE QUALIDADE DE DADOS E GAMIFICAÇÃO EM PROJETOS DE CROWD SCIENCE

Diego de Almeida Zanon

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia de Sistemas e Computação, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia de Sistemas e Computação.

Orientadores: Jano Moreira de Souza

Maria Gilda Pimentel Esteves

Rio de Janeiro

Dezembro de 2018

APLICAÇÃO DE TÉCNICAS DE QUALIDADE DE DADOS E GAMIFICAÇÃO EM
PROJETOS DE CROWD SCIENCE

Diego de Almeida Zanon

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO
LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA
(COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE
DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE
EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

Examinada por:

Prof. Geraldo Zimbrão da Silva, D. Sc.

Dr.^a Maria Gilda Pimentel Esteves, D. Sc.

Prof.^a Adriana Santarosa Vivacqua, D. Sc.

RIO DE JANEIRO, RJ - BRASIL

DEZEMBRO DE 2018

Zanon, Diego de Almeida

Aplicação de Técnicas de Qualidade de Dados e Gamificação em Projetos de Crowd Science / Diego de Almeida Zanon – Rio de Janeiro: UFRJ/COPPE, 2018.

XI, 84 p.: il.; 29,7 cm.

Orientadores: Jano Moreira de Souza

Maria Gilda Pimentel Esteves

Dissertação (Mestrado) – UFRJ/ COPPE/ Programa de Engenharia de Sistemas e Computação, 2018.

Referências Bibliográficas: p. 75-78.

1. Crowd Science. 2. Gamificação. 3. Qualidade de dados. I. Souza, Jano Moreira et al. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Sistemas e Computação. III. Título.

Dedicatória

À minha esposa Carol
Aos meus pais Reinaldo (*in memoriam*) e Lourdes

Agradecimentos

Aproveito esse espaço para agradecer de forma geral a todos que, direta ou indiretamente, contribuíram para que eu conseguisse concluir a minha formação de mestre.

À minha querida esposa Carol que sempre esteve do meu lado nos momentos mais difíceis e por fazer a minha vida melhor.

Ao meu pai, em memória, e à minha mãe por sempre me motivarem a estudar e a seguir este caminho.

Agradeço à Gilda Esteves por ter me apresentado ao tema deste projeto e por toda a orientação, motivação e parceria nestes anos.

Ao professor Jano Moreira de Souza pela sua orientação, ensinamentos e atenção dada nas nossas longas reuniões e discussões.

Ao professor Eduardo Barbosa pelas conversas e o incentivo.

Aos professores Marcio Antelio e Geraldo Xexéo e à aluna de mestrado Airine Carmo pelo apoio na realização do experimento deste trabalho.

Aos alunos do CEFET de Itaguaí que participaram do experimento e me ajudaram a levantar métricas para defender esta dissertação.

Aos professores participantes da banca pelo seu tempo, atenção e críticas.

Aos colegas Rafael Quintanilha, Daniel Rechtman, Thiago Ferraz e Luiz Oliveira pelas contribuições técnicas e apoio no desenvolvimento do código e infraestrutura da plataforma.

Ao meu amigo e ex-chefe, Bruno Tonioni, e a minha antiga empresa Chemtech, por apoiar o meu desejo de realizar o mestrado e me dar flexibilidade nos horários de trabalho para que eu pudesse assistir às aulas e desenvolver o meu projeto de pesquisa.

Agradeço também ao IBGE, minha empresa atual, pelo apoio e incentivo.

E à sociedade brasileira por ajudar a custear o ensino superior de qualidade e gratuito com o objetivo de formar cidadãos mais capacitados para o desenvolvimento do nosso país.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

APLICAÇÃO DE TÉCNICAS DE QUALIDADE DE DADOS E GAMIFICAÇÃO EM PROJETOS DE CROWD SCIENCE

Diego de Almeida Zanon

Dezembro/2018

Orientadores: Jano Moreira de Souza

Maria Gilda Pimentel Esteves

Programa: Engenharia de Sistemas e Computação

Crowd Science é um novo paradigma para colaboração que possui o objetivo de viabilizar, através da participação aberta de um número grande de cidadãos não cientistas, pesquisas científicas que dependem de trabalho humano e que não seriam possíveis por restrições econômicas ou de tempo. A partir do crescente interesse acadêmico nesta área, este trabalho optou por pesquisar como ferramentas de qualidade de dados podem auxiliar na *eficácia* de iniciativas de crowd science, tornando os resultados mais confiáveis e assertivos, e como ferramentas de gamificação podem auxiliar na *eficiência* dos projetos, aumentando a retenção de voluntários e a quantidade de contribuições feitas por cada um. Após a discussão destes dois temas, que incluiu uma investigação sobre as ferramentas disponíveis em gamificação e qualidade dos dados e como aplicá-las no campo de pesquisa de crowd science, esta dissertação descreve como foi conduzida a seleção e a implementação de novas ferramentas na plataforma Fast Science, e como foi realizado um teste A/B que indicou que a gamificação implementada aumentou em 4 vezes o número de contribuições sem que a qualidade de dados fosse prejudicada.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

APPLYING DATA QUALITY AND GAMIFICATION TECHNIQUES
IN CROWD SCIENCE PROJECTS

Diego de Almeida Zanon

December/2018

Advisors: Jano Moreira de Souza

Maria Gilda Pimentel Esteves

Department: Computer and Systems Engineering

Crowd Science is a new paradigm for collaboration that has the goal of enabling, through open participation of a great number of non-scientists citizens, scientific researches that depends on human workforce and that would not be feasible considering economic or time constraints. Considering the growing academic interest in this area, this work has researched how some data quality tools can help in the *effectiveness* of crowd science initiatives, making the results more reliable and assertive, and how gamification tools can help in the *efficiency* of those projects, increasing the retention of volunteers and the amount of contributions made by each of them. After discussing these two subjects, which includes a research of the gamification and data quality tools available and how to apply them in the crowd science field, this work demonstrates how the selection and development of the new tools was done in the Fast Science platform, and how an A/B test was performed which indicated that the gamification used has increased by 4 times the number of contributions without harming the data quality.

Sumário

Capítulo 1 – Introdução	1
1.1 <i>Definição de Crowd Science</i>	1
1.2 <i>Benefícios de Crowd Science</i>	3
1.3 <i>Aumentando as chances de sucesso de um projeto de Crowd Science</i>	5
1.4 <i>Motivação para uso da Plataforma Fast Science</i>	9
1.5 <i>Questões de Pesquisa</i>	11
1.6 <i>Metodologia</i>	11
Capítulo 2 – Revisão da Literatura	15
2.1 <i>Qualidade de Dados</i>	15
2.1.1 <i>Definição de Qualidade de Dados</i>	15
2.1.2 <i>Qualidade de Dados em Crowd Science</i>	18
2.1.3 <i>Melhorando a Qualidade e a Confiabilidade dos Dados</i>	19
2.1.4 <i>Definindo a Qualidade Através da Multidão</i>	21
2.2 <i>Gamificação</i>	23
2.2.1 <i>Definição de Gamificação</i>	23
2.2.2 <i>Gamificação em Crowd Science</i>	24
2.2.3 <i>Aplicando Gamificação</i>	27
2.2.4 <i>Dando recompensas</i>	31
2.2.5 <i>Ranqueamento dos jogadores</i>	31
Capítulo 3 – Proposta e Implementação	33
3.1 <i>Histórico e estado inicial do Fast Science antes deste trabalho</i>	33
3.2 <i>Qualidade de Dados</i>	36
3.2.1 <i>Sistema de validação de contribuições</i>	36
3.2.2 <i>Sistema de auditoria</i>	39
3.2.3 <i>Sinalizadores de problemas</i>	40
3.2.4 <i>Notificações</i>	41
3.2.5 <i>Melhorias na execução de tarefas</i>	41
3.2.6 <i>Detector de texto sem sentido</i>	42
3.2.7 <i>Estatísticas de contribuições</i>	45
3.3 <i>Gamificação</i>	47
3.3.1 <i>Ranqueamento</i>	47
3.3.2 <i>Recompensas</i>	48
3.3.3 <i>Autopromoção</i>	50

3.3.4 Barra de progresso	50
3.4 Implementação e Arquitetura.....	51
3.5 Modelo de dados.....	53
Capítulo 4 – Experimento.....	55
4.1 Objetivos.....	55
4.2 Metodologia.....	55
4.2.1 Planejamento.....	55
4.2.2 Preparação.....	56
4.2.3 Recrutamento e execução	57
4.2.4 Perfil dos Participantes.....	58
4.2.5 Avaliação	59
4.3 Resultados para Gamificação.....	60
4.3.1 Quantidade de tarefas executadas	60
4.3.2 Formulário de avaliação	60
4.4 Resultados para Qualidade de Dados.....	64
4.4.1 Formulário de avaliação.....	64
4.4.2 Quantidade de contribuições inválidas.....	66
4.4.3 Tarefas de auditoria	68
Capítulo 5 – Conclusão	69
5.1 Questões de Pesquisa	69
5.2 Avaliação das melhorias para Qualidade.....	70
5.3 Avaliação das melhorias para Motivação.....	72
5.4 Trabalhos Futuros	73
5.4.1 Lei de Benford.....	73
5.4.2 Ferramentas de comunicação e divulgação	74
Referências Bibliográficas.....	75
Anexo 1 – Formulário de Avaliação para o grupo com gamificação.....	79
Anexo 2 – Formulário de Avaliação para o grupo sem gamificação	82

Índice de Figuras

Figura 1 – Fluxograma de processos adaptado de PEFFERS (2007).....	12
Figura 2 – Gráfico de média móvel (BORNFELD e RAFAELI, 2017).....	28
Figura 3 – Pagina inicial da Plataforma Fast Science.....	34
Figura 4 – Lista de projetos disponíveis por tipo (coleta ou processamento)	34
Figura 5 – Workflow de contribuição no projeto “Jogos e Emoções”	35
Figura 6 – Tela de validação.....	37
Figura 7 – Após analisar a contribuição, o pesquisador aceita ou rejeita.....	38
Figura 8 – Funcionalidade de auditoria	39
Figura 9 – Sinalizando problemas durante uma contribuição.....	40
Figura 10 – Mensagem de alerta que o gestor do projeto recebe	41
Figura 11 – Tela para controle de alertas criados e as suas respostas	41
Figura 12 – Alerta de mensagem com texto incompreensível.....	43
Figura 13 – Contribuições no tempo	46
Figura 14 – Contribuições por voluntário.....	46
Figura 15 – Ranking implementado na plataforma Fast Science	48
Figura 16 – Funcionalidade de notificações de recompensas.....	49
Figura 17 – Plano de fundo e insígnias temáticas.....	49
Figura 18 – Indicador mostra progresso para a terceira medalha.....	50
Figura 19 – Visão das camadas da arquitetura e os seus módulos	51
Figura 20 – Diagrama de componentes para os módulos do AngularJS	52
Figura 21 – Diagrama de classes da implementação do React.....	53
Figura 22 – Modelo simplificado do banco de dados.....	54
Figura 23 – Distribuição de voluntários por gênero	58
Figura 24 – Distribuição de voluntários por idade	58
Figura 25 – Comparativo de execução de tarefas	60
Figura 26 – Resultados do formulário com gamificação.....	61
Figura 27 – Resultados do formulário sem gamificação	62
Figura 28 – Diagrama de caixa do resultado	65

Índice de Tabelas

Tabela 1 – Comparativo entre iniciativas de crowd science.....	10
Tabela 2 – Análise de qualidade de dados.....	16
Tabela 3 – Lista das afirmativas elaboradas para avaliar as funcionalidades.....	59
Tabela 4 – Resultados dos formulários de feedback	61
Tabela 5 – Análise de contribuições válidas no experimento.....	68

Capítulo 1 – Introdução

Esta dissertação é uma continuação do trabalho iniciado pela Dra. Maria Gilda Pimentel Esteves em sua tese de doutorado (ESTEVES, 2016), onde o sistema Fast Science (<https://fastscience.com.br>) foi desenvolvido com o objetivo de fomentar e viabilizar tecnicamente a concepção e execução de projetos científicos com a participação de multidões - *Crowd Science*. Sendo um dos programadores que colaborou para o desenvolvimento inicial do sistema Fast Science, o autor desta dissertação buscou dar continuidade a pesquisa investigando e propondo soluções tecnológicas direcionadas para dois grandes desafios descritos na literatura: (1) aumentar a confiabilidade e a qualidade dos dados fornecidos pelos participantes; (2) promover mecanismos que estimulem o engajamento e a manutenção de uma comunidade motivada a participar de projetos de crowd science.

Após esta pesquisa inicial, foi possível implementar ferramentas de gamificação e qualidade de dados para adicionar ao Fast Science mecanismos para uma melhor retenção de voluntários e assertividade no resultado final dos projetos. O objetivo dessa dissertação, então, é mostrar a investigação realizada sobre as ferramentas disponíveis em gamificação e qualidade de dados e como aplicá-las no campo de pesquisa de crowd science e quais foram os resultados obtidos com as mesmas.

Assim sendo, este trabalho irá iniciar com definição de crowd science, seguindo com a apresentação da motivação do trabalho, explicando o por quê e o interesse de se construir melhores sistemas para crowd science, prosseguindo com a apresentação da revisão da bibliografia e a análise de ferramentas e técnicas para qualidade de dados e gamificação. Na sequência será apresentado um histórico do projeto Fast Science e como as ferramentas citadas neste estudo foram implementadas. Por fim, será apresentado um experimento de análise e *feedback* da implementação e uma conclusão apresentando o resultado que foi obtido e trabalhos futuros.

1.1 Definição de Crowd Science

Segundo ESTEVES (2016), o termo *Crowd Science* foi utilizado pela primeira vez na literatura por YOUNG (2010), o qual destacou que a análise de dados na astronomia não era mais realizada apenas por pesquisadores, mas também com a participação de uma multidão de usuários conectados à internet. Através do uso de

plataforma de crowdsourcing e de interfaces web amigáveis, amadores passaram a ser convidados a visualizar dados e rotular imagens de galáxias e já havia relatos de contribuições importantes, como a de uma professora de colégio na Holanda que descobriu uma nuvem de gás incomum, a qual poderia explicar o ciclo de vida dos quasares. YOUNG (2010) definiu que este novo paradigma de colaboração, que permite a inclusão de participantes amadores em projetos científicos, deveria ser chamado de *Crowd Science* e não estava restrito apenas a astronomia, mas a qualquer área da ciência. Este mesmo autor afirmou que crowdsourcing é uma solução natural para muitos dos problemas que os cientistas estão lidando, especialmente aqueles que envolvem a análise ou processamento de uma grande quantidade de dados.

FRANZONI et al. (2014) descreveram quais são as características essenciais para que um projeto seja considerado de crowd science: a participação é aberta a um grande conjunto de potenciais colaboradores e os resultados e dados intermediários são compartilhados com todos. De modo geral, todos os projetos deveriam ter no mínimo estas características. Na prática, alguns não conseguem ter um número grande de voluntários e muitos não disponibilizam os seus resultados intermediários com a comunidade.

Obviamente, para um projeto ser caracterizado como de crowd science, além de seguir os requisitos de aberto a voluntários e com dados compartilhados, é necessário que o mesmo tenha algum objetivo científico. Afinal, sem ciência, temos apenas um projeto de crowdsourcing genérico com apoio de uma multidão.

Segundo ESTEVES (2016), crowd science refere-se ao fenômeno dos projetos inovadores de pesquisa científica que usam crowdsourcing como mecanismo para atrair um número grande de voluntários ampliando-se, desta forma, a equipe do projeto para além dos limites das instituições acadêmicas. Através de uma chamada aberta, pessoas comuns, não cientistas, podem participar de projetos científicos reais e colaborar com a ciência executando tarefas anteriormente realizadas apenas por cientistas.

Outro termo que se confunde com crowd science é Citizen Science. Na literatura não há uma distinção clara entre os dois e ambos são, frequentemente, usados como sinônimos. A ciência cidadã (*citizen science*) convida amadores a participarem do desenvolvimento científico da mesma maneira que a ciência feita pelas multidões - crowd science. A principal diferença é que os projetos de crowd science realizam necessariamente uma chamada aberta e utilizam uma plataforma intermediária de crowdsourcing para facilitar o processo de coleta e processamento de dados. Por esta

razão, em alguns projetos de crowd science de coleta de dados, a plataforma pode transformar o cidadão participante no próprio alvo de pesquisa como, por exemplo, solicitando que o mesmo forneça dados pessoais ao responder questões sobre a sua saúde. Segundo MILACH (2017), na ciência cidadã o papel do participante é diferente de se voluntariar para participar fornecendo dados pessoais. Na ciência cidadã, os cidadãos recolhem e analisam dados sobre o mundo que eles observam em torno deles. Por conseguinte, o “cidadão cientista” desempenha parte do papel do pesquisador, somando uma força de trabalho extra, mas nunca como objeto da pesquisa. Além disso, é comum ver projetos de ciência cidadã que não necessitam de ferramentas digitais, como exemplo, projetos escolares onde as crianças são orientadas para ajudarem a coletarem informações em campo.

1.2 Benefícios de Crowd Science

Segundo COOK (2011): “...uma razão para o repentino olhar para crowd science é que ela oferece uma resposta criativa para um problema central do século XXI: muita informação”. Crowd science representa uma nova abordagem para a condução dos estudos científicos modernos, com especial interesse para as etapas de **obtenção de dados** e a **análise** dos mesmos.

Atualmente, muitas pesquisas modernas necessitam trabalhar com quantidades de dados que eram inimagináveis décadas atrás. *Big Data* é um termo que tem sido utilizado para caracterizar projetos onde a quantidade de dados é exageradamente grande. Porém, este termo não prevê e não define um valor absoluto para *grande*. Por exemplo, um terabyte pode ser muito para um sistema que processa textos de redes sociais, mas é uma quantidade pequena para um sistema que processa vídeos de alta qualidade. Portanto, o tamanho do *Big Data* é definido (WARD et al., 2013) como o tamanho que, para o seu sistema, ferramentas tradicionais deixam de fazer sentido ou que inviabilizem o seu uso.

Crowd science não está necessariamente relacionado a Big Data, mas crowd science faz sentido quando, para o projeto de ciência, a forma tradicional de obtenção e processamento de dados deixa de fazer sentido ou passa a ser inviável. Crowd science é usado quando a dimensão dos dados não permite que o cientista possa coletar ou processar os mesmos apenas com a força de trabalho de sua equipe contratada, onde os recursos e meios comuns passam a ser insuficientes para a realização da pesquisa científica.

Em alguns casos, apesar de não ser viável trabalhar com os dados manualmente, é possível escrever programas de computador para automatizar tarefas, reconhecer padrões, e assim, obter e processar grandes quantidades de dados de forma rápida e prática. Isso é possível e muito utilizado, por exemplo, com a criação de *bots* para fazerem *web scrapping* de redes sociais com o objetivo de obtenção de dados ou a criação de programas de inteligência artificial que extraem e processam informações a partir de imagens. O problema é que nem sempre a pesquisa científica pode reaproveitar ferramentas que já foram criadas para outros fins ou não é viável economicamente ou não é possível desenvolver uma nova ferramenta para o fim específico que se necessita dentro de um prazo aceitável. Além disso, nem tudo pode ser resolvido apenas com o uso de computadores. Em alguns casos, os dados estão disponíveis fisicamente (na natureza, por exemplo) e é necessário o trabalho de uma pessoa para transcrevê-los para um formato digital.

Em 2005, VON AHN cunhou o termo “Computação Humana” para discutir como aproveitar melhor a interação homem-máquina para resolver problemas que necessitam dessa combinação ou que são mais otimizados quando resolvidos desta forma. Na computação tradicional, o homem formata um problema com uma descrição formal e o computador executa uma rotina para resolvê-lo, mas na computação humana, os papéis são invertidos, onde o computador delega aos humanos para que estes executem determinadas tarefas que o mesmo não pode realizar. No final, após o processamento por humanos, o computador agrega, interpreta e integra o resultado à resolução final do problema.

Em 2006, o termo “crowdsourcing” foi cunhado pelo jornalista Jeff Howe e definido como “o ato de terceirizar, através de uma chamada aberta, um trabalho realizado por funcionários de uma empresa para um grupo grande e indefinido de pessoas, utilizando a Internet” (HOWE, 2006). Alinhando os conceitos de computação humana com crowdsourcing para finalidades científicas, é possível usar esse novo paradigma de colaboração para alcançar resultados que antigamente não seriam possíveis. Para isso, voluntários, geralmente amadores com interesse no tema de pesquisa, podem ser convocados para doarem um pouco do seu tempo para ajudarem na pesquisa científica através da realização de tarefas simples, como tirar fotos (coleta) ou catalogar dados (processamento).

Convidar a comunidade a colaborar é também uma maneira de promover os objetivos de uma pesquisa. Por exemplo, se um sistema é criado para que as pessoas

contribuam com a identificação de lugares onde existem pessoas que contraíram determinada doença para mapeamento do contágio e proliferação, as pessoas envolvidas como voluntárias da pesquisa acabam tendo mais informações sobre o problema e ajudam a divulgar as formas de combate ao transmissor da doença para evitar a sua proliferação. Como é o caso do projeto Flu Survey (<https://flusurvey.net>) que foi desenvolvido em 2009 durante a epidemia de gripe suína com o objetivo de mapear as pessoas infectadas na Europa e acompanhar as evoluções e estágios da doença com o preenchimento voluntário de questionários por parte das pessoas doentes.

Em contrapartida, projetos de crowd science também trazem benefícios para os participantes. Além do aprofundamento do conhecimento no tema de interesse, os participantes, em alguns casos, conseguem ter mais exposição e realizar uma autopromoção nas comunidades que se formam em torno de cada projeto. Essas comunidades permitem a criação de contatos e relacionamentos sociais e profissionais entre pessoas com interesses semelhantes, graças à criação do projeto de pesquisa participativa. Entre os principais benefícios identificados na literatura podemos citar: aprendizagem, interação social, reconhecimento, desafio intelectual ou simplesmente diversão e lazer (FRANZONI & SAUERMAN, 2014).

1.3 Aumentando as chances de sucesso de um projeto de Crowd Science

Uma das questões recorrentes na literatura sobre crowd science é como conseguir obter sucesso e atrair dezenas ou centenas de voluntários quando o trabalho não é remunerado e usando, para isso, apenas a motivação pessoal dos participantes.

A Wikipédia não é um exemplo de crowd science, pois a maioria dos artigos elaborados não possuem cunho científico e os que possuem são criações espontâneas da comunidade, sem relacionamento com uma necessidade específica de um pesquisador. Porém, a Wikipédia é um dos melhores exemplos de crowdsourcing onde o trabalho de uma multidão de usuários não remunerados pode sim ser bem sucedido e ter alta qualidade. A questão é por que ela funcionou tão bem quando outras muitas tentativas similares falharam?

De acordo com Larry Sanger (SANGER, 2015), cofundador da Wikipédia, o sucesso dela pode ser atribuído a fatores como os descritos abaixo:

1. **Início com um grupo de pessoas bem intencionadas:** KELLING (1982) elaborou a “Teoria das Janelas Quebradas” que explica que, quando um edifício possui

janelas quebradas e estas não são consertadas, isso incita vândalos a quebrarem outras janelas e danificarem outras partes do edifício. Da mesma forma, quando o edifício está inteiramente bem conservado, a tendência de ele ser atacado por vândalos é muito menor. O que extraímos desta teoria é que, ao iniciar um projeto colaborativo com um grupo pequeno de pessoas focadas em criar conteúdo de alta qualidade, o resultado tende a ser continuado por outras pessoas que ingressam posteriormente no projeto. Da mesma forma, se um projeto é mal iniciado, as características ruins do mesmo tendem a ser replicadas durante a sua evolução.

2. **Conteúdo livre e gratuito:** desde o início, a Wikipédia sempre promoveu a geração de conteúdo livre e gratuito, o que motivou as pessoas a colaborarem de forma altruísta, pelo bem da humanidade. Se o conteúdo fosse vendido e com a intenção de enriquecer um grupo de pessoas, a colaboração não teria sido tão bem vista.
3. **Todos podem colaborar:** a Wikipédia sempre foi aberta a todos, sem elitismo ou discriminação. Isso contribuiu para que, pessoas de diferentes culturas, idades e experiências pudessem contribuir nos mais variados tópicos, permitindo expandir o registro do conhecimento humano. Além disso, as regras sempre estabeleceram que o conteúdo precisa ser *neutro*, não permitindo opiniões ou que a informação tivesse um viés para satisfazer determinados grupos.
4. **Tecnologia facilitando o uso:** sem um site bem feito e com boa usabilidade, a Wikipédia não teria alcançado o mesmo sucesso. Se o site tivesse problemas de disponibilidade, performance, perda de conteúdo ou fosse complexo para colaborar, muitas pessoas seriam desestimuladas a contribuir e o conteúdo perderia qualidade.

Além destes fatores, é possível sugerir um quinto motivo para o sucesso da Wikipédia, conforme descrito abaixo:

5. **Autopromoção:** ao colaborar, mesmo de forma anônima, é importante que o conteúdo tenha uma fonte de referência bibliográfica. E este pode vir a ser um dos fatores motivadores para que pesquisadores colaborem com conteúdo, para que as suas publicações possam ser listadas na seção “Referência” e o acesso ser impulsionada a partir da sua contribuição. Propagandas e spam não são permitidos na plataforma, mas referências bem colocadas são bem vindas.

Outro caso de sucesso é o Stack Exchange (<https://stackexchange.com>). Esta é uma rede de “Q&A” (perguntas e respostas) com mais de 170 comunidades dos mais diferentes tópicos, como música, biologia, jogos, engenharia, etc., sendo a comunidade

do Stack Overflow (<https://stackoverflow.com>), sobre questões de programação, a mais popular e com uma base de conhecimento de mais de 10 milhões de perguntas. O Stack Overflow foi a primeira comunidade e começou com uma estratégia um pouco diferente da Wikipédia. Em vez de focar em conteúdo anônimo, as colaborações (perguntas ou respostas) são identificadas pelos usuários que as criaram e, estes, possuem um perfil público. Neste perfil, se o usuário desejar, ele pode colocar o seu nome verdadeiro e sua foto, além de links para os seus sites pessoais e o seu currículo. Essa ideia incentivou um grande número de pessoas a criar conteúdo de qualidade para vincular o seu nome a resolução de problemas complexos de programação e, assim, mostrar o seu domínio do conhecimento em determinada área para servir de portfólio para aumentar as suas chances de ser bem sucedido em entrevistas de emprego.

Além de ter o apelo da autopromoção, o Stack Overflow também incentiva as contribuições através da gamificação. Ao contribuir, as outras pessoas votam de forma anônima, atribuindo +1 ou -1 e, assim, os criadores de conteúdo recebem pontos e medalhas.

Quando uma pessoa passa a ter um número elevado de pontos, a plataforma entende que este contribuinte foi bem avaliado pela comunidade e este passa a estar qualificado a realizar tarefas de moderação que outros usuários mais novos não podem realizar por não terem pontos suficientes. Isso gera aos maiores contribuintes uma sensação de “poder”, pois eles passam a ter ferramentas para controlarem o tipo de conteúdo que deve permanecer visível ao público. Esse tipo de mecânica existe desde o início do Stack Overflow, que a implementou como forma de garantir mais qualidade as contribuições. A plataforma criou um conjunto de ferramentas para que os colaboradores mais experientes pudessem julgar, eliminar, recuperar e melhorar os conteúdos postados. Além disso, é possível que um colaborador novato possa sugerir melhorias e correções no conteúdo de outro usuário e estas sugestões entram em uma fila para serem avaliadas por colaboradores mais experientes que irão aprová-las ou rejeitá-las.

De todas as ferramentas para controle de qualidade, a que mais se destaca é o sistema de votação. Uma pergunta pode ter respostas criadas por vários usuários, mas é a comunidade que define qual é a resposta mais correta e que melhor responde à pergunta que foi feita. Esta definição é através da contagem entre os votos positivos e negativos. Na lista de respostas, estas são ordenadas da maior pontuação para a menor, destacando assim o conteúdo que a comunidade mais gostou. Segundo ANDERSON et

al. (2012), o objetivo nesse sistema de votação não é destacar a resposta que mais ajuda a pessoa que *perguntou* e sim destacar a resposta que será mais útil à *comunidade*, pois outra pessoa poderá ter o mesmo questionamento e, ao consultar a plataforma, irá encontrar a mesma pergunta que gostaria de criar já com a resposta destacada.

Esse tipo de padrão de interação e votação não é comum na internet de maneira geral e era menos ainda quando o Stack Overflow foi lançado em 2008. O que é comum é o sistema de fóruns, onde uma comunidade se organiza criando tópicos e novas postagens são ordenadas cronologicamente e não por votos ou *likes*. Em crowd science, este modelo de fórum também é o mais comum sendo adotado, por exemplo, na plataforma Zooniverse (<https://www.zooniverse.org>), onde cada projeto pode ter o seu próprio fórum e as postagens dentro de um tópico são ordenadas cronologicamente, mesmo sendo possível votar (marcar como *helpful*) nas postagens mais pertinentes.

Outro conceito interessante relacionado a contribuição de voluntários é o Crowdfunding. Muitos sites, como o Kickstarter (<https://kickstarter.com>), usam a vontade das pessoas de ajudarem para que investidores consigam tornar ideias em produtos. Em alguns casos, as pessoas doam dinheiro para, se a ideia vingar, elas possam receber uma unidade do produto com um preço mais baixo do que elas pagariam em uma loja, mas em muitos outros casos, as pessoas doam apenas pelo altruísmo e pela esperança de que o mundo será melhor e mais interessante se determinada ideia puder ser concretizada em algo útil para as outras pessoas. O mesmo processo vem acontecendo com a ciência, como é o caso de duas plataformas específicas de *crowdfunding* para a ciência: Experiment.com (<http://experiment.com>) e Crowd.Science (<http://crowd.science>) que passam a mensagem de que as pessoas podem contribuir para um mundo melhor ao financiar a ciência que irá beneficiar a todos.

Portanto, concluindo, existem várias formas de motivar as pessoas a contribuírem em projetos de crowd science, o importante é identificar, dentro do público-alvo, quais as estratégias são mais eficazes. Pode ser que o simples fato de que o campo de pesquisa gera interesse ao cidadão cientista, ele venha a desejar transformar o seu hobby em algo útil colaborando gratuitamente, mas adicionar outras ferramentas para alimentar o seu desejo como gamificação ou recompensa como autopromoção, reconhecimento, interação social, ou dar ao voluntário prazer ao resolver desafios, trazem resultados significativos (NOV et al., 2010).

1.4 Motivação para uso da Plataforma Fast Science

A Web 2.0, onde os sites enfatizam o conteúdo gerado por usuários, traz a possibilidade de unir voluntários e pesquisadores (HUDSON-SMITH et al. 2009). Entretanto, os cientistas ainda precisam investir tempo e dinheiro na construção e manutenção de um website ou aplicativo e isso pode representar um esforço adicional que irá coibir a criação de uma iniciativa de crowd science.

Uma análise dos projetos de crowd science existentes revelou que existe uma gama muito grande de projetos similares, mas os seus códigos fontes não estão disponíveis para reuso em outras iniciativas. Por exemplo, The Seafloor Explorer (www.seafloorexplorer.org) e The Galaxy Zoo (www.galaxyzoo.org) são projetos de áreas de pesquisa muito distintas (oceanografia e astronomia), mas ambos possuem a mesma tarefa que é apresentar uma imagem e solicitar que a comunidade de voluntários a classifique seguindo um tutorial. Nos dois casos, qualquer pessoa pode contribuir sem necessidade de ter conhecimentos técnicos na área. Estes sites possuem um objetivo comum e o implementaram usando ferramentas semelhantes, mas cada um precisou desenvolvê-las por conta própria. Se um cientista quiser criar um novo projeto para catalogar imagens de pássaros, ele não teria uma forma de reaproveitar o sistema que foi desenvolvido para estes sites.

Olhando para o cenário atual, foi verificado que apenas a plataforma Fast Science, fruto da tese de doutorado de ESTEVES (2016), permite a hospedagem, criação e reuso de projetos de crowd science tanto para a coleta de dados como o processamento de dados científicos. As demais plataformas existentes são especializadas ou para hospedar projetos de processamento de dados, como as plataformas CrowdCrafting (<https://crowdcrafting.org>) e o Zooniverse (www.zooniverse.org), ou para a coleta de dados, como a iNaturalist (<https://www.inaturalist.org>) e o EpiCollect (<http://www.epicollect.net>), não sendo possível a realização das etapas de coleta e processamento dos dados em um mesmo ambiente. Essa vantagem, aliada ao conhecimento e facilidade de acesso ao código da plataforma Fast Science e ser esta uma iniciativa brasileira que permite que cientistas reutilizem ou criem tarefas a partir de componentes pré-existentes, foram fatores decisivos para a escolha deste sistema como infraestrutura para a realização deste estudo.

Além disso, para o desenvolvimento desta pesquisa de mestrado, as plataformas acima citadas foram analisadas junto com outros projetos populares de crowd science para entender quais são as estratégias utilizadas para garantir a qualidade dos dados e quais mecanismos para gamificação costumam ser implementados. Na tabela abaixo, este comparativo é apresentado, onde o tipo “Individual” significa que o sistema foi criado para apenas um projeto e “Plataforma” quando o sistema foi criado para hospedar diversos projetos distintos.

Tabela 1 – Comparativo entre iniciativas de crowd science

Iniciativa	Tipo	Qualidade	Gamificação
CitSci	Plataforma	Não possui	Não possui
CrowdCrafting	Plataforma	Tutoriais; comentários nas tarefas; redundância; estatísticas	Ranking
eBird	Plataforma	Material de treinamento; filtros detectam dados anômalos e encaminham para especialistas validarem	Colecionar imagens de observadas
EpiCollect	Plataforma	Não possui	Não possui
FoldIt	Individual	Algoritmo atribui uma pontuação para cada contribuição	Ranking; pontuação; medalhas
iNaturalist	Plataforma	Validação do nome das espécies; validação por consenso e por especialista	Ranking; colecionar imagens de espécies observadas
Old Weather	Individual	Incentivos para manter a transcrição do mesmo diário de bordo (qualidade aumenta porque grafia é conhecida)	Ranking; insígnias
Project Noah	Plataforma	Espécies definidas por consenso	Medalhas; barra de progresso; colecionar imagens de espécies observadas

Projeto Urubu	Individual	Especialistas validam	Não possui
SETI@home	Individual	Algoritmo determina a contribuição	Medalhas; progresso; usuário em destaque
Zooniverse	Plataforma	Treinamento; redundância de contribuições para obter consenso	Não possui

Observa-se que independente do objetivo ou do tipo da iniciativa (plataforma ou individual) alguns requisitos são comuns, como por exemplo: necessidade de treinamento, redundância das contribuições e elementos de gamificação: medalhas, indicativo de progresso e mecanismos para o reconhecimento daqueles que melhor contribuem para o projeto.

1.5 Questões de Pesquisa

Com base nesta introdução e no objetivo de aumentar a motivação dos voluntários e a qualidade das contribuições, as seguintes questões de pesquisa (QP) são levantadas com o intuito de guiar o estudo em questão:

QP1: A gamificação usada ajuda a aumentar a motivação dos voluntários?

QP2: A gamificação usada ajuda a aumentar a quantidade de contribuições?

QP3: A gamificação usada afeta a qualidade de dados?

O interesse na QP1 é saber se a gamificação, da forma implementada, motivou os voluntários a se interessarem mais por crowd science e a QP2 se essa motivação gerou mais contribuições. A diferença entre a QP1 e a QP2 é que, um voluntário mais motivado, não necessariamente contribui mais vezes. Ele pode usar a plataforma de outras formas ou usar a motivação para divulgar para outras pessoas.

Na QP3, o interesse é saber se a gamificação implementada afetou a qualidade, considerando que também foram implementadas ferramentas de qualidade de dados para evitar e dificultar que os voluntários insiram dados inválidos.

1.6 Metodologia

Para que uma pesquisa possa ser considerada sólida e relevante, é necessário apresentar qual foi a metodologia utilizada para que esta possa ser avaliada e, possivelmente, contestada. No caso, este trabalho utilizou a metodologia de *Design Science Research* (DSR) com base em DRESCH et al. (2015).

DRS é uma metodologia orientada a solução de problemas onde dado um determinado contexto, deseja-se determinar quais artefatos podem ser desenvolvidos ou melhorados para ter uma solução, não sendo esta necessariamente uma solução ótima ou perfeita, apenas melhor do que existe atualmente. Nesta definição, entende-se por “artefato” como todo elemento construído pelo homem e que, para esta construção, necessitou de inteligência e conhecimento.

Enquanto a Ciência Natural está preocupada em entender fenômenos complexos, descobrir como a natureza funciona e justificar os seus mecanismos, Design Science está preocupada em projetar sistemas que irão usar mecânicas já conhecidas e estudadas da natureza para a criação de artefatos que irão resolver problemas existentes.

DRS também apresenta um conceito de validade pragmática, onde a solução proposta precisa ser avaliada considerando fatores como custo/benefício, particularidades do contexto onde ela é aplicada, e as reais necessidades dos interessados na solução. Portanto, não basta encontrar uma solução. É necessário investigar a sua viabilidade e os seus impactos.

Segundo LACERDA et al. (2013) e PEFFERS et al. (2007), a utilização de DRS utiliza o seguinte fluxograma de processos:

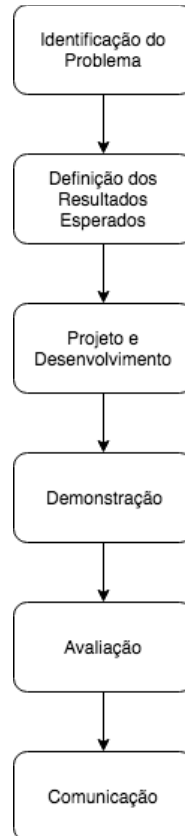


Figura 1 – Fluxograma de processos adaptado de PEFFERS (2007)

A partir desse fluxograma, podemos analisar cada etapa do processo de pesquisa conforme abaixo:

Identificação do Problema: trata-se de uma etapa de conscientização, onde o problema de pesquisa é identificado e as suas soluções são analisadas. Os problemas de interesse desta pesquisa são a dificuldade de se conseguir voluntários para contribuir altruisticamente com projetos de crowd science e como tratar a qualidade das contribuições que são realizadas por pessoas que não são especialistas na área e não possuem capacitação técnica.

Definição dos Resultados Esperados: após a identificação dos problemas de interesse desta pesquisa, é necessário definir quais os resultados esperados. Após o desenvolvimento de funcionalidades (artefatos) para promover gamificação, espera-se que a quantidade média de contribuições que um voluntário realiza na plataforma aumente de forma considerável. Entretanto, o que significa “um aumento considerável” não é definido pois não se sabe, a priori, qual impacto quantitativo o conjunto de ferramentas estudadas deve causar na plataforma. Além disso, outro resultado esperado é que, após a implementação dos artefatos de qualidade de dados, também haverá um aumento na porcentagem contribuições válidas, definindo como uma contribuição inválida as que são compostas por dados essenciais não preenchidos ou preenchidos com valores sem sentido.

Projeto e Desenvolvimento: após o levantamento dos artefatos para solucionar o problema, deve-se projetar como os mesmos serão implementados. Neste caso, o [Capítulo 3](#), “Proposta e Implementação”, destacará quais funcionalidades identificadas na literatura podem ser aplicadas no Fast Science e como as mesmas serão desenvolvidas considerando as modificações no código, na arquitetura e no modelo de dados do sistema. Para dar suporte a esta fase será realizado no [Capítulo 2](#), “Revisão da Literatura”, um levantamento de artigos relacionados com gamificação e qualidade dos dados, visando identificar e selecionar as técnicas que foram utilizadas com sucesso em outros projetos de crowd science e que sugerem que as mesmas podem ser combinadas na plataforma Fast Science para alcançar resultados semelhantes.

Demonstração: trata-se do uso dos artefatos desenvolvidos para solucionar o problema identificado. Esta demonstração pode ser feita através de um experimento, caso de uso ou mesmo uma simulação. No caso, esta pesquisa realizará um experimento com o objetivo de quantificar se os artefatos geraram ou não ganhos na quantidade e

qualidade das contribuições. Sobre este experimento, o mesmo é descrito no [Capítulo 4](#), “Experimento”.

Avaliação: com base nos resultados do experimento, é possível avaliar se os objetivos foram ou não alcançados. Neste trabalho, como o experimento foi projetado para dois grupos distintos, onde o primeiro grupo acessa a plataforma sem gamificação e o segundo grupo com gamificação, deseja-se observar no segundo grupo uma quantidade maior de contribuições. Além disso, para ambos os grupos, como a plataforma foi testada com as ferramentas de qualidade já implementadas, uma avaliação quantitativa da qualidade das contribuições será realizada e comparada com resultados antigos de outros testes realizados na plataforma antes da inclusão dos artefatos de melhoria de qualidade.

Comunicação: trata-se da forma que os resultados são apresentados, identificando os impactos dos artefatos na solução do problema, sejam estes positivos ou negativos. Esta etapa será tratada no [Capítulo 5](#), “Conclusão”.

Capítulo 2 – Revisão da Literatura

O objetivo desta dissertação é mostrar como plataformas de crowd science podem ser mais *eficientes* com o uso de ferramentas de gamificação e como podem ser mais *eficazes* com o uso de técnicas para qualidade de dados. Gamificação permite uma maior *eficiência* porque aumenta a taxa de retenção e a quantidade de contribuições por voluntários, o que permite ter uma maior produção com a mesma quantidade de pessoas. Já a qualidade de dados permite uma maior *eficácia* porque o resultado final será mais assertivo e confiável. Com isto em mente, as duas seções seguintes irão mostrar uma análise do que existe na literatura atual para que, no capítulo seguinte, possa apresentar o que será utilizado para avaliação.

2.1 Qualidade de Dados

2.1.1 Definição de Qualidade de Dados

Qualidade é um termo muito estudado no meio empresarial, com o objetivo de desenvolver processos que irão permitir a criação de produtos ou serviços que irão gerar mais satisfação aos clientes. No que tange a ciência, “qualidade” não está relacionada a “satisfação” do cientista e sim se os dados representam ou não de forma fiel uma determinada realidade ou se tratam de uma amostragem significativa da realidade para que a conclusão de uma determinada pesquisa não tenha um viés errado. Se uma pesquisa é enviesada significa que outro pesquisador, obtendo dados de outra maneira, terá resultados e conclusões diferentes, o que não é de interesse da ciência. POPPER (1959) defendeu que ocorrências únicas, sem reprodutibilidade, não são de interesse da ciência. Portanto, é necessário estabelecer critérios na obtenção e no tratamento de dados para que outro cientista, seguindo as mesmas práticas, consiga chegar a mesma conclusão.

Segundo GAMBLE et al. (2011), para que um dado possa ser utilizado em uma pesquisa, é necessário avaliar o mesmo sob três questionamentos:

- O dado é bom quando comparado por padrões e normas? (qualidade)
- O dado parece bom considerando a sua fonte? (confiança)
- O dado parece se encaixar nas necessidades da pesquisa? (utilidade)

GAMBLE et al. (2011) defendem que estas perguntas devem ser respondidas observando os dados em três dimensões: qualidade, confiança e utilidade. Assim, a qualidade do dado em si, dos critérios utilizados para a obtenção do dado, não é suficiente. É necessário avaliar também a sua confiabilidade e utilidade. GAMBLE et al. (2011) definem estas três dimensões desta forma:

- **Qualidade:** conjunto de características específicas e padronizadas que são aplicadas apenas ao dado de origem e não a como o mesmo será utilizado. Por exemplo, se um objeto foi pesado com um instrumento preciso, o dado (valor do peso) terá **acurácia** e, portanto, qualidade na medição. Outras características incluídas neste grupo são: consistência, corretude e atualidade.
- **Confiabilidade:** trata de características relacionadas ao processo em si que gera uma melhor **reputação** sobre os dados. Exemplos: objetividade, confiança, segurança e autoridade.
- **Utilidade:** são as características que interessam ao consumidor da informação e tratam da sua **relevância**. Como exemplos, podem ser citados: acessibilidade, disponibilidade, facilidade de entendimento, valor agregado, custo e aplicabilidade.

A tabela abaixo mostra um exemplo de como seria aplicar estas facetas para avaliar os dados em um projeto de crowd science com o objetivo de coletar fotos de pássaros para identificar e catalogar a fauna de uma determinada região.

Tabela 2 – Análise de qualidade de dados

Dimensão	Característica	Forma de mensurar
Qualidade	Acurácia	É possível observar os pássaros com clareza nas fotos?
Qualidade	Consistência	Todas as fotos são de pássaros?
Qualidade	Corretude	Qual é o índice de erros de identificação?
Qualidade	Atualidade	As fotos são recentes?
Confiabilidade	Reputação	Os voluntários contribuem de boa fé?
Confiabilidade	Objetividade	As identificações quanto as espécies são claras?
Confiabilidade	Confiança	Os voluntários são capazes de identificar as espécies?
Confiabilidade	Autoridade	Os voluntários possuem capacitação ou experiência profissional na área?
Utilidade	Relevância	Os resultados são de interesse da pesquisa?

Utilidade	Compreensibilidade	Os resultados são de fácil compreensão?
Utilidade	Custo	Qual é o custo ao gerir e analisar os resultados do projeto?
Utilidade	Aplicabilidade	Os resultados podem ser utilizados na pesquisa?

Como o nosso interesse é desenvolver ferramentas de software que ajudam a melhorar a qualidade dos dados, analisando a tabela anterior, observamos as seguintes ações que podem ser tomadas:

- **Acurácia:** permitir que os voluntários ou o moderador do projeto atribua uma nota de grau de qualidade a cada resultado.
- **Consistência:** permitir que os voluntários ou o moderador do projeto marque os resultados que são inconsistentes.
- **Corretude:** permitir que os voluntários ou o moderador do projeto avaliem ou atribuam uma nota a cada resultado.
- **Atualidade:** permitir que o moderador informe aos voluntários qual é o “prazo de validade” dos dados enviados pelos mesmos para que estes sejam válidos.
- **Reputação:** criar ferramentas que identifiquem manualmente e/ou automaticamente quais voluntários devem ser excluídos ou reconhecidos de acordo com as suas contribuições realizadas.
- **Objetividade:** oferecer campos de seleção como *dropdowns* ou *radio buttons*.
- **Confiança:** fornecer ferramentas de auditoria para mensurar o quanto um determinado voluntário está capacitado a contribuir. Se o mesmo falhar uma quantidade de vezes inaceitável, as suas contribuições podem ser descartadas e o voluntário advertido ou excluído.
- **Autoridade:** permitir que o moderador informe o grau mínimo de instrução que um voluntário precisa ter para que as suas contribuições sejam aceitas.
- **Compreensibilidade:** permitir que os resultados possam ser exportados de maneira que facilite ao pesquisador utilizar os mesmos.

2.1.2 Qualidade de Dados em Crowd Science

SHEPPARD et al. (2011) fizeram um levantamento sobre qualidade de dados em projetos de ciência cidadã e, concluíram que, apesar de todos os benefícios que estes projetos oferecem, uma preocupação comum aos pesquisadores é da qualidade de dados ser presumida como inferior. O que esses autores oferecem é um conjunto de medidas para reverter esta má impressão. Dentre elas, destacam-se:

- Redução de potenciais fontes de erros: erros acontecem, em geral, por falha na **coleta** ou na **entrada de dados**. Erros de coleta ocorrem quando o instrumento de medição está mal calibrado ou o método de obtenção não é adequado. Já os erros de entrada ocorrem quando os dados precisam ser transcritos manualmente. O que os autores identificaram foi que, em determinado projeto de ciência cidadã, a coleta dos dados era feita em pares onde um estudante lia o instrumento, falava em voz alta e o outro anotava. Posteriormente, outra pessoa lia os dados anotados e transcrevia para uma planilha Excel. Neste caso, a tecnologia reduz a fonte de erros, pois permite que os dados coletados pelo instrumento sejam armazenados e enviados automaticamente. Por exemplo, quando uma tarefa de crowd science permite que uma foto seja tirada utilizando um aplicativo de um dispositivo móvel e sua localização seja obtida de forma automática e enviada como resposta.
- Identificação de erros: duas formas apontadas para identificar erros nos datasets foram: **conhecimento do domínio** e **detecção de anomalias**. Os autores defenderam a importância de que, alguém com expertise na área, faça uma amostragem dos dados e avalie se os mesmos estão coerentes. Uma pessoa experiente pode perceber que determinadas informações estão erradas e descartá-las. Se a taxa de análise/descarte for baixa, os dados possuem uma boa confiabilidade. Sobre detecção de anomalias, uma técnica a ser aplicada pode ser a de ordenar os resultados ou plotar os mesmos em gráficos e ver se existem pontos muito distantes da curva esperada.
- Tratamento de erros: ao identificar que determinado dado está errado, é necessário avaliar o que fazer com o mesmo. Em geral, é melhor descartar um dado ruim, e ficar com ausência de uma medição no período, do que manter o dado errado. Em outros casos, pode ser melhor substituir o valor ruim pela média dos dados no período. Outro fato importante é que, nem

toda anomalia deve ser descartada. Existem sim fenômenos que podem gerar resultados muito distantes do esperado e nem por isso os valores estão errados. Neste caso, permitir que um dado possua um comentário associado pode ajudar. Por exemplo, se um voluntário mediu um valor muito estranho, ele pode adicionar informações que justificam o motivo do dado ter sido tão anômalo.

Além disso, SHEPPARD et al. (2011) evidenciaram que, ao aumentar o conjunto de práticas que garantam a qualidade, o envolvimento da comunidade de voluntários e os objetivos educacionais são maiores e não menores.

2.1.3 Melhorando a Qualidade e a Confiabilidade dos Dados

ALABRI et al. (2010) destacam que voluntários com baixa capacitação, além de contribuírem com dados errados podem gerar eventualmente resultados enganosos e que irão encaminhar uma pesquisa para um rumo incorreto. O que eles propõem, para isso, é um framework para combinar melhorias de qualidade de dados e métricas de confiança, sendo aplicada especificamente para dados gerados por multidões.

Para melhoria de qualidade de dados, eles propõem, primeiramente, catalogar as métricas de qualidade e observar quais ferramentas podem ser implementadas no sistema para potencializar cada característica. Esta abordagem é semelhante a que foi proposta por GAMBLE et al. (2011), sendo a principal diferença o fato de ALABRI et al. (2010) focar em qualidade para projetos com multidões.

No experimento que ALABRI et al. (2010) fizeram, eles conseguiram melhorar a qualidade dos dados em 70%, após observarem que a maioria dos erros eram gerados devido aos seguintes fatores:

- Falta de análise de consistência e validação.
- Falta de extração automatizada dos dados.
- Falta de identificação do usuário que gerou a contribuição. Todos os dados eram anônimos.
- Ausência de um modelo de dados.
- Falta de feedback aos voluntários sobre a qualidade da contribuição que estes estão fazendo.
- Falta de ferramentas de gráficos e visualização de dados.

Eles sugerem também a implementação de métricas de confiabilidade. Para isso, eles citam GOLBECK (2009) que avaliou que a reputação dos usuários é a que define a

confiabilidade de um sistema. Em sites de comércio eletrônico como eBay e Amazon, os usuários possuem uma pontuação que varia de acordo com a avaliação por parte de outros usuários em relação a compra e venda de produtos. A partir do modelo de reputação descrito por GOLBECK, eles definiram o próprio modelo.

Para o **cálculo de reputação** de um usuário, um conjunto de atributos precisa ser avaliado. Neste caso, eles consideraram os atributos abaixo, dando notas e pesos para cada um deles para calcular a reputação do usuário.

- Votação direta entre membros.
- Reputação através da nota da resposta de questionários.
- O grau de instrução do voluntário.
- O papel do voluntário na pesquisa.
- A qualidade dos dados que o voluntário contribuiu no passado.
- A quantidade de dados já contribuídos pelo voluntário.
- Frequência e período de contribuições.
- A quantidade de treinamentos realizados pelo voluntário.
- Algoritmo para inferir confiabilidade através do relacionamento entre membros.

Além disso, eles trataram o cenário onde o usuário inicia na plataforma sem uma reputação já que o sistema não tem como mensurar os atributos anteriores. Nesse caso, eles implementaram um sistema de confiança em rede. Este algoritmo é descrito como uma tentativa de atribuir uma reputação provisória para novos membros de acordo com o grau de semelhança que o mesmo possui com outros membros. Por exemplo, se um novo usuário possui mestrado em determinada área e realiza certos treinamentos, ele possui uma grande chance de apresentar contribuições de mesma qualidade que outro usuário com o mesmo perfil, sendo a sua reputação inicial equiparada a do usuário semelhante. Posteriormente, a sua reputação é ajustada de acordo com a maneira que ele interage com o sistema e como o sistema avalia segundo os atributos listados anteriormente.

Por último, ALABRI et al. (2010) frisam a importância do sistema oferecer ferramentas de visualização para que dados espúrios e anomalias possam ser retirados do dataset de resultados. Esta recomendação também foi feita por SHEPPARD et al. (2011), conforme descrito anteriormente.

Outro artigo interessante realizado neste contexto de melhoria de qualidade de dados foi escrito por WIGGINS et al. (2011), onde uma pesquisa realizada com 73 pessoas que participaram de projetos de crowd science permitiu a criação de um framework para melhoria de qualidade com 18 mecanismos recomendados. Destes, destacam-se os seguintes mecanismos:

- **Tarefas repetidas:** múltiplos participantes avaliam a mesma tarefa para aumentar a confiança de uma resposta.
- **Tarefas de controle:** avaliar respostas dos participantes em tarefas onde a resposta correta já é conhecida. O objetivo é identificar participantes que geram contribuições ruins.
- **Treinamento de participantes:** adição de tutoriais para orientar os voluntários.
- **Revisão de experts:** revisão de dados por experts na área de conhecimento.
- **Remoção de anomalias:** remoção automatizada ou manual de dados anômalos.
- **Feedback:** informar participantes sobre o seu desempenho com o objetivo de educá-los para que melhorem as suas contribuições ou motivá-los a continuar.
- **Triangulação de dados:** comparar respostas com dados de telemetria. Por exemplo, usar a posição GPS do smartphone para indicar o local.
- **Normalização de dados:** métodos estatísticos para identificar e corrigir dados espúrios.

2.1.4 Definindo a Qualidade Através da Multidão

Uma proposta alternativa para melhorar a qualidade dos resultados é deixar que a própria multidão defina o que é qualidade. Nesta abordagem, a qualidade do conteúdo produzido por uma comunidade é tão boa quanto a inteligência coletiva dos participantes, como numa sociedade democrática e igualitária. Neste modelo, o próprio grupo define os resultados que são válidos e os que não são.

RODRIGUEZ et al. (2007) propuseram um sistema chamado Smartocracy, onde cada pessoa vota em uma determinada opção e, depois, as relações entre as pessoas são avaliadas para determinar os pesos de cada voto. Estes pesos dependem da quantidade de relacionamentos. Se um usuário possui muitos relacionamentos, então existem

muitas pessoas que confiam nele e, por isso, seu voto possui um peso maior no sistema. Testes foram feitos comparando esta estratégia com outras técnicas de votação, e viram que, dessa forma, para escolher “1 opção dentro de 13 soluções”, essa foi a estratégia com a maior correlação de votos por pessoas do mesmo grupo. Ou seja, em um grupo grande de amigos, eles pensavam de forma mais parecida.

Uma estratégia semelhante é utilizada pelo sistema de busca do Google. Se existem muitos links apontando para uma determinada página, é porque esta página possui uma relevância maior e, por isso, é destacada para ser mais facilmente encontrada. Além disso, essa página mais relevante, quando aponta para outra página, gera um peso maior no apontamento e faz com que a página apontada receba parte de sua relevância.

No contexto de crowd science, o que pode ser utilizado é: se um usuário possui muitos relacionamentos com outros membros, as suas contribuições terão um nível de confiança maior do que a de outro usuário sem relacionamentos.

Outro sistema de qualidade definido pela multidão foi proposto por ANTELIO et al. (2012) e este chama-se Qualitocracy. A ideia é que, como foi discutido por GAMBLE et al. (2011), em projetos de crowd Science é necessário avaliar os dados sob múltiplos aspectos para poder definir se os mesmos são bons ou ruins. Adicionalmente, é importante ter em mente que esta análise depende do contexto, já que para alguns projetos a corretude do nome de uma espécie, por exemplo, pode ser fundamental e, em outros, a localização da informação pode ser mais importante.

A framework Qualitocracy propõe que, após iniciar um experimento de crowd science, um grupo de cientistas irá avaliar a qualidade dos dados obtidos sob diferentes dimensões de qualidade. Cada um dará uma nota em cada quesito para avaliar se, por exemplo, uma determinada foto está suficientemente nítida. Após uma rodada inicial de análise, um sistema de avaliação de votos coletivos deve ser utilizado. Neste caso, o Smartocracy é citado como uma possibilidade. A votação coletiva irá identificar, para cada cientista, qual dimensão de qualidade ela vota de maneira mais parecida com o que é esperado pelo coletivo de cientistas. Por exemplo, se um cientista vota no quesito “nitidez da foto” de maneira coerente ao esperado pelo coletivo de cientistas, futuramente, em outras análises, ele pode ser requisitado para avaliar apenas a característica “nitidez da foto”. Já em outro caso, se um cientista deu notas para o quesito “corretude do nome da espécie” de forma muito diferente do classificado por

outros cientistas, então ele não deve avaliar novamente essa característica em futuras análises porque a sua opinião diverge da maioria.

O objetivo do Qualitocracy é desenvolver um sistema de qualidade definido pelo coletivo de cientistas e manter o andamento de um determinado projeto otimizando o tempo dos cientistas ao avaliar os dados que devem ser aproveitados e os que devem ser descartados.

2.2 Gamificação

2.2.1 Definição de Gamificação

De fato, um dos maiores desafios de crowd science é manter os voluntários motivados a contribuir de forma altruísta, sem remuneração, e a proposta é usar gamificação para ajudar a minimizar este problema. A gamificação ajuda porque ela traz elementos da mecânica de jogos para sistemas comuns e jogos são atividades que dão prazer porque permitem que as pessoas estimulem a competição e tenham a satisfação de serem melhores do que outras pessoas (ou a um computador), ou estimulem a imaginação, permitindo uma “fuga” da vida real, indo para uma realidade alternativa onde os seus desejos e vontades são realizáveis.

O jogo mais famoso desenvolvido para um projeto de crowd science é o FoldIt (<https://fold.it>). Nele, o jogador é apresentado a uma estrutura molecular e tem como objetivo resolver um quebra-cabeça onde ele precisa encontrar formas de dobrar proteínas com o objetivo de encontrar pontos de maior estabilidade. O que foi observado, na prática, é que humanos possuem mais capacidade em encontrar padrões e soluções para esse tipo de problema do que os algoritmos atuais. Enquanto os voluntários se divertem em resolver os problemas e se destacam nos rankings de melhores colocados, os resultados desta pesquisa ajudam no tratamento de diversas doenças e os idealizadores do projeto já receberam diversos prêmios pela ideia.

Entretanto, gamificação não precisa de um “jogo” propriamente dito para ser utilizada. Inclusive, DETERDING et al. (2011) definiram gamificação como “o uso de elementos de jogos e técnicas de design de jogos fora do contexto de jogos” e WERBACH (2014) como “o processo de tornar as atividades mais parecidas com jogos”. Portanto, o projeto de crowd science não precisa desenvolver um jogo, mas precisa aplicar técnicas que estimulem, proporcionem prazer e gerem o “vício”, a recompensa e o reconhecimento característicos dos jogos. Como dito por

STOYANOVA et al. (2017), gamificação introduz a dinâmica de jogos em processos e em diversas áreas com o objetivo de motivar, engajar e atrair.

Para um entendimento mais preciso da definição de gamificação feita por DETERDING et al. (2011), o artigo de GROH (2012) começa explicando o que é um “jogo”. Para isso, ele segue CALLOIS et al. (2001) no conceito de que “jogar” possui duas vertentes: *paidia* e *ludus*. *Paidia* é o “brincar” no sentido livre, de forma espontânea e criativa. Já o *ludus* é o “jogar” seguindo regras claras e pré-definidas. Gamificação encontra-se na categoria “*ludus*”, onde os elementos utilizados exigem do “jogador” o cumprimento de regras pré-estabelecidas.

Quanto ao termo “fora do contexto”, GROH (2012) explica que as técnicas de criação de jogos são utilizadas em aplicações sérias, que não apresentam componentes “jogáveis”. Estas aplicações, quando “gamificadas”, passam a ter regras para a obtenção de recompensas digitais, como pontuação, classificação ou medalhas.

Ainda sobre o termo Gamificação, conforme levantado por MORA et al. (2015), o termo foi definido pela primeira vez em 2002 por Nick Pelling como “a aplicação de design de interfaces gráficas parecidas com jogos para transformar aplicações eletrônicas mais agradáveis e rápidas”, mas a ideia em si não é tão nova assim. A organização “The Boys Scouts of America”, uma grande organização de jovens escoteiros, começou a usar medalhas para simbolizar status e mérito em 1911. O objetivo na época já era de motivar os jovens escoteiros através de premiações simbólicas por terem aprendido uma determinada habilidade, como “primeiros socorros” ou “cientista de plantas”.

Outra referência do passado é em relação ao judô que começou em 1907 a utilizar faixas coloridas para indicar o grau de habilidade dos praticantes. O que também caracteriza um simbolismo com objetivo de motivar o treino e continuidade da prática.

2.2.2 Gamificação em Crowd Science

Gamificação possui o objetivo de tornar as aplicações mais interessantes e promover a retenção dos usuários, mas nem sempre esse objetivo é alcançado. Em alguns casos, alguns usuários podem abandonar a plataforma por preferirem uma interface mais séria ou madura (PRESTOPNIK, 2012). Além disso, uma preocupação importante é em relação a qualidade de dados, já que algumas pessoas acabam contribuindo o mais rápido possível apenas para cumprir metas e ganhar pontos virtuais, indo contra os objetivos de uma pesquisa (BOWSER, 2013).

EVELEIGH et al. (2013) fizeram um levantamento de prós e contras na aplicação de gamificação em projetos de crowd science após analisarem o seu uso no projeto Old Weather (<https://www.oldweather.org>), que foi concebido para que voluntários pudessem ajudar a transcrever manuscritos antigos de observações meteorológicas.

Os cientistas responsáveis pelo Old Weather reconheceram que transcrição manual de textos é uma tarefa chata e cansativa e, por isso, decidiram aplicar gamificação em seu projeto. Eles criaram um sistema de ranking de acordo com a quantidade de contribuições de cada participante para encorajamento. Os usuários se associavam a um “navio” (representando um grupo de dados) e iniciavam na categoria “Cadete” e progrediam para “Tenente” e depois para “Capitão”. O “Capitão” era o voluntário com o maior número de contribuições naquele grupo, podendo perder a posição para outro voluntário mais assíduo, semelhante com a estratégia de retenção do Foursquare (<https://foursquare.com>) que usava o conceito de “Prefeito” do lugar.

Essa estratégia, além de incentivar a continuidade das contribuições, tinha o objetivo de incentivar o usuário a permanecer contribuindo em um mesmo conjunto de dados para que o mesmo se familiarizasse com a letra de determinada pessoa em um diário de bordo.

Uma pesquisa feita com 545 pessoas permitiu as seguintes conclusões:

- **Pontos Positivos:**

- **Identificação:** o tema de “marinheiros” e “transcrever diários de bordo” permitiu que pessoas interessadas na área tivessem mais vontade de ajudar o projeto.
- **Acompanhamento:** alguns usuários disseram que o sistema de pontuação ajudava a dar uma noção de quanto eles já haviam contribuído e não apenas para se exibirem.
- **Competição:** o sentimento de competir, o desafio e a satisfação pessoal de superar outros competidores e se tornar o “capitão de um navio” deu uma motivação extra para continuarem a contribuir.

- **Pontos Negativos:**

- **Quantidade em vez de qualidade:** alguns voluntários perceberam que estavam contribuindo mais rápido do que deviam apenas para acumular pontos e sabiam que assim estariam atrapalhando o próprio projeto.

- **Desmotivação:** competidores com baixa pontuação se sentiram desmotivados por estarem muito distantes dos competidores mais assíduos e isso fez com que alguns ficassem desestimulados.
- **Estresse:** manter a posição de “capitão” exigia dedicação e acompanhamento. Em alguns momentos, isso era estressante e cansativo, tirando o prazer da atividade.
- **Desconfiança:** alguns usuários temiam que outros usuários poderiam estar trapaceando para subirem mais rapidamente nos rankings e esse sentimento era desestimulante.

Para balancear os pontos positivos e negativos, EVELEIGH et al. (2013) sugerem:

- Para cada mecanismo de pontuação, criar mais categorias e metas. Uma maior quantidade de estágios faz com que o usuário sinta, a todo momento, que a sua contribuição está tendo valor o tempo todo.
- Criar ferramenta de feedback e valorizar mais as contribuições corretas do que a quantidade de contribuições. Permitir que os voluntários verifiquem a qualidade das próprias contribuições e corrijam as ruins.
- Reconhecer que os participantes possuem níveis diferentes de interesse. Para alguns, a temática e “histórias” podem ser chatas e apenas um obstáculo para a contribuição. Portanto, permitir que o colaborador desative ou esconda algumas funcionalidades pode ser interessante. Além disso, criar tarefas para grupos ou concursos pode ajudar a revigorar o interesse de participantes antigos.

Como foi observado no projeto Old Weather, os próprios usuários reconhecem que, às vezes, contribuem rápido demais apenas para ganhar pontos. Esse tipo de gamificação pode ser nociva para a qualidade dos dados. Por isso, é importante tomar alguns tipos de cuidados. Por exemplo, ao invés de dar pontos para cada contribuição, poderia ter sido implementado que apenas contribuições validadas por outro usuário que seriam bonificadas e, em caso de erros, haveria penalidades. Nesse caso, talvez a gamificação perdesse um pouco do seu atrativo, mas em troca ela não causaria problemas na qualidade.

2.2.3 Aplicando Gamificação

Antes de aplicar gamificação em um sistema, é muito importante pensar na motivação para o seu uso e nas consequências. A implementação impacta muito se a resposta dos participantes será ou não positiva. Uma determinada estratégia pode funcionar para um público-alvo, mas não ser efetiva para outro tipo de público. Para entender quais ferramentas temos a nossa disponibilidade, PEREIRA (2014) identificou 8 mecânicas que podem ser aplicadas:

- **Pontuação:** as atividades podem oferecer pontos como recompensa. Isso ajuda o usuário a ter uma sensação de avanço.
- **Níveis:** ao conhecer os níveis disponíveis, o colaborador irá criar mentalmente metas para atingi-los.
- **Placares (rankings):** o usuário consegue comparar a sua posição com a de outras pessoas e, assim, criar metas pessoais.
- **Medalhas:** são recompensas que são dadas ao completar tarefas específicas. Servem como um direcionador, para que um comportamento específico seja incentivado.
- **Ambientação:** é a apresentação de um novo usuário ao sistema de gamificação.
- **Desafios/Missões:** dão ao colaborador uma meta difícil a ser alcançada para uma recompensa maior por um esforço mais significativo.
- **Ciclos de engajamento:** eventos permitem trazer de volta e reengajar “jogadores” que haviam se desestimulado com a plataforma. Uma novidade no sistema incentiva a retomada das contribuições.
- **Feedback/Reforço:** permitem que o colaborador se situe no “jogo” e possa enxergar como o seu desempenho está sendo avaliado.

O site Stack Overflow (<https://stackoverflow.com>) não é uma rede de crowd science, e sim um crowdsourcing de questões e respostas técnicas de computação, mas é um caso de sucesso com mais de 10 milhões de questões. Assim sendo, ele acaba sendo estudado por diversos pesquisadores e hoje existem dezenas de artigos citando as técnicas que ele aplica para qualidade de dados e gamificação.

Como a sua base de dados é gratuita e disponibilizada online para qualquer um (exceto os dados pessoais dos usuários), várias pesquisas fazem análises estatísticas da eficácia das medalhas. Por exemplo, BORNFIELD & RAFAELI (2017) estudaram o

impacto da introdução das medalhas Excavator e Archaeologist que foram criadas em agosto de 2011 para recompensar os usuários que editavam posts antigos (com mais de 6 meses de idade). O resultado foi que a quantidade de edições nesta época teve um aumento muito significativo, que está relacionado a introdução destas recompensas. Abaixo, o gráfico de média móvel usa uma janela de 180 dias e mostra na linha vermelha vertical a época em que a introdução das medalhas ocorreu.

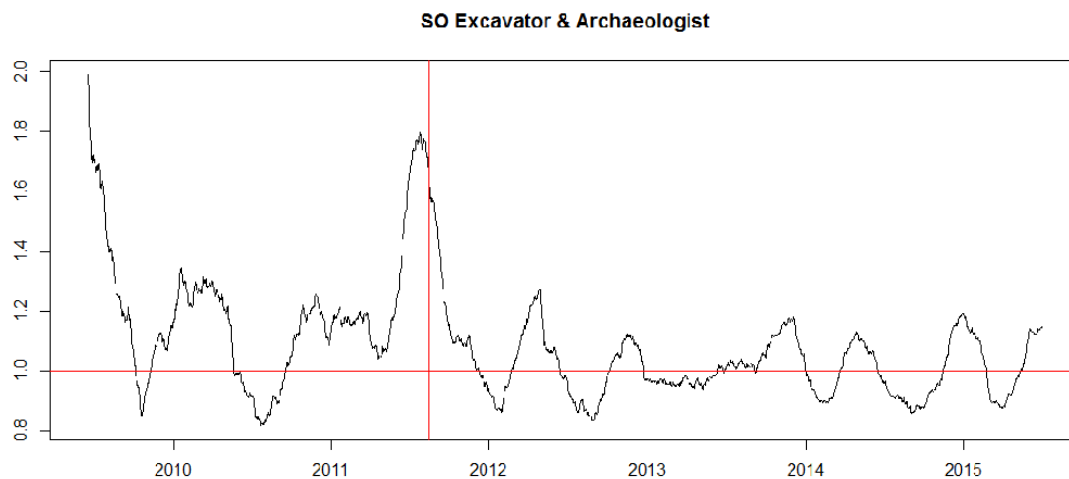


Figura 2 – Gráfico de média móvel (BORNFELD e RAFAELI, 2017)

Considerando que é verdade que a introdução de medalhas afeta no comportamento dos colaboradores, outros autores, como GHARIBI e MALEKZADEH (2017), incentivam a criação de sistemas de recomendação de medalhas. Os resultados da modelagem matemática realizada revelou que recomendar uma medalha atingiu um nível de acerto de 70% dos casos e recomendar duas medalhas elevou para 80% de acertos. Um sistema como o Stack Overflow, com 97 tipos diferentes de medalhas, pode fazer com que o usuário fique perdido em relação a como contribuir. As recomendações servem para mostrar quais medalhas se adequam mais ao perfil do usuário e quais tarefas ele tem maior facilidade de completar.

No Stack Overflow, existem três níveis para as medalhas: bronze, prata e ouro. As medalhas do tipo bronze são as mais fáceis de serem conseguidas. Estas servem para “educar” novos usuários sobre como o site funciona. Por exemplo, a medalha Autobiographer é concedida quando o usuário preenche o próprio perfil e Citizen Patrol quando o usuário utiliza a ferramenta de “sinalização” para marcar posts que precisam de intervenção de um moderador. Estes tipos de medalham se enquadram na mecânica de Ambientação, conforme definido por PEREIRA (2014). Já as medalhas do tipo prata

e ouro são mais difíceis de serem conseguidas e servem para dar metas aos usuários e educarem eles a seguirem um comportamento desejado.

O Stack Overflow também oferece eventos anuais de motivação. O WinterBash ocorre todo final de ano e presenteia os usuários com “chapéus” engraçados que os participantes podem adicionar aos seus avatares para decoração. Esse evento usa chapéus no lugar de medalhas, mas com a diferença de serem recompensas por comportamentos muito mais específicos e difíceis de serem obtidos. Respeitando quem não gosta desse tipo de brincadeira e prefere um site mais sério, este evento pode ser desativado pelo usuário e o tema tradicional não será alterado para ele.

Como participante assíduo do Stack Overflow por mais de 6 anos, descrevo as seguintes características do site como as suas principais ferramentas de gamificação para fidelizar a sua base de usuários:

- **Perfil público:** como é possível associar as contribuições técnicas com o seu usuário e preencher um perfil público com dados de contato e currículo, muitos usuários enxergam a possibilidade de serem contactados por empresas oferecendo emprego. Isto serve como um motivador adicional para que estes contribuam com mais qualidade. Existe, inclusive, uma extensão do site chamada de Stack Overflow Jobs para esta finalidade.
- **Ranking:** os melhores usuários da plataforma, por semana, mês, ano e ao total, são listados. Da mesma forma que o perfil público, além de uma boa posição no ranking gerar uma satisfação pessoal, ela aumenta as chances do seu perfil ser encontrado para receber uma oferta de emprego.
- **Pontuação/Reputação:** boas contribuições são avaliadas por outros usuários que votam e dão pontos para os outros colaboradores. Muitas pessoas criam metas pessoais e ficam satisfeitas quando conseguem cruzar certas faixas de pontuação, como exemplo: “mil pontos” ou “dez mil pontos”.
- **Funcionalidades travadas:** com o aumento da reputação (pontuação), novas funcionalidades são destravadas. Por exemplo, para que um usuário possa apagar uma contribuição ruim de outro usuário, ele precisa ter uma reputação alta para executar a ação. A sensação de “poder” que é dada com o aumento da quantidade de pontos ajuda na motivação.

- **Recompensas:** um usuário pode oferecer uma quantia significativa de pontos para que outros usuários o ajudem a resolver determinada questão. Essa quantia adicional que é ofertada é atrativa e recebe destaque no site.
- **Medalhas/Eventos:** conforme já foi discutido, o sistema oferece um grande número de medalhas para guiar o comportamento dos usuários e motivá-los.

Da mesma forma que observado por EVELEIGH (2013) no projeto Old Weather JIN et al. (2015) também identificaram problemas com a estratégia de gamificação do Stack Overflow. Uma expressão comum na plataforma é a “Fastest Gun in the West” que faz alusão aos duelos armados no faroeste, comparando com os usuários que respondem novas questões da forma mais rápida possível. O motivo é que a ação que dá maior pontuação é quando o usuário responde uma questão e esta é avaliada como certa por outro usuário. O problema é que é necessário que outro usuário, registrado no site, e com pontuação maior que 15, visualize a sua resposta. Como existem mais de 10 milhões de perguntas, a chance de estes usuários visualizarem é bem pequena, mas acaba sendo maior quando a questão é recente. Como existem centenas de usuários monitorando as novas questões para adicionar as suas respostas, o usuário que for mais rápido acabará recebendo a atenção de mais “avaliadores” e, conseqüentemente, mais pontos. Esse comportamento causa os seguintes efeitos negativos ao site:

- Respostas corretas, mas extremamente curtas e objetivas. Estas respostas às vezes deixam de considerar alguns cenários ou uma explicação simples demais pode não ajudar um usuário com pouco conhecimento no assunto.
- Questões simples são priorizadas. Os usuários deixam de contribuir em questões mais difíceis porque a resposta será demorada e, passada a “novidade”, poucas pessoas estarão visualizando (e votando) quando a resposta for postada.

YIN et al. (2015) analisaram 100 mil postagens de um intervalo de tempo de 4 meses e identificaram que esse problema de respostas rápidas afeta cerca de 8% da comunidade e que estes criam respostas, em média, com menos de 6 minutos. Além disso, quem usa desta prática possui uma tendência de usar tags (marcações nas postagens) diferentes da maioria. Isto é, como as tags das questões também afetam as medalhas recebidas, o uso de tags também é gamificado.

2.2.4 Dando recompensas

Existem vários tipos de recompensas virtuais que podem ser ofertadas: pontos, níveis, títulos, medalhas, troféus, etc. A questão é como definir *o que* dar, *quando* dar e *quanto* dar de recompensa. A seguir, uma análise para definir cada tópico:

- **O que dar:** depende da temática do projeto. No exemplo discutido anteriormente, o Old Weather tratava de relatórios de navegação, então foi natural que as recompensas fossem de títulos de “Cadete”, “Tenente” e “Capitão”.
- **Quando dar:** existe um consenso na indústria de jogos que diz que as recompensas precisam seguir uma função crescente e não-linear, isto é, para cada medalha ou nível alcançado, a próxima recompensa deve ter uma dificuldade maior de ser conseguida (ZICHERMANN & CUNNINGHAM, 2011). Além disso, as recompensas em gamificação devem ser dadas principalmente para manipular o jogador a realizar tarefas chatas ou que não são interessantes. Se uma tarefa já é do agrado dele, completá-la já seria uma satisfação e, por si só, uma recompensa.
- **Quanto dar:** recompensar em excesso torna o artifício sem graça (ZICHERMANN & CUNNINGHAM, 2011), por isso é importante conhecer o seu público-alvo, suas motivações e fazer testes controlados. Uma estratégia é testar diferentes parâmetros e configurações com pequenas amostragens do público-alvo e usar o feedback para “calibrar” como e quantas recompensas devem ser dadas para determinadas ações.

2.2.5 Ranqueamento dos jogadores

Um ranking de usuários é uma proposta visual muito simples, onde cada jogador consegue se comparar com facilidade em relação aos demais jogadores. Além de estimular a competição direta entre eles.

Entretanto, quando um usuário consegue se distanciar demais de outros competidores, isso acaba gerando desmotivação nos demais jogadores (WERBACH e HUNTER, 2012). Para tratar isso, existem estratégias como criar diversos rankings (semanais, mensais) ou por contexto (projeto A, projeto B).

Em jogos que possuem uma competição contínua e direta entre jogadores, como por exemplo, o ranking das melhores seleções de futebol da Fifa, alguns algoritmos

foram elaborados para que partidas mais difíceis, que envolvem jogadores mais bem colocados, valessem mais pontos e que o cálculo geral de pontos reduz a pontuação quando uma partida foi vencida no passado distante. Esses tipos de algoritmos não foram estudados porque dificilmente fazem sentido em projetos de crowd science.

Capítulo 3 – Proposta e Implementação

Como visto na literatura, existem diversos elementos de jogos e ferramentas interessantes que podem ser empregadas em projetos de crowd Science com o objetivo de aumentar a confiabilidade e qualidade dos dados e ao mesmo tempo promover mecanismos que estimulem a participação e o engajamento dos participantes. Neste trabalho, a proposta é pinçar as ideias que parecem ser mais aderentes as necessidades mapeadas e implementá-las na plataforma Fast Science já existente.

3.1 Histórico e estado inicial do Fast Science antes deste trabalho

O Fast Science foi desenvolvido para exemplificar e avaliar as propostas defendidas na tese de ESTEVES (2016). Ele foi avaliado por um grupo de pesquisadores que criaram projetos na plataforma, captaram voluntários e responderam a um questionário para identificar a utilidade e a facilidade de uso, incluindo os pontos fortes e fracos do sistema desenvolvido, sob a visão do gestor do projeto.

Em resumo, a plataforma foi lançada com as seguintes funcionalidades:

- Telas com a apresentação do portal e informações sobre o mesmo.
- Funcionalidades de segurança, permitindo o cadastro de usuários, login, logout e recuperação de senhas.
- Perfil do usuário
- Sistema de criação de projetos e workflow de tarefas, reuso de workflows, criação de tutoriais e exportação das contribuições.
- Sistema para contribuição em projetos e execução dos workflows de tarefas.
- Cadastro de parceiros.
- Funcionalidades de interação com comunidades (blog e eventos).



Figura 3 – Pagina inicial da Plataforma Fast Science



Figura 4 – Lista de projetos disponíveis por tipo (coleta ou processamento)

Figura 5 – Workflow de contribuição no projeto “Jogos e Emoções”

Nas funcionalidades iniciais, não existia nenhuma ferramenta de gamificação e as ferramentas implementadas para melhorar a qualidade dos dados foram: o sistema de redundância, treinamento de participantes e a triangulação de dados (uso da posição GPS para indicar o local de coleta dos dados). O sistema de redundância foi criado para permitir que o pesquisador configurasse, em projetos de processamento de imagens, uma determinada quantidade de colaborações por imagem que seriam necessárias para que a imagem não precisasse mais ser avaliada por outros usuários. Por exemplo, se o pesquisador escolhe o valor 5 para imagens de entrada, significa que 5 usuários distintos precisam analisar a mesma imagem antes que esta seja retirada e o resultado possa ser validado. A vantagem é que ter 5 respostas distintas para o mesmo dado de entrada permite que o pesquisador valide a resposta que obtiver a maior redundância, identifique e elimine respostas anômalas e reavalie o dado quando houverem muitas respostas divergentes.

Em relação ao treinamento dos participantes, outra prática de controle de qualidade, defendida por GAMBLE et al. (2011), existe na plataforma Fast Science a opção para o pesquisador ou gestor do projeto preencher uma seção, com textos e imagens, para servir de informativo ou tutorial aos usuários. Assim, o voluntário

pode se informar sobre o projeto, sobre os objetivos e como o mesmo deve proceder para executar as tarefas, servindo como um breve treinamento e capacitação.

Além dessa análise do estado atual da Plataforma Fast Science, para o início desta dissertação, novas versões de código e melhorias gerais de layout foram implementadas e a plataforma foi migrada para outra infraestrutura de servidores, continuando hospedada na Amazon Web Services, mas migrando de servidores Windows para Linux com o objetivo de reduzir os custos.

Usando como ponto de partida a plataforma pré-existente, este trabalhou desenvolveu como artefatos um conjunto de melhorias que serão descritas a seguir.

3.2 Qualidade de Dados

A partir das sugestões de ferramentas encontradas na literatura, as seguintes funcionalidades foram implementadas na plataforma Fast Science visando aumentar a confiabilidade e qualidade dos dados fornecidos pelos participantes:

- Sistema de validação de contribuições
- Sistema de auditoria
- Sinalizadores de problemas
- Notificações
- Melhorias na execução de tarefas
- Detector de texto sem sentido
- Estatísticas de contribuições

3.2.1 Sistema de validação de contribuições

Segundo GAMBLE et al. (2011), a existência de um perfil de “moderador” é necessária para avaliar as contribuições e rejeitar as ruins. Nesse caso, a plataforma Fast Science possui a ideia de “pesquisador” que é o usuário que cria projetos e fica responsável pelo mesmo, sendo este aquele que irá validar as contribuições.

GAMBLE et al. (2011) também defende que o moderador possa atribuir uma nota de qualidade a cada contribuição, indicando um grau de confiança aos dados. Com base nos projetos existentes na plataforma, esta funcionalidade de definir nota não pareceu ser muito útil. No caso, foi implementado apenas as opções de “aprovar” ou “rejeitar”.

A importância da reputação dos dados também é defendida por GAMBLE et al. (2011). Nesse caso, sempre que um pesquisador rejeitar uma contribuição, o usuário é

marcado como “não confiável” para aquele projeto. Dessa forma, quando o pesquisador exportar os resultados no final do seu projeto, ele terá como filtrar dados que possuem uma reputação duvidosa. Isto é, se o usuário contribuiu de forma errada uma vez, ele não é confiável. Entretanto, esta marcação vale apenas para as contribuições atuais do usuário. Segundo WIGGINGS et al. (2011), os usuários precisam ser avisados e “educados” quando cometem erros. Dessa forma, quando o pesquisador rejeita uma contribuição, ele é obrigado a justificar e essa mensagem será enviada como feedback para o voluntário. Dessa forma, as próximas contribuições do mesmo não estarão marcadas como suspeita, a não ser que outro erro seja cometido.

SHEPPARD et al. (2011) defendem que o moderador tenha a possibilidade de corrigir dados errados durante a sua validação. Essa funcionalidade foi implementada, mas não exatamente como sugerido. No caso, como a mensagem de validação é exportada nos resultados, o pesquisador pode usá-la para justificar porque determinado dado está errado e qual seria o valor correto. Estes mesmos autores também defendem que o próprio voluntário possa marcar a sua contribuição com uma mensagem de alerta para avisar ao pesquisador do projeto em caso de problemas na medição ou obtenção do dado. Como o tempo do pesquisador é limitado, ele não necessariamente terá tempo de validar todos os dados, e irá usar o seu tempo apenas para fazer uma validação por amostragem para ter uma impressão da qualidade geral. Por isso, a funcionalidade de validação foi implementada para que as contribuições “sinalizadas” com alertas tivessem prioridade.



Figura 6 – Tela de validação



Adicione uma mensagem para o usuário:

Esta contribuição foi rejeitada pois (...)

Rejeitar Aprovar

Figura 7 – Após analisar a contribuição, o pesquisador aceita ou rejeita

Uma outra proposta de GAMBLE et al. (2011) seria de implementar um sistema de pontuação para os voluntários onde a reputação dos usuários seria transferida para a reputação dos dados, isto é, dados produzidos por um usuário com mais pontos seriam mais confiáveis. Isso poderia ser usado também no cenário onde ocorre divergência entre voluntários, fazendo com que um critério de desempate seja a pontuação. Apesar de um sistema de pontuação ter sido implementado com o objetivo de promover gamificação, esta informação não é usada na plataforma para fins de qualidade.

Outras iniciativas de crowd science possuem estratégias diferentes para a funcionalidade de validação de contribuições. Por exemplo, o Projeto Urubu (http://cbee.ufla.br/portal/sistema_urubu). Nele, o voluntário tira foto de animais mortos por atropelamento em estradas e envia para o portal. A foto é registrada e um conjunto de avaliadores especialistas, também voluntários, fazem uma validação da taxonomia do animal morto para identificar se o voluntário autor da foto registrou o animal corretamente. Por fim, os administradores do portal recebem os dados ajustados e fazem uma segunda validação para que, enfim, o dado seja utilizado.

Já no site iNaturalist (<https://www.inaturalist.org>), a validação é feita pela própria comunidade e no formato de consenso. O objetivo do iNaturalist é catalogar organismos (animais, vegetais, etc.) para registrar sua distribuição e ocorrência ao longo do tempo. Para isso, o voluntário tira uma foto e envia para o site junto com uma identificação. Por exemplo, ele pode dizer que trata-se de um inseto. Outros usuários irão analisar a foto e fazer a validação do tipo de ser vivo encontrado, confirmando e adicionando informações mais detalhadas. Se outra pessoa concordar, a informação é confirmada. Agora se outras pessoas classificarem de uma forma divergente, e se esta

nova classificação tiver 2/3 de concordância entre o total de avaliadores, a informação de consenso é alterada.

3.2.2 Sistema de auditoria

GAMBLE et al. (2011) e WIGGINGS et al. (2011) defendem a importância de se implementar um sistema de auditoria para identificar voluntários “ruins” e descartar as suas contribuições. Como o Fast Science divide os projetos em dois grupos, os de Coleta de dados e os de Processamento de dados, só foi possível automatizar a auditoria para os projetos de processamento. O motivo é que, para uma auditoria automatizada, é necessário que a resposta já seja conhecida a priori. Isso não é possível em um projeto de Coleta onde o usuário cria novas informações.

Nos projetos de processamento, a ideia é que o pesquisador irá configurar “tarefas e controle”. Isto é, tarefas que são extremamente simples de serem resolvidas e que o voluntário só irá preencher com dados errados se ele estiver distraído ou não estiver minimamente capacitado para a complexidade do projeto. Neste caso, errando uma tarefa de auditoria, todas as suas contribuições anteriores são marcadas como “suspeitas”, cabendo ao pesquisador eliminá-las se desejar. Estas tarefas de controle aparecem para o voluntário como se fosse uma tarefa comum, na frequência de 10% das vezes que ele acessa o projeto para contribuir.



Figura 8 – Funcionalidade de auditoria

Para projetos de Coleta, pode-se entender que foi implementada uma “auditoria manual” através da funcionalidade de validação de dados que foi explicada na seção anterior. Quando o gestor do projeto valida as contribuições, ele está *auditando* os usuários, já que ao identificar contribuições ruins, as contribuições dos usuários serão marcadas como “suspeitas”.

3.2.3 Sinalizadores de problemas

Uma das abordagens que SHEPPARD et al. (2011) defendem é que o sistema precisa permitir que o voluntário informe quando o mesmo detecta um possível problema. Pode ser uma imagem errada no dataset do pesquisador ou pode ser uma dúvida do voluntário. Outro exemplo: se o usuário faz uma medição em um dia com temperatura extremamente atípica, seja frio, calor, ou umidade, pode ser que o seu instrumento faça uma medida estranha e, nesse caso, caberia ao pesquisador determinar o descarte do dado ou o seu ajuste.

A imagem a seguir mostra como o botão de “sinalizar tarefa” foi criado e, na sequência, como a notificação aparece para o gestor do projeto:

Identificação da cor mais relevante

Cena de Jogo

Se você encontrou algum problema com esta tarefa ou deseja reportar algo, utilize o campo abaixo:

O problema encontrado foi (...)

Salvar

Qual a cor que te chamou mais a atenção?

Amarelo

Caso tenha escolhido "Outra", preencher abaixo

← Início Sinalizar tarefa Avançar →

Figura 9 – Sinalizando problemas durante uma contribuição

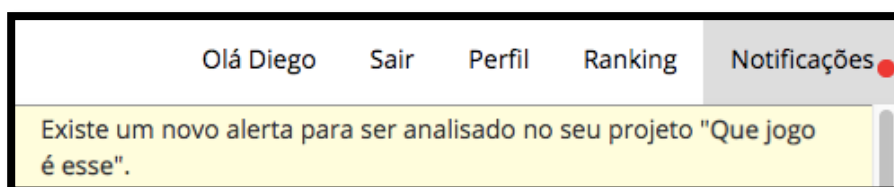


Figura 10 – Mensagem de alerta que o gestor do projeto recebe

Para esta funcionalidade, implementou-se também uma tela para que o voluntário possa acompanhar os alertas criados e as respostas dos mesmos:

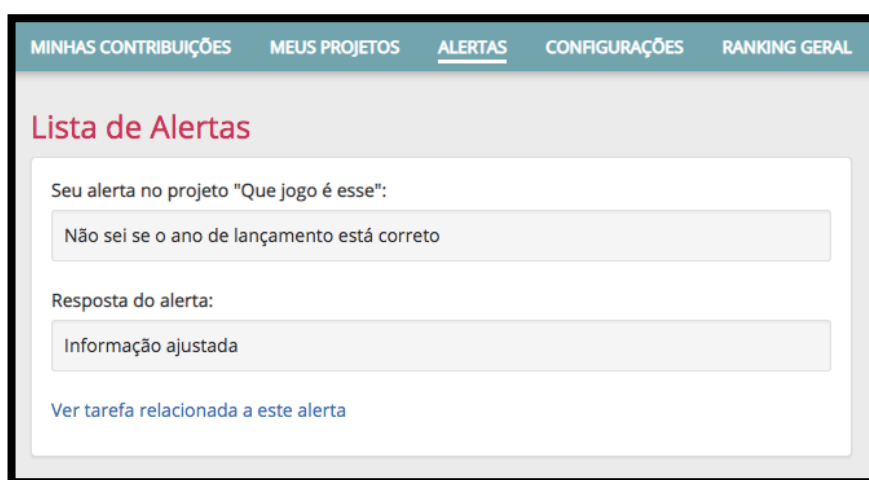


Figura 11 – Tela para controle de alertas criados e as suas respostas

3.2.4 Notificações

O sistema de notificações foi criado porque grande parte das contribuições ruins são originadas por uma pequena porcentagem de colaboradores e acontecem por falta de *feedback* (ALABRI et al., 2010). Com este objetivo, sempre que um gestor rejeita uma contribuição, ele é obrigado a escrever uma mensagem que será enviada para o voluntário através de uma notificação. Da mesma forma, sempre que o voluntário falha em uma auditoria, ele é notificado da falha e uma mensagem informa que ele precisa ter mais atenção.

3.2.5 Melhorias na execução de tarefas

Segundo SHEPPARD et al. (2011), além de permitir que o pesquisador identifique e trate erros, é necessário criar meios para que seja dificultado a criação e proliferação de erros. Para isso, as seguintes funcionalidades foram implementadas:

- **Componente de CPF:** quando o pesquisador cria uma tarefa o sistema disponibiliza uma série de componentes que poderão ser selecionados conforme os objetivos do projeto como, por exemplo: caixa de texto, checkbox, select box, envio de imagens, localização geográfica, entre outros. Um novo componente adicionado foi o de CPF para que, em projetos que o voluntário cadastra um questionário respondido, como foi o projeto “Bebeu Água?”, ter uma máscara de CPF possibilita que apenas dados válidos sejam aceitos. Um outro componente que poderia ser criado seria para máscara de telefone, mas não houve necessidade que justificasse a sua criação.
- **Campos Obrigatórios:** quando o Fast Science foi lançado, não havia sido configurada uma opção para que o pesquisador definisse quais campos das tarefas eram obrigatórios. Por isso, algumas contribuições realizadas estavam “em branco” ou sem informações essenciais para a tarefa ser válida. Adicionar essa opção força que o voluntário tenha mais atenção na hora de executar a tarefa.
- **Barreira de 5 segundos:** observou-se também, ao analisar contribuições antigas, que algumas foram feitas muito rapidamente, onde o voluntário abria uma nova tarefa, marcava qualquer valor disponível em um select box e finalizava a mesma. Para remediar este cenário, adicionou-se uma barreira de 5 segundos onde o voluntário só pode ir para uma nova etapa ou finalizar após aguardar 5 segundos, exigindo que o mesmo, no mínimo, tenha 5 segundos de atenção antes de marcar qualquer informação.

3.2.6 Detector de texto sem sentido

Esta melhoria não veio como sugestão da literatura e sim a partir de uma ferramenta de código aberto encontrada no GitHub (<https://github.com>) chamada de “Gibberish Detector” (<https://github.com/rrenaud/Gibberish-Detector>) desenvolvida por Rob Neuhaus. A ferramenta, originalmente desenvolvida em Python, foi portada para JavaScript para ser executada no frontend em vez de ser executada no backend.

O objetivo da ferramenta é detectar quando um texto sem sentido é digitado. Como exemplos: “asdfasdf” ou “aodjufowje”. Esse tipo de texto ocorre quando o voluntário não quer preencher os dados corretamente e digita caracteres aleatórios apenas para fugir de uma restrição do tipo “campo obrigatório”. Esse tipo de comportamento é mais frequente em projetos que possuem ferramentas de gamificação,

onde o usuário “burla” a ideologia do projeto para inserir qualquer informação sem sentido o mais rápido possível apenas para ganhar pontos.

A ferramenta foi implementada na plataforma para todos os campos de texto. Quando o usuário preenche o campo, antes da informação ser enviada para o servidor, a análise do “gibberish detector” é executada e, quando um texto sem sentido é encontrado, a contribuição não é enviada e uma mensagem de alerta orienta o usuário a revisar o texto digitado.

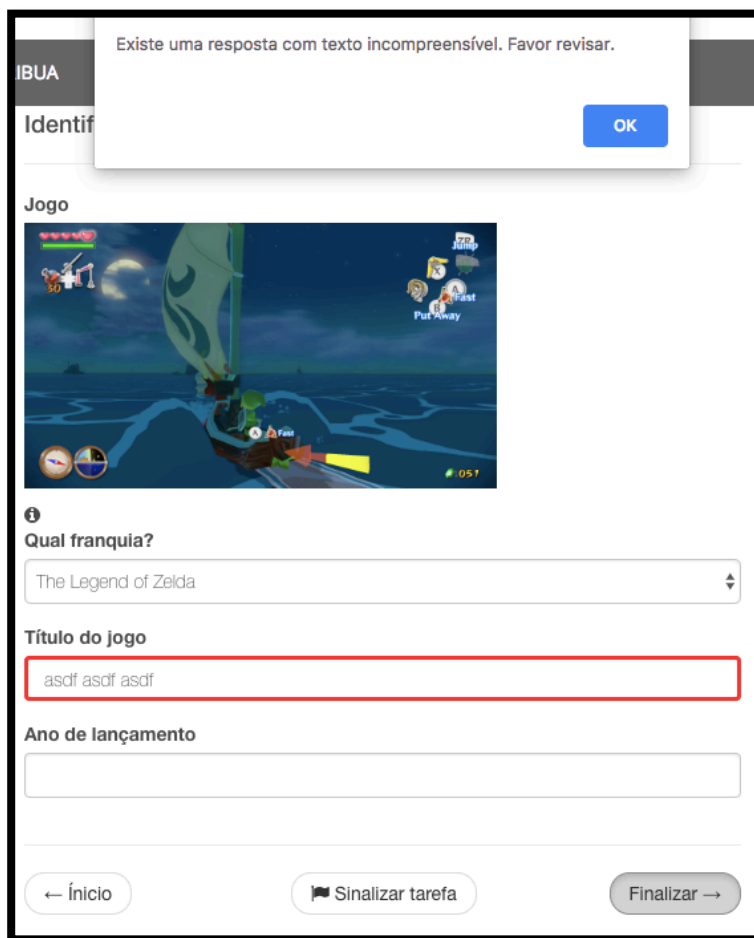


Figura 12 – Alerta de mensagem com texto incompreensível

Essa ferramenta funciona em duas etapas: primeiro, a ferramenta precisa ser treinada recebendo uma informação de texto bem longa (como um livro) para que ela aprenda a frequência que determinados caracteres aparecem ao lado de outros caracteres. A seguir, textos válidos e inválidos são fornecidos para que ela seja calibrada e identifique quais são os limiares que devem ser aceitos e quais devem ser rejeitados.

Esse treinamento com um texto longo serve para gerar uma análise estatística de uma Cadeia de Markov, que trata-se de um processo estocástico onde para determinar a

probabilidade do próximo estado, a única informação necessária é conhecer o estado atual. Por exemplo, digamos que analisando o texto de um livro, sempre que uma letra “a” aparece em uma palavra, a chance de outra letra “a” aparecer em seguida seja de 0,001% e de uma letra “b” aparecer seja de 2%. Logo, após realizar este levantamento estatístico, se um novo texto for analisado e o mesmo tiver duas letras “a” em sequência, a ferramenta pode julgar esse novo texto como “inválido”, já que a ocorrência é tão inesperada que provavelmente é inválida. Porém, se existem duas letras “a” em sequência, mas diversas outras letras em sequências consideradas como “normais”, o somatório de probabilidades alcançará o valor alto e a frase como um todo será aceita. Assim, erros de digitação não é um problema.

Para a ferramenta saber qual é o limiar de probabilidade que deve aceitar e qual que deve rejeitar, ela precisa em um segundo passo ser treinada com uma lista de textos definidos como válidos e inválidos. Um exemplo de texto válido utilizado foi “esta foto representa um pássaro” e um texto inválido foi “ifoqwefk asdlffad”.

Para o treinamento com texto longo, a ideia inicial era usar algum livro famoso da literatura brasileira. O problema é que a maioria das grandes obras está protegida pelas leis autorais e não podem ser usadas livremente, pois a lei brasileira determina que os direitos autorais valham por até 70 anos após a morte do autor. Um dos autores com livros em domínio público é o Machado de Assis, mas após realizar o treinamento da ferramenta usando os livros “Dom Casmurro” e “Memórias Póstumas de Brás Cubas”, verificou-se que as palavras e estruturas de frase utilizadas hoje em dia são muito diferentes das utilizadas pelo o autor na época e alguns falso-positivos, textos válidos marcados como inválidos, ocorreram. Por isso, utilizou-se para o treinamento a monografia escrita pelo autor como trabalho de conclusão de curso: “Monitoração de Processos Industriais Utilizando o Protocolo OPC UA e um Aplicativo para Smartphones Android”, UFRJ, Escola Politécnica, Departamento de Eletrônica, 2011.

O resultado obtido neste treinamento ficou bem melhor do que o anterior, inclusive para palavras em inglês, visto que a monografia tratava de tecnologias que são citadas com os nomes originais em inglês. Porém, apesar da calibração para palavras ter ficado boa, verificou-se que para nomes próprios ela ainda gerava falso-positivos. Esse problema foi resolvido adicionando ao treinamento a lista dos 5570 municípios do Brasil, usando como fonte da informação o site do IBGE (<https://ibge.gov.br>).

Além dessa ferramenta “gibberish detector”, outras possibilidades haviam sido estudadas, mas não se mostraram tão eficazes. Estas foram:

- **Dicionário de palavras:** a estratégia seria carregar todas as palavras do dicionário português e inglês e, caso uma frase não usasse uma dessas palavras, ela poderia ser marcada como inválida. Essa estratégia mostrou-se claramente ineficaz porque as pessoas costumam cometer erros de digitação ou usar nomes próprios que não estão mapeados.
- **Entropia de Shannon:** na teoria da informação, entropia é definida como uma medida de incerteza associada a um determinado estado, sendo a Entropia de Shannon uma quantificação dessa incerteza. Esse algoritmo é utilizado para medir a entropia de palavras e, constata-se que, em geral, palavras de baixa entropia como “asdfasdf”, “123123123” e “qqqqq”, podem ser consideradas inválidas. O problema é que esta técnica acaba gerando muitos falso-negativos e não possui um resultado muito bom.
- **Análise do desvio padrão:** outra possibilidade estudada foi analisar um texto longo e medir o desvio padrão de três medidas: frequência de caracteres únicos em cada palavra, frequência de vogais em cada palavra e quantidade média de letras por palavra. A partir destes valores, comparar com o texto sendo analisado e avaliar se o mesmo é válido ou inválido. Esse algoritmo não foi testado neste trabalho, mas segue como uma possibilidade.

3.2.7 Estatísticas de contribuições

No artigo escrito por WIGGINGS et al. (2011), ele recomenda 18 mecanismos para controle de qualidade. Um deles diz que é necessário prover ao moderador da plataforma alguma forma para avaliação estatística para identificar e corrigir erros espúrios. Esse mecanismo não foi implementado. Cabe ao pesquisador de cada projeto, ao exportar os resultados das contribuições, analisar os dados para procurar por valores fora do padrão. Porém, o que foi implementado no sistema foi uma tela para exibir uma estatística de quantas contribuições foram feitas ao longo do tempo e de quanto cada usuário contribuiu. O valor agregado nestes gráficos é permitir que o pesquisador analise:

1. **Contribuições x Tempo:** muitas contribuições em datas isoladas pode representar sazonalidade e, em alguns casos, pode significar que os resultados não são muito confiáveis por não estarem representando uma amostra fiel do comportamento de um fenômeno ao longo do tempo. Por exemplo, se muitas medidas forem feitas em dias quentes, pode-se tomar

conclusões que não se aplicam a dias chuvosos. Além disso, esse gráfico permite mostrar a eficácia de campanhas, por exemplo, quando se divulga um projeto em uma mídia social, pode-se mensurar o alcance que foi obtido.

2. **Contribuições x Voluntário:** muitas contribuições feitas por um único voluntário deixa a amostra enviesada. Nesse caso, o pesquisador pode identificar o problema e aumentar a divulgação de seu projeto para tentar promover a diversidade de contribuições. Outra finalidade é poder identificar o percentual de pessoas que contribuem em comparação com a quantidade de usuários cadastrados.

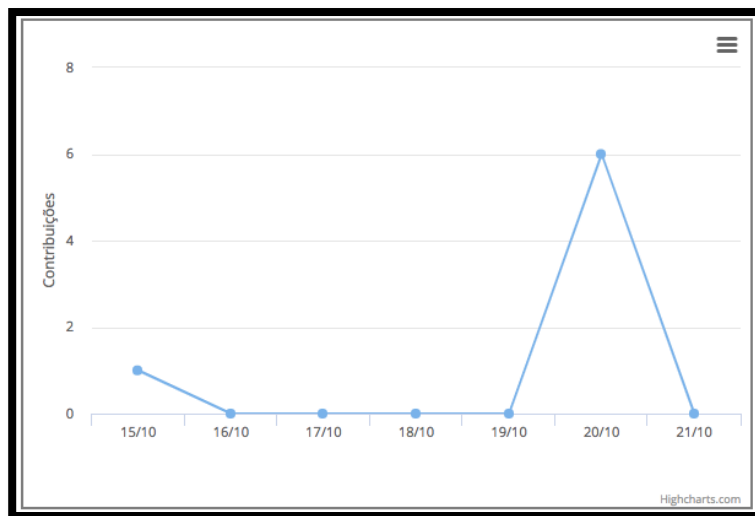


Figura 13 – Contribuições no tempo

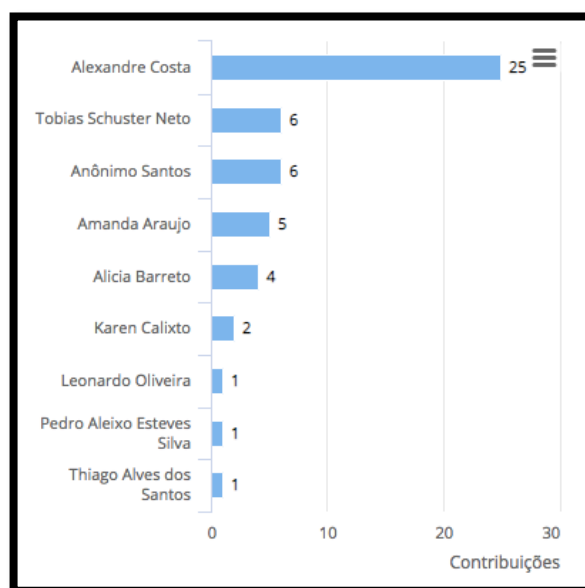


Figura 14 – Contribuições por voluntário

3.3 Gamificação

A partir das sugestões de ferramentas encontradas na literatura, as seguintes funcionalidades foram implementadas na plataforma Fast Science:

- Ranqueamento
- Recompensas
- Autopromoção
- Barra de progresso

3.3.1 Ranqueamento

Listar usuários ordenando do melhor colocado para o pior é uma proposta muito comum em iniciativas com gamificação. Rankings estimulam a competitividade e trazem satisfação, pelo menos para os melhores colocados. Conforme WERBACH e HUNTER (2012), da mesma maneira que o ranking consegue estimular um competidor, ele desestimula quem está mal posicionado ou quem acabou de entrar no “jogo”. Para tratar esse problema, eles sugerem a criação de vários tipos de ranking para que, mesmo que a pessoa não consiga ser a melhor no geral, para que ela tenha a chance em se destacar em algum contexto.

No Fast Science, além do ranking geral da plataforma, implementou-se também rankings por cada projeto. E, para cada ranking, divisões entre “Total”, “Mensal” e “Semanal”. Como forma de estimular os novos competidores em relação aos “veteranos”, uma colocação entre os 3 melhores no ranking semanal ou mensal dá pontos extras no ranking geral. Logo, uma pessoa que acabou de começar, se ela for a melhor, pelo menos no ranking semanal, ela já conseguirá ter um bônus para, pelo menos, se aproximar dos melhores colocados do geral. Uma observação é que se implementou o conceito de “bônus temporário”. Ser o melhor da semana dá 10 pontos no ranking total, mas ao perder a colocação de melhor da semana, esses 10 pontos desaparecem. A proposta é que, se os bônus fossem permanentes, um usuário antigo iria acumulá-lo diversas vezes e um usuário novato teria ainda mais dificuldade de alcançá-lo, ficando assim, ainda mais desestimulado.

Ranking

Total

Mensal

Semanal

Total

Posição	Nome	Pontos	Medalhas
1	Alexandre Costa Novo Usuário	222	3 ● 3 ● 6 ●
2	Stefanie Luz	76	1 ● 2 ● 2 ●
3	Anônimo Santos Novo Usuário	64	5 ●
4	Steffany Cardoso	55	1 ● 2 ● 2 ●
5	Renata Goulart	55	1 ● 2 ● 2 ●
6	Franciele Ribeiro Reis	43	2 ● 2 ●
7	Lidiane Bauwalet	40	1 ● 2 ●
8	carla stevenson	37	2 ● 2 ●
9	Amanda Araujo Novo Usuário	36	4 ●
10	karolaine pinto	35	1 ● 2 ●
11	andressa moreira	33	1 ● 2 ●
12	Leonardo Oliveira Novo Usuário	33	9 ●

Figura 15 – Ranking implementado na plataforma Fast Science

3.3.2 Recompensas

Sempre que um usuário contribui com uma atividade de um projeto na plataforma Fast Science, ele recebe 1 ponto. Quando o cientista avalia a contribuição como válida, o usuário ganha mais 2 pontos por ter feito uma boa contribuição, mas perde 10 pontos se o cientista a rejeitar. Este sistema de pontuação é utilizado para o ranqueamento dos voluntários.

Como existem tarefas com complexidades bem distintas, às vezes não é justo remunerar ambas da mesma forma. Uma estratégia seria permitir que o gestor do projeto informasse o grau de complexidade de sua tarefa e quantos pontos ela remuneraria. Para simplificar, a estratégia utilizada foi: se nos componentes da tarefa existir um componente para “upload de imagem”, esta tarefa valerá 10 vezes mais do que uma tarefa comum pela necessidade de o usuário ir “à campo” para completar a mesma.

Medalhas virtuais de ouro, prata e bronze também são entregues conforme o usuário vai progredindo nos projetos, sendo uma medalha de bronze na primeira contribuição, prata na décima e ouro na vigésima quinta. E se o usuário contribuir muitas vezes em uma determinada área da ciência, ele também ganha medalhas adicionais por isso. Segundo PEREIRA (2014), o mecanismo de notificação é importante, por isso o usuário recebe uma mensagem quando ganha uma nova medalha:



Figura 16 – Funcionalidade de notificações de recompensas

EVELEIGH et al (2013) defendem que as recompensas devem seguir a temática da plataforma para aumentar a imersão. Com essa proposta, foram implementadas duas funcionalidades no perfil: a primeira é o plano de fundo que muda conforme área da ciência que o usuário mais colaborou e a segunda são “insígnias” destacando os “níveis” em cada área. Na imagem a seguir, o plano de fundo de uma floresta destaca que o usuário colaborou mais em projetos da área “Ciências Naturais”. Se ele tivesse contribuído mais em “Ciências Exatas”, a imagem seria de uma galáxia e, para “Ciências Humanas”, a imagem seria de uma biblioteca.



Figura 17 – Plano de fundo e insígnias temáticas

Ao lado do avatar (foto de perfil) são exibidas as insígnias que destacam qual nível o usuário está em cada uma das 3 grandes áreas. Os níveis criados foram: “aprendiz”, “experiente”, “sênior” e “sábio”.

3.3.3 Autopromoção

Seguindo o exemplo do site Stack Overflow (<https://stackoverflow.com>) para usar a autopromoção como ferramenta de motivação, criou-se uma página de perfil público que é acessada por um link disponível nos rankings. Nessa página, o usuário pode adicionar uma foto, uma mensagem e links para promover as suas redes sociais, conforme pode ser visto na figura anterior.

Um voluntário de um projeto de crowd science geralmente é um amador naquela área de pesquisa e, como tal, costuma usar o tempo livre se dedicando ao assunto. Muitas vezes ele possui blogs, coleções de fotos ou projetos relacionados ao tema e dar a ele um espaço para promovê-los é uma forma de recompensação pelo trabalho voluntário. Outra observação é que, quantos mais pontos na plataforma, melhor é a posição do usuário no ranking e maior visibilidade ele vai ter. O que sugere que uma pessoa interessada em autopromoção irá contribuir mais.

3.3.4 Barra de progresso

Segundo PEREIRA (2014), é necessário fornecer ao usuário algum mecanismo de acompanhamento e *feedback* para que ele saiba quanto falta para atingir um próximo nível. A partir disso, criou-se uma barra de progresso para o usuário saber quanto falta para a próxima medalha:



Figura 18 – Indicador mostra progresso para a terceira medalha

3.4 Implementação e Arquitetura

O sistema Fast Science foi implementado utilizando os frameworks AngularJS e React no frontend. O motivo de utilizar dois frameworks distintos é porque, usando o conceito de microserviços, cada serviço pode ser individualizado e ficar independente de linguagem, sendo que compete ao desenvolvedor definir e implementar o serviço com as ferramentas mais apropriadas para cada caso. Por isso, o sistema foi iniciado com AngularJS, mas durante o desenvolvimento do módulo do Workflow de tarefas, optou-se por utilizar bibliotecas em React que possuíam um resultado visual melhor.

Para o backend, utilizou-se a linguagem Python e a framework web Django para implementar serviços REST acessados pelo frontend e utilizou-se a ORM (Object-Relational Mapping) do Django para fazer consultas ao banco de dados a partir do código Python, sem a necessidade de escrever queries SQL. A vantagem do ORM foi de simplificar a criação de consultas, a legibilidade do código e o seu reuso em diferentes partes do sistema.

Para o banco de dados, o PostgreSQL foi utilizado por ser uma solução eficiente e open source.

A imagem abaixo mostra a visão das camadas da arquitetura, mostrando os módulos “user”, “project” e “workflow” do frontend e o módulo único no backend chamado de “fast_science_app” que possui os componentes “user”, “project” e “workflow” para atender às requisições do frontend.

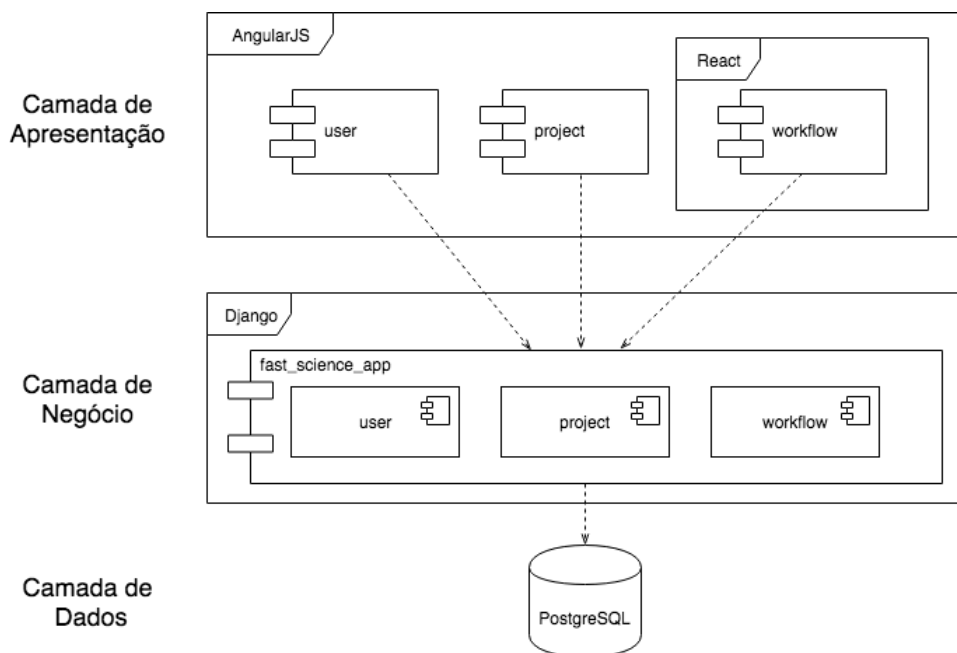


Figura 19 – Visão das camadas da arquitetura e os seus módulos

Na figura abaixo, mostra-se uma visão mais detalhada do código implementado com a framework AngularJS, onde o ator Usuário acessa a um módulo do sistema através de uma requisição para uma View. A partir desta requisição, o Controller irá tratar a lógica do sistema, requisitando, se necessário, acesso ao backend através do componente Services, para popular o Model que será exibido na View. Esta arquitetura segue o conceito definido pelo MVC (Model-View-Controller).

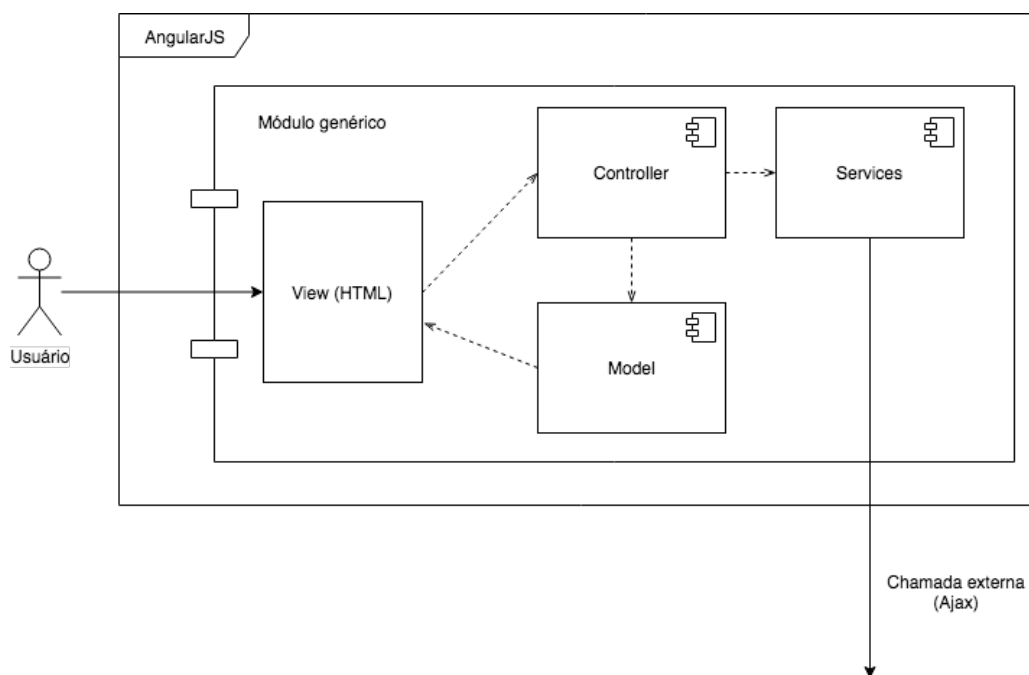


Figura 20 – Diagrama de componentes para os módulos do AngularJS

Para o código que utilizou o React, a framework Redux foi utilizada para implementar uma arquitetura onde as classes de View acionam as Actions que possuem a lógica de acesso do backend e são responsáveis por popular as Stores que são os conjuntos de dados que são usados pelas Views para exibição nas telas.

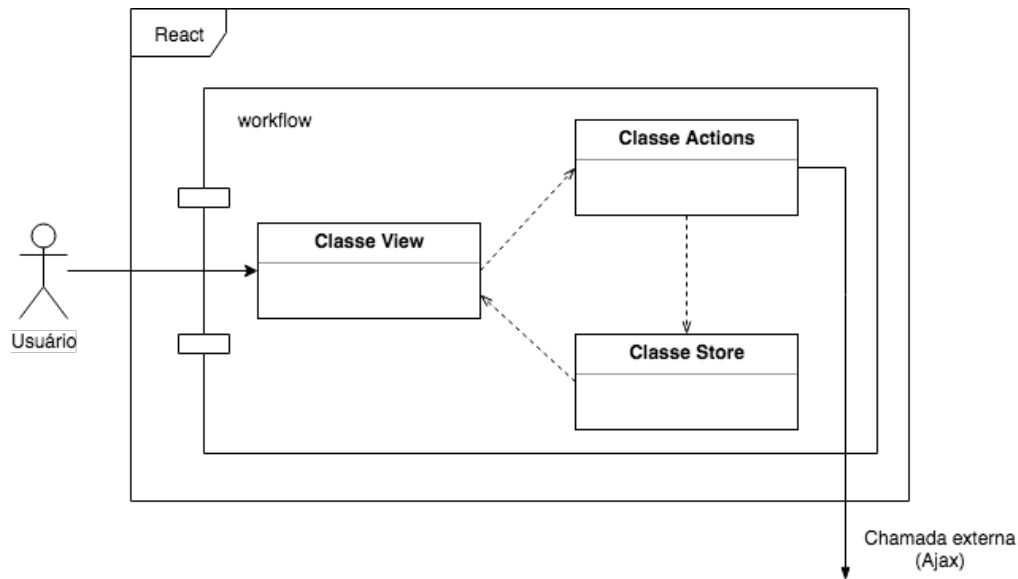


Figura 21 – Diagrama de classes da implementação do React

Com esta arquitetura, o sistema foi implantado em produção utilizando uma instância Linux na nuvem da AWS (Amazon Web Services), onde o Apache foi utilizado como servidor web.

3.5 Modelo de dados

Os dados do sistema foram modelados com o uso de 42 tabelas. Para simplificar e facilitar o entendimento, a imagem abaixo mostra apenas 12 tabelas exibindo apenas com as chaves primárias e estrangeiras.

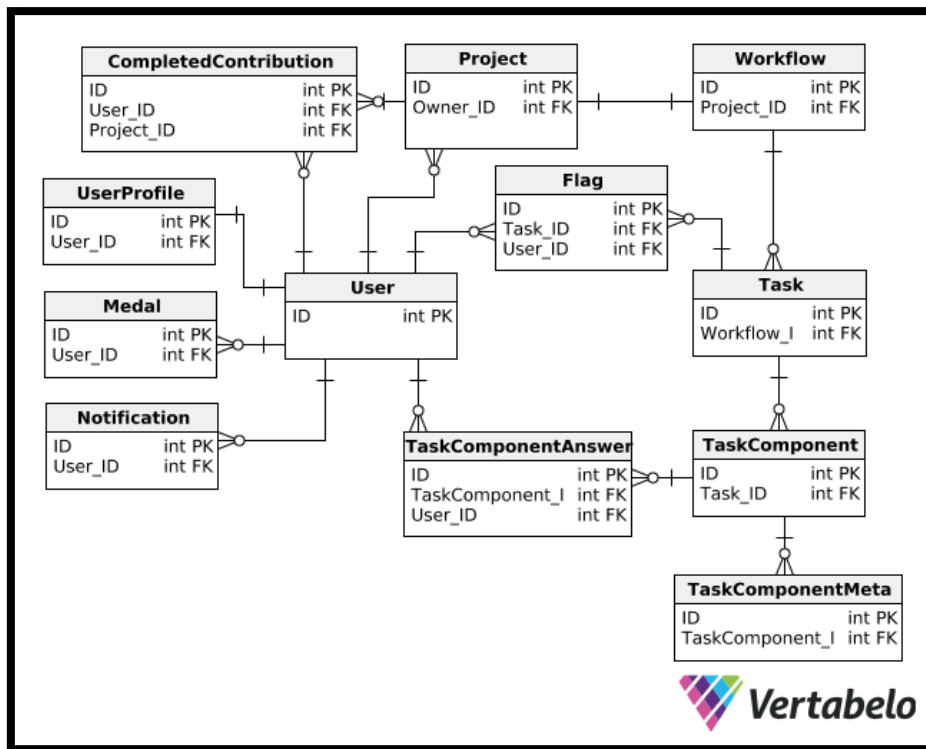


Figura 22 – Modelo simplificado do banco de dados

Analisando o modelo a partir da tabela de usuários (“User”), é possível ver que esta se relaciona com as tabelas à esquerda de medalhas (“Medal”) e notificações (“Notification”) na relação “um para N”, onde um usuário pode possuir diversas medalhas e diversas notificações. Em relação ao perfil, estes possuem cardinalidade “um para um” em relação ao perfil.

Observando a relação do usuário com o projeto, entende-se que um projeto possui apenas um usuário como proprietário (“owner”) e um projeto possui apenas um workflow de tarefas. Em relação ao workflow, observa-se que o mesmo é composto de diversas tarefas, onde cada tarefa possui diversos componentes e pode ter diversas sinalizações (“Flags”) criadas por usuários. Entre os componentes, cada um possui múltiplos metadados e múltiplas respostas de usuários distintos.

Por último, a tabela “CompletedContribution” possui um relacionamento “um para N” com usuários e projetos, onde um usuário pode ter várias contribuições em vários projetos. Esta tabela é redundante, já que a “TaskComponentAnswer” já armazena dados de respostas nos projetos, mas esta redundância ajudou a simplificar algumas consultas e a trazer melhor desempenho.

Capítulo 4 – Experimento

4.1 Objetivos

O experimento foi idealizado com o objetivo de validar se as técnicas utilizadas nesta pesquisa podem realmente contribuir com iniciativas de crowd science e se estas realmente provocaram os ganhos esperados.

Partindo desta visão, os objetivos do experimento são:

- Avaliar se as ferramentas de gamificação aplicadas conseguiram gerar mais contribuições.
- Avaliar se a gamificação implementada afetou a qualidade dos dados.
- Identificar quais ferramentas tiveram mais êxito e quais foram subutilizadas.

A metodologia utilizada (vide [seção 1.6](#)) inclui a quarta e quinta etapas do método de design science proposto por PEFFERS et al. (2007). A quarta etapa é a “Demonstração”, ou seja, corresponde ao uso dos artefatos implementados na plataforma Fast Science para que os mesmos possam ser validados ou refutados na quinta etapa de “Avaliação”. Nesta etapa, dados qualitativos foram coletados e os resultados obtidos foram então comparados com os requisitos definidos na segunda etapa do método “Definição dos Resultados Esperados”.

4.2 Metodologia

4.2.1 Planejamento

Para a execução do experimento, optou-se pelo uso do método de teste A/B. Um teste A/B é usado quando existem 2 opções para se implementar uma determinada funcionalidade em um sistema e deseja-se identificar qual delas é a melhor a partir da avaliação do comportamento dos usuários (KOHAVI e LONGBOTHAM, 2015). Por exemplo, a versão A de um sistema poderia ter um botão de “comprar” na cor cinza e a versão B poderia ter o mesmo botão destacado na cor vermelha. Neste cenário, um grupo de pessoas veria o sistema na sua versão A e outro grupo na versão B. Após um determinado tempo, avalia-se a quantidade de compras que foram feitas no sistema a partir de cada versão e determina-se assim qual cor ajudou a gerar mais vendas e esta será a escolhida para a versão final.

No teste A/B, em geral, testa-se apenas uma variável por vez, mas também pode-se testar um conjunto de funcionalidades desde que estas estejam agrupadas em um mesmo contexto. No caso, o contexto é gamificação. Portanto, o grupo A testado visualizou o sistema Fast Science com todas as modificações de gamificação desenvolvidas neste trabalho enquanto que o grupo B visualizou o sistema onde as funcionalidades de gamificação estavam desativadas ou escondidas.

Para ambos os grupos, a versão já contemplava as melhorias para qualidade de dados. Para avaliar estas melhorias, utilizou-se como referência a qualidade mensurada de contribuições antigas na plataforma em comparação com a qualidade medida das contribuições realizadas neste experimento.

4.2.2 Preparação

Antes da realização deste experimento, a plataforma Fast Science possuía 14 projetos cadastrados. Entretanto, a grande maioria dos projetos existentes são destinados a coleta de dados ambientais como os abaixo exemplificados:

- **Monitoramento de lixo em praias:** mapear a composição do lixo em praias.
- **Insetos do Brasil:** mapear insetos e espécies a partir de fotos obtidas em diferentes áreas geográficas.
- **Meio Ambiente:** questionário para mapear padrões de comportamento das pessoas.
- **Erosão Costeira:** dada uma foto de uma região costeira, identificar se houve erosão e marcar os elementos apropriados.

Levando-se em consideração que o recrutamento dos participantes teve como público-alvo alunos de graduação em engenharia e, sabendo-se que motivação em contribuir depende de identificação pessoal com o tema do projeto, optou-se por criar três novos projetos de processamento de dados a fim de ampliar o leque de escolhas em função de preferências pessoais.

- **Jogos e Emoções:** dada uma cena de um jogo de videogame, dizer quais emoções são observadas, quais cores chamam mais atenção e quais elementos da imagem contribuem para estes sentimentos.
- **Que jogo é esse?:** dada uma cena de um jogo de videogame, identificar de qual franquia o jogo faz parte, qual é o título do mesmo e o ano de lançamento.

- **Fact Checking:** encontrar notícias falsas (“fake news”) e as notícias que comprovam que a notícia original é falsa.

Cabe ressaltar, que a criação do projeto “Jogos e Emoções” teve como base em uma necessidade real científica da pesquisadora Airine Carmo e as contribuições realizadas neste projeto devem ser utilizadas futuramente na sua dissertação de mestrado.

4.2.3 Recrutamento e execução

O experimento foi realizado com a colaboração de alunos pertencentes à 2 turmas de graduação do Centro Federal de Educação Tecnológica Celso Suckow da Fonseca (CEFET) de Itaguaí, sendo uma turma de graduação em engenharia mecânica e outra de engenharia de produção. O experimento foi conduzido com a ajuda do professor Marcio Antelio que disponibilizou alguns minutos de uma de suas aulas para que fosse possível apresentar o tema de crowd science, explicar o experimento e a importância da participação do cidadão visando motivá-los a contribuir com os projetos de crowd science disponíveis na plataforma Fast Science.

Foi solicitado aos participantes que eles acessassem a plataforma Fast Science, criassem uma conta de usuário e selecionassem entre os 17 projetos existentes, as iniciativas que eles considerassem mais interessantes para contribuir. Para exemplificar o funcionamento, o sistema foi apresentado em sala de aula junto com uma demonstração de como executar as tarefas. Uma lista foi disponibilizada para que os alunos pudessem preencher os seus dados pessoais. Neste momento, foram coletados o nome, e-mail, gênero e idade.

Com o objetivo de realizar o teste A/B, 38 alunos foram convidados, tendo sido distribuídos de forma aleatória com 19 convites para acessarem o ambiente que representava o grupo A e 19 para o grupo B. Entretanto, apenas 24 participaram efetivamente durante as 2 semanas de experimento, tendo sido 12 participantes no ambiente A e 12 no ambiente B. Para motivar as pessoas a participarem e a usarem o sistema, 5 e-mails foram enviados para cada aluno durante o acompanhamento do experimento. Para o grupo sem gamificação, foi dito que quem contribuísse na plataforma participaria de um sorteio para ganhar brindes (bombons) e no grupo com gamificação, foi dito que os brindes (bombons) seriam dados aos melhores colocados.

Uma das questões levantadas é se presentear os voluntários é algo que invalida ou não o experimento. O argumento seria que os bombons representam uma motivação

extrínseca, isto é, um fator externo que ofusca as motivações intrínsecas, aquelas que vêm de dentro da pessoa e que estão aliadas aos interesses pessoais dela. O contra-argumento é que os bombons são de baixo valor e foram dados no final do experimento apenas como forma de agradecimento e para gerar empatia, da mesma maneira que bombons foram dados a todos os alunos no dia que o experimento foi apresentado. Se, em vez de oferecer bombons, tivesse sido oferecida uma pontuação extra na disciplina, aí sim os alunos teriam tido uma motivação extrínseca muito forte para contribuir, principalmente os alunos com notas mais baixas, e o experimento teria se desvirtuado.

4.2.4 Perfil dos Participantes

O perfil dos voluntários é formado por jovens adultos, com idade entre 17 e 25 anos, de maioria masculina, e com formação de ensino superior incompleto, sendo estudantes de graduação em engenharia. Segue abaixo a distribuição dos 24 voluntários que efetivamente contribuíram realizando tarefas na plataforma de acordo com o gênero e a idade.

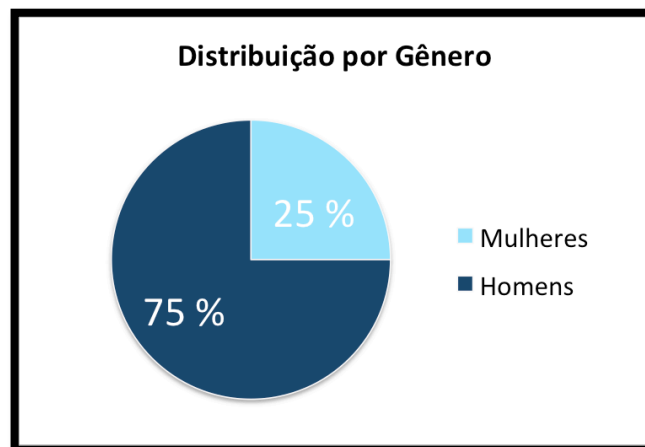


Figura 23 – Distribuição de voluntários por gênero

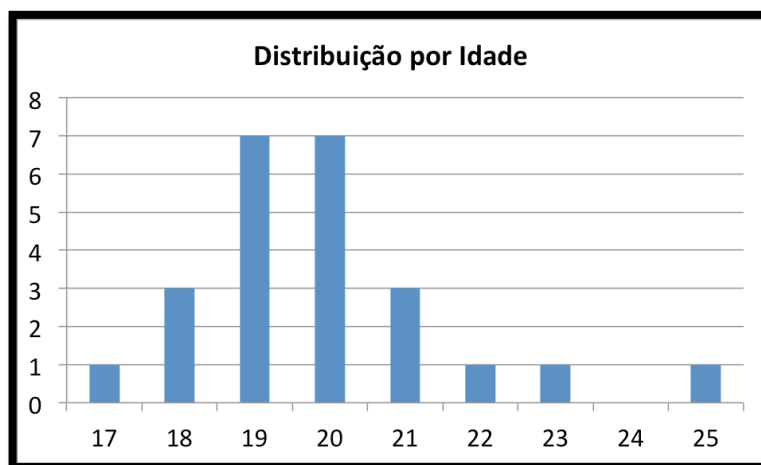


Figura 24 – Distribuição de voluntários por idade

4.2.5 Avaliação

No final das duas semanas, os alunos que contribuíram foram convidados a preencher um formulário de avaliação como forma de feedback sobre o experimento. Neste formulário, 8 afirmações foram apresentadas e os participantes responderam utilizando uma escala Likert de 1 a 5, variando de: “Discordo totalmente”, “Discordo parcialmente”, “Não discordo nem concordo”, “Concordo parcialmente” e “Concordo totalmente”.

As afirmações que foram apresentadas seguem abaixo.

Tabela 3 – Lista das afirmativas elaboradas para avaliar as funcionalidades

Grupo com gamificação	Grupo sem Gamificação
1) A existência de um ranking geral destacando os usuários que mais contribuíram, aumentou a minha motivação em participar	1) Eu acredito que a existência de um ranking geral, destacando os usuários que mais contribuíram, aumentaria a minha motivação em participar
2) Eu acredito que colaborei mais por existir um ranking e querer ficar mais bem colocado(a)	2) Eu colaboraria mais se existisse um ranking porque eu ia querer ficar bem colocado(a)
3) Receber pontos e medalhas aumentou a minha satisfação pessoal em contribuir	3) Receber pontos e medalhas aumentariam a minha satisfação pessoal em contribuir
4) A presença da barra de progresso, indicando uma nova medalha, aumentou a minha motivação para contribuir mais	4) A presença de uma barra de progresso, indicando uma nova medalha, aumentaria a minha motivação para contribuir mais
5) Ao ver a lista dos usuários no ranking, eu tive interesse em acessar o perfil dos participantes	5) Eu gostaria de visualizar a lista de participantes para acessar o perfil deles
6) O perfil público ajuda a conhecer os participantes e a acessar as suas redes sociais	6) Um perfil público ajudaria a conhecer os participantes e a acessar as suas redes sociais
7) Receber notificações sobre a minha participação é um retorno importante para mim	7) Receber notificações sobre a minha participação seria um retorno importante para mim

8) Ter a possibilidade de "Sinalizar a tarefa" e fazer um comentário ajuda a criar um canal de comunicação importante entre o participante e o pesquisador gestor do projeto

Além destas afirmações, uma pergunta opcional “Deseja fazer algum comentário?” foi incluída em ambos os questionários bem como solicitado que os participantes concordassem que estavam cientes que os dados seriam divulgados de forma agregada e anônima (Anexos 1 e 2). Os resultados obtidos estão descritos a seguir.

4.3 Resultados para Gamificação

4.3.1 Quantidade de tarefas executadas

Após 2 semanas de experimento, 12 pessoas do grupo com gamificação executaram 197 tarefas e 12 pessoas do grupo sem gamificação executaram 53 tarefas. Considerando que obteve-se quase 4 vezes mais contribuições utilizando as técnicas de gamificação em comparação ao grupo sem gamificação, os resultados ajudam a responder à Questão 2 da pesquisa pois evidenciam que o uso de técnicas de gamificação contribuiu para o aumento na quantidade de contribuições.

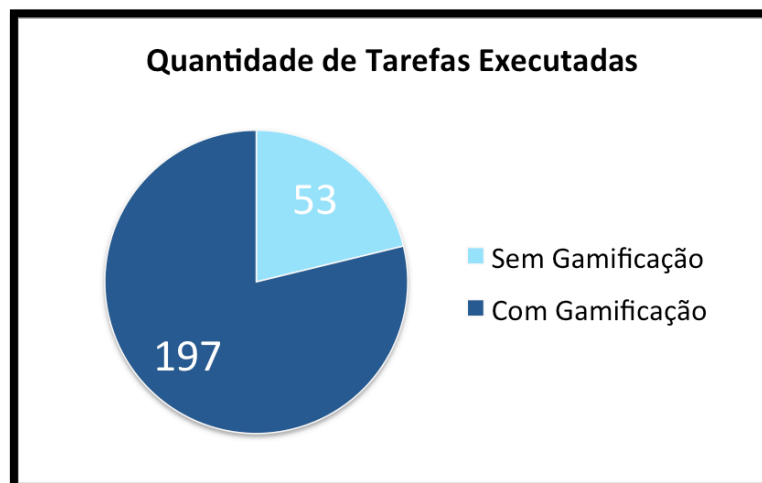


Figura 25 – Comparativo de execução de tarefas

4.3.2 Formulário de avaliação

Após o término do período de contribuições, um formulário de feedback foi enviado para os alunos e, neste, 7 pessoas de cada grupo responderam, totalizando 14 respostas dos 24 participantes. Cada grupo recebeu um conjunto de 7 afirmativas listadas na Tabela 3 e que foram avaliadas utilizando-se uma escala Likert de 5 pontos.

O resultado encontra-se na tabela abaixo, considerando-se que “Discordo totalmente” valia 1 ponto e “Concordo totalmente” valia 5 pontos. Neste caso, uma pontuação de valor 3 significa “Não discordo nem concordo”.

Tabela 4 – Resultados dos formulários de feedback

Tópicos avaliados em cada afirmativa	Média com gamificação	Média sem gamificação
1) Ranking gera motivação	4,6	2,9
2) Ranking gera mais contribuições	3,6	2,7
3) Pontos e medalhas geram satisfação	4,1	3,1
4) Barra de progresso de medalhas gera motivação	3,7	4,0
5) É interessante acessar o perfil dos participantes	2,4	3,0
6) É interessante encontrar as redes sociais dos participantes pelo perfil	2,6	3,0
7) Notificações sobre a participação são importantes	3,6	4,1

Além da tabela de médias, os gráficos abaixo ajudam a analisar os resultados obtidos. Os gráficos representam diagramas de caixa (*boxplots*), onde o ponto central marca o valor médio, o retângulo marca o erro padrão e os limites marcam o desvio padrão de cada resposta. Na legenda horizontal, “AF” significa “Afirmiação”, representando qual afirmiação foi avaliada.

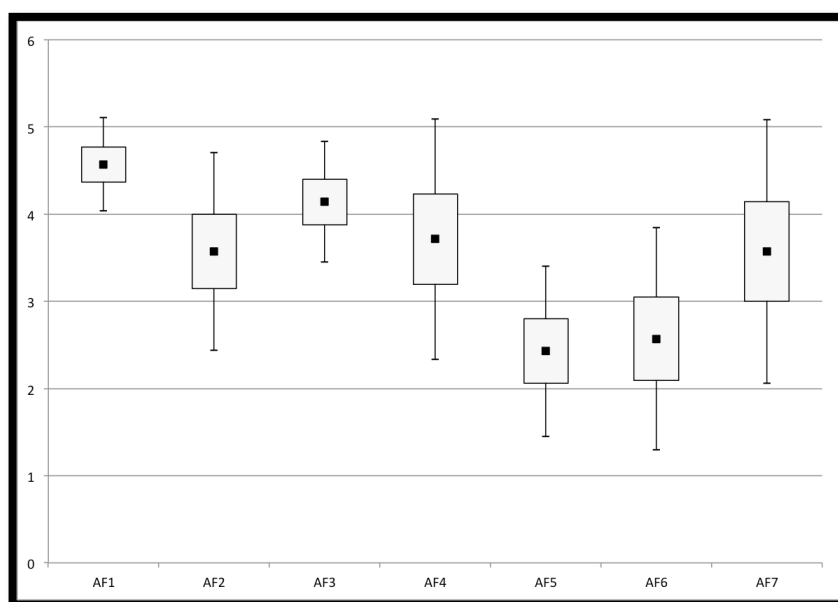


Figura 26 – Resultados do formulário com gamificação

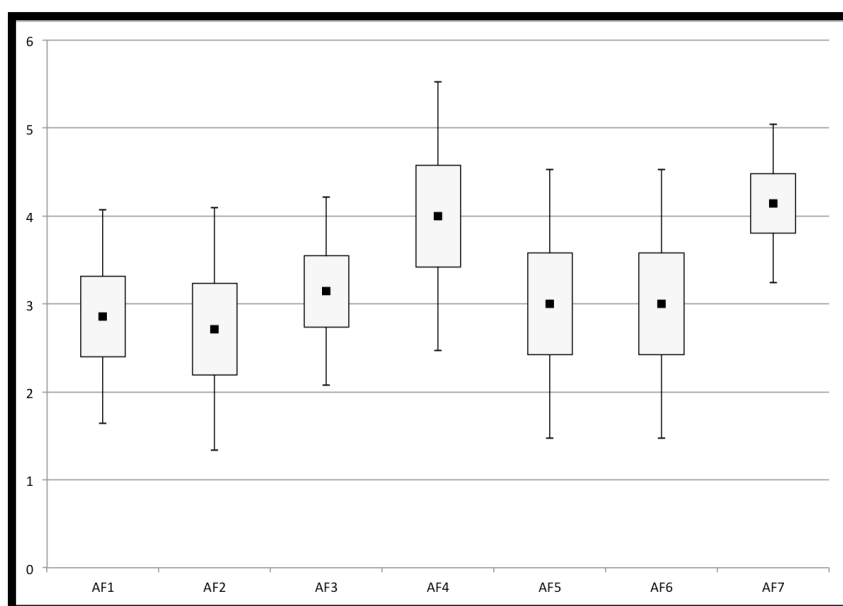


Figura 27 – Resultados do formulário sem gamificação

Para cada requisito implementado, buscou-se elaborar uma afirmativa que auxiliasse no entendimento da sua utilidade através do grau de importância dado pelo participante. Esse conjunto de respostas, ao final, ajudarão a responder as duas grandes questões de pesquisa desta dissertação, **QP1**: A gamificação usada ajuda a aumentar a motivação dos voluntários? E **QP2**: A gamificação usada ajuda a aumentar a quantidade de contribuições?

O primeiro conjunto de afirmativas buscou avaliar se a existência de um ranking geral, destacando os usuários que mais contribuíram, aumentaria a motivação individual em participar e, conseqüentemente, aumentaria a quantidade de contribuições. Os resultados mostraram que todos os participantes que tiveram a experiência de ter acesso ao ranking responderam que concordam parcialmente ou concordam totalmente. Ao contrário, os participantes que contribuíram utilizando a plataforma sem o ranking foram indiferentes, revelando uma tendência neutra sobre se a existência de um ranking iria ou não motivá-los (média 2,9).

Já a segunda afirmação, sobre se o ranking gera ou não mais contribuições, buscou avaliar se a existência do ranking e o desejo de querer ficar mais bem colocado motiva a pessoa a colaborar mais e, por consequência, aumenta o número de contribuições. Novamente, os participantes que não tiveram a experiência de ter um ranking revelaram uma opinião indiferente (média 2,7). Ao contrário, um número maior de participantes que utilizaram o ambiente com gamificação e tiveram a experiência enriquecida pela presença do ranking mostraram uma tendência maior a concordar. Na

prática, os resultados quantitativos revelaram que esse grupo teve 4 vezes mais contribuições, indicando que o ranking influencia de forma positiva na motivação e conseqüentemente, no aumento da quantidade de contribuições (Questões 1 e 2 desta dissertação).

Sobre as afirmações 3 e 4, que avaliam a importância de duas novas ferramentas de recompensa implementadas, as respostas indicam que os participantes tiveram uma ligeira tendência a concordar que o seu uso ou a possibilidade do seu uso geram mais motivação a contribuir. A afirmativa 3 tem o objetivo de aferir se receber pontos e medalhas aumentam ou podem aumentar a satisfação pessoal em contribuir. Novamente, o grupo que participou do ambiente com gamificação e teve a experiência de receber este estímulo responderam em sua maioria, 86%, que “concordo totalmente” ou “concordo parcialmente”. Ao passo que o grupo que não teve acesso a esse recurso, e por isso teve que imaginar o seu impacto na sua satisfação pessoal, apontaram para uma resposta mais neutra (média 3,1). Já a afirmativa 4 buscou investigar se a presença da barra de progresso (Figura 18) para obtenção de uma nova medalha pode causar um efeito positivo ou uma motivação adicional para que o participante contribua mais, visando obter essa recompensa. Interessante reparar que para essa funcionalidade a média dos grupos foi a que mais se aproximou, mostrando que tanto para o grupo com gamificação quanto para o grupo sem gamificação esse requisito é um requisito importante. Em ambos os grupos, 71% disseram “concordo totalmente” ou “concordo parcialmente”.

As afirmativas 5 e 6 foram elaboradas para avaliar as funcionalidades implementadas para permitir que usuários tenham acesso ao perfil de outros usuários. A intenção destas ferramentas era aumentar a interação social entre os membros do mesmo projeto ou entre os usuários da plataforma Fast Science, mas os resultados mostram uma tendência neutra em relação a esta melhoria. Interessante observar que o grupo que não teve acesso ao perfil público de outros usuários ficou neutro nas duas afirmativas (média 3,0), mas o grupo que teve acesso respondeu “discordo parcialmente” ou “discordo totalmente” em 57% das respostas em relação ao interesse em acessar o perfil de outros participantes e deram uma resposta um pouco mais neutra, mas com tendência a discordar (média 2,6), em relação a afirmativa de que o perfil público ajuda a conhecer os participantes e acessar as suas redes sociais. Embora ainda pouco conclusivo, estas melhorias talvez necessitem de um tempo maior de uso para que esses

interesses sejam despertados entre os participantes ou necessitem de melhorias na mecânica e visual.

A última afirmação, a sétima, busca averiguar a importância do recebimento de notificações como, por exemplo, sobre as contribuições aprovadas e rejeitadas pelos pesquisadores e as recompensas recebidas (pontos e medalhas). Os resultados mostram uma tendência dos participantes a concordarem que as notificações sobre a participação são importantes. Em ambos os grupos, 71% dos participantes responderam que “concordam parcialmente” ou “concordam totalmente”.

Por fim, se considerar que as respostas do grupo sem gamificação representam a expectativa da existência de uma determinada ferramenta e as respostas do grupo com gamificação representam se a expectativa foi alcançada ou não, pode-se concluir que:

1. Ranqueamento, pontos e medalhas superaram as expectativas.
2. A barra de progresso para obtenção de medalhas atendeu à expectativa.
3. Perfis, redes sociais e notificações ficaram abaixo das expectativas.

4.4 Resultados para Qualidade de Dados

4.4.1 Formulário de avaliação

No formulário de feedback, apenas uma afirmação foi feita em relação a qualidade de dados, já que a maioria das ferramentas implementadas não eram do conhecimento dos participantes. A afirmação foi em relação a funcionalidade de “sinalizar tarefa” e com média de 4,7 nas respostas, os participantes tenderam a responder que “Concordam totalmente” com a importância da funcionalidade como ferramenta de comunicação entre o voluntário e o gestor para, por exemplo, esclarecer dúvidas, indicar problemas ou relatar alguma observação ou descoberta.

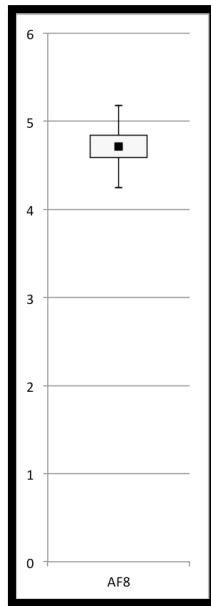


Figura 28 – Diagrama de caixa do resultado

Entretanto, nenhum dos participantes utilizou a ferramenta, mesmo havendo necessidade em pelo menos 4 momentos que ocorreram devido a existência de 2 imagens que não carregaram na terceira etapa do projeto “Jogos e Emoções”. Esse erro ocorreu com dois voluntários que não utilizaram o botão “Sinalizar tarefa” para alertar ao gestor do projeto que haviam imagens com problemas de exibição. Esse resultado, embora ainda pouco significativo, pode ser um indicativo que a finalidade da ferramenta não está suficientemente clara e/ou o layout precisa ser melhorado.

No questionário, havia também uma pergunta opcional: “Deseja fazer algum comentário?”, mas apenas uma pessoa a respondeu. Este voluntário disse que achou que uma das iniciativas de crowd science estava com imagens e perguntas confusas e que isso havia atrapalhado na contribuição e que ele sentiu falta de um tutorial explicando como proceder. No caso, a plataforma Fast Science já possui uma ferramenta de tutorial, com o título de “Instruções” na página principal de cada projeto. Provavelmente o aluno não percebeu a existência da funcionalidade ou o texto escrito não estava suficientemente claro.

Diversos autores, como LE et al. (2010), ressaltam que para garantir a qualidade dos dados é necessário que o participante seja treinado e, para isso, existem várias formas de treinamento: vídeos, tutoriais, cursos presenciais, perguntas e respostas entre outros. Por exemplo, a plataforma CrowdCrafting (<https://crowdcrafting.org>) e a Zooniverse (<https://www.zooniverse.org>) possuem uma estratégia para tratar este

problema. Ao criar um projeto, existe a opção de criar um tutorial on-line que o usuário, ao contribuir pela primeira vez em uma iniciativa de crowd science, ele é obrigado a ler um passo-a-passo que o ensina na prática a como contribuir.

4.4.2 Quantidade de contribuições inválidas

Quem define se uma contribuição é válida ou inválida é o gestor do projeto. Em geral, uma contribuição incorreta não necessariamente é uma contribuição inválida. Por exemplo, se em um projeto de identificação de insetos um usuário diz que determinado besouro é da espécie “besouro-hércules” quando na verdade o mesmo é da espécie “besouro-golias”, o erro é pequeno, mostra que o voluntário teve boa vontade e tal erro geralmente é removido automaticamente quando se têm várias pessoas avaliando o mesmo inseto. Neste caso, o gestor pode considerar a contribuição como *válida* e adicionar um comentário para que o dado seja corrigido posteriormente. Porém, se o voluntário diz que o besouro é da espécie “astronauta”, ele está propositalmente fazendo uma contribuição ruim que precisará ser descartada, sendo assim uma contribuição *inválida*.

Mas como evitar esse tipo de contribuição? Uma das formas é apresentar uma lista de opções válidas para que o usuário apenas tenha que selecionar uma opção. No exemplo acima, se o projeto é sobre insetos, uma lista pré-definida das espécies deverá ser exibida para evitar que informações inválidas ou com erro de digitação sejam enviadas. Entretanto, informações incorretas, mesmo que selecionadas de uma lista pré-definida, sempre poderão ocorrer e uma etapa de validação das informações sempre é necessária.

Buscando reduzir o número de informações inválidas e contribuições maliciosas, algumas funcionalidades foram implementadas. A primeira funcionalidade foi evitar que um usuário responda um campo texto com um dado incompreensível, como “asdf”, evitando o recebimento de uma contribuição inválida. Outra funcionalidade foi permitir que um campo seja marcado como “obrigatório”. Entretanto, ainda existem tipos de contribuições inválidas que não foram tratadas, por exemplo, quando o usuário se contradiz respondendo em um campo que “determinada informação existe na imagem”, mas em outro campo onde ele deveria dizer que informação é esta, ele deixa o campo em branco. Neste caso, o campo não poderia ser marcado como “obrigatório”, pois não necessariamente ele terá valor a preencher em todos os casos. Este tipo de “dado

inválido” que é identificado quando um campo depende de outro campo, só foi percebido após a etapa de avaliação dos resultados, por isso não foi implementado.

Portanto, o objetivo principal que as ferramentas de qualidade implementadas tentaram obter foi reduzir a quantidade de contribuições inválidas forçando que campos obrigatórios sejam preenchidos e que tenham dados inteligíveis. Para avaliar a necessidade da implementação dessas funcionalidades, foi feita uma análise do número de contribuições inválidas registradas na plataforma Fast Science, antes deste experimento, considerando os seguintes parâmetros:

- **Incompletas:** quando campos importantes da tarefa não foram preenchidos.
- **Incompreensíveis:** respostas com texto ininteligível, indicando pressa, descaso ou quando o usuário está apenas testando a plataforma.
- **Erradas:** quando a resposta está completamente incorreta e que demonstra que o voluntário não está capacitado a executá-la ou que está agindo de má fé.

Foram analisadas 640 contribuições, realizadas por 164 participantes distribuídos em 14 projetos. É importante ressaltar que, nesta avaliação, foram considerados inválidos todos os campos sem informação, o que pode ser um viés porque a opção de selecionar o campo como “obrigatório” não existia e, portanto, não temos certeza que o seu preenchimento era obrigatório. Mas, a título de uma avaliação sobre o percentual de completude dos campos oferecidos e que, entendemos que eram de interesse do gestor do projeto ter essas informações preenchidas, o resultado final indicou que 191 (29,8%) das contribuições estavam incompletas, sendo consideradas como inválidas. Portanto, são contribuições que não teriam sido inseridas na plataforma caso a melhoria de campos obrigatórios já tivesse disponível. Em relação as contribuições de texto incompreensível, foram encontradas 12 contribuições nesta categoria, representando 1,9% do total que poderiam ter sido evitadas.

Como a base de projetos é diferente em relação às contribuições feitas em experimentos no passado e no experimento atual e o perfil dos voluntários, idade e grau de instrução, também mudou, não é possível comparar a qualidade dos dados entre o passado e o presente. Porém, é possível comparar a qualidade dos dados entre o grupo A e o grupo B do experimento atual para responder a **QP3**: A gamificação usada afeta a qualidade de dados?

O resultado obtido foi que o grupo com gamificação teve 4,6% ($\pm 0,4\%$) de contribuições inválidas contra 11,3% ($\pm 0,3\%$) do grupo sem gamificação, considerando um intervalo de confiança de 95%. Este resultado indica que a qualidade final foi um pouco melhor no grupo com gamificação, respondendo a QP3 de forma que a qualidade dos dados não foi prejudicada pela gamificação.

Tabela 5 – Análise de contribuições válidas no experimento

	Contribuições Válidas	Contribuições Inválidas	Total
Com Gamificação	188 (95%)	9 (5%)	197
Sem Gamificação	47 (89%)	6 (11%)	53

4.4.3 Tarefas de auditoria

O comportamento dos usuários na execução das tarefas de auditoria também foi avaliado. Como visto neste trabalho, as tarefas de auditoria são tarefas muito simples e tem como o objetivo identificar automaticamente voluntários que não estão contribuindo corretamente ou que não estão prestando atenção. Das 15 tarefas de auditoria realizadas, os voluntários passaram em 11 e erraram em 4. Destas 4, um usuário errou 1 auditoria e outro usuário errou 3 auditorias.

Foi verificado que o usuário que errou 1 auditoria, após cometer o erro e ter um pop-up chamando a atenção para a falha, ele fez mais 2 contribuições no mesmo projeto e estas foram rejeitadas pelo gestor por estarem com dados errados. Esse conjunto de interações evitou que mais contribuições inválidas fossem adicionadas ao projeto. Quanto ao outro usuário que errou 3 auditorias, depois do primeiro erro e de ter sido chamado a atenção pelo pop-up, ele ainda fez mais 17 contribuições no mesmo projeto, errando mais 2 auditorias e tendo 6 contribuições rejeitadas pelo gestor por estarem erradas. Este voluntário estava no grupo de gamificação e disputava as primeiras posições do ranking. Ou seja, melhorias devem ser implementadas nesta ferramenta de auditoria para a conscientização dos voluntários.

Podemos concluir que a funcionalidade de auditoria cumpriu o seu papel, pois ajudou a identificar 2 usuários que realizaram contribuições ruins e que precisaram ser descartadas. Nesse ponto, a ferramenta indica na exportação que estas contribuições são “suspeitas” e facilita que o gestor as remova.

Capítulo 5 – Conclusão

Esta dissertação teve o foco de pesquisa direcionado para dois grandes desafios científicos que estão relacionados com a dificuldade de conseguir obter e manter voluntários contribuindo com as iniciativas de crowd science e, ao mesmo tempo, manter a qualidade dos dados produzidos. A principal contribuição deste estudo foi a investigação, seleção e implementação de novas funcionalidades na plataforma Fast Science para dar suporte a esses dois desafios. As ferramentas implementadas foram avaliadas por dois grupos controlados de voluntários que experimentaram, durante duas semanas, a plataforma Fast Science através da execução de um teste A/B (com e sem elementos de gamificação). Em ambos os casos, as ferramentas de controle de qualidade dos dados foram aplicadas. A resposta do experimento indicou que os resultados esperados foram obtidos: a quantidade de contribuições por usuário aumentou graças a gamificação e a qualidade dos dados não foi prejudicada se comparada ao experimento sem gamificação. As ferramentas de controle de qualidade implementadas mostraram-se eficientes principalmente para evitar o registro de textos sem sentido, campos nulos e respostas mais padronizadas diminuindo desta forma a probabilidade de registros de contribuições inválidas. O sistema de validação facilitou a identificação de contribuições ruins e dos seus respectivos autores, permitindo desta forma o estabelecimento de um canal para troca de experiência e aprendizado entre os gestores dos projetos e os voluntários.

5.1 Questões de Pesquisa

Revisitando as questões de pesquisa descritas no Capítulo 1:

QP1: A gamificação usada ajuda a aumentar a motivação dos voluntários?

QP2: A gamificação usada ajuda a aumentar a quantidade de contribuições?

QP3: A gamificação usada afeta a qualidade de dados?

No capítulo anterior, vimos um indicativo que **QP1** foi confirmada: gamificação aumenta a motivação. Os voluntários responderam ao formulário afirmando que concordam que se sentiram mais motivados devido a existência do ranking, o qual foi a principal ferramenta de gamificação. Já a **QP2** também foi confirmada porque o grupo com gamificação contribuiu quatro vezes mais do que o grupo sem gamificação.

As duas questões foram separadas porque a origem da motivação pessoal em contribuir mais vezes pode estar relacionada a fatores diversos do reconhecimento pessoal ou autopromoção associada aos efeitos do uso de gamificação. O voluntário poderia estar motivado a contribuir, por exemplo, devido a motivações intrínsecas do tipo: senso de propósito, altruísmo ou lazer que poderiam da mesma forma resultar em um aumento de contribuições. Mas o que vimos, de fato, foi que os voluntários se interessaram pelo ranking, pontos e medalhas revelando que sentiram mais satisfação e motivação ao completar as tarefas e, conseqüentemente, contribuíram mais vezes do que o grupo sem gamificação.

Já a **QP3** foi negada pelo experimento. Os resultados mostraram que a qualidade foi até um pouco melhor no grupo com gamificação e, portanto, da forma que a gamificação foi implementada não houve redução na qualidade dos dados.

A seguir uma avaliação final é apresentada, mostrando os resultados obtidos em cada um dos itens implementados e agrupados em melhorias para a qualidade dos dados e motivação dos voluntários.

5.2 Avaliação das melhorias para Qualidade

Um conjunto de ferramentas foram implementadas para melhorar a qualidade dos dados ao longo de todo o processo de criação, execução e avaliação dos dados. No geral, o resultado foi positivo, mas algumas ferramentas mostraram-se mais eficientes do que outras. Abaixo uma crítica de cada uma destas:

- **Sistema de validação de contribuições:** na versão anterior do sistema, não existia uma forma visual e prática para o pesquisador analisar e validar contribuições. Era necessário exportar todas as tarefas e analisar um Excel com milhares de resultados. Por exemplo, em uma das iniciativas da plataforma, “Bebeu Água?”, o Excel de resultados exportado tinha 12 mil linhas. Por isso, a existência dessa ferramenta agregou bastante valor não apenas no processo de validação, mas também na comunicação entre os gestores de projetos e os voluntários.
- **Sistema de auditoria:** o objetivo maior é ajudar a conscientizar os voluntários “flagrados” a contribuírem de forma mais cuidadosa ou orientá-los a não mais cometerem erros. O sistema funcionou de forma adequada, foi possível encontrar com mais facilidade 2 usuários que estavam contribuindo de forma muito errada e facilitou a eliminação das suas contribuições.

Entretanto, observou-se que mesmo sendo notificados usuários podem continuar a contribuir de forma indesejável e punições mais restritivas, como por exemplo, bloqueio de acesso devem ser propostas para trabalhos futuros.

- **Sinalizadores de problemas:** apesar de poder ter sido colocado em prática em 4 momentos devido a problemas na exibição de imagens do dataset, a ferramenta não foi utilizada pelos voluntários. Uma das hipóteses para este comportamento é que a implementação não foi suficientemente intuitiva e seria necessário melhorar a sua apresentação. Outra possibilidade é, sabendo-se que não basta somente implementar uma ferramenta e esperar que os usuários a utilizem livremente, esforços no sentido de melhor entender o público alvo de cada projeto e os mecanismos que podem motivá-los a colaborar devem ser conduzidos nas investigações futuras.
- **Notificações:** o objetivo era notificar o usuário para informá-lo se este estava contribuindo da forma correta ou não a partir da avaliação das tarefas executadas para incentivar quem estava contribuindo da forma certa e ajudar a corrigir quem estava contribuindo mal. A análise da eficácia desta funcionalidade mostrou que, das poucas pessoas que erraram, estas continuaram errando mesmo após as notificações. Não ficou claro se realmente foi um problema de design da ferramenta, pois mesmo mostrando pop-ups informando em destaque que o usuário precisava ter mais atenção, estes continuaram contribuindo de forma errada. Provavelmente, se após cada alerta, o sistema tivesse exibido o tutorial do projeto para o voluntário, o resultado teria sido melhor.
- **Melhorias na execução de tarefas:** permitir que o gestor do projeto configure quais campos são obrigatórios foi uma melhoria simples de ser implementada e que teve um resultado extremamente satisfatório. Identificou-se que o maior “vilão” das contribuições inválidas eram tarefas executadas de forma incompleta e que esta funcionalidade ajudou a eliminar o problema. Além disso, outra melhoria na execução foi adicionar a barreira de “5 segundos” para que cada etapa do workflow de tarefas tenha no mínimo 5 segundos de atenção do voluntário antes de ser finalizada. Porém, tanto na versão anterior do sistema como na versão atual, não se tem registro de data-hora que cada campo é preenchido e sim a data-hora que o conjunto com todos os campos preenchidos do workflow são recebidos pelo servidor.

Por isso, não foi possível confirmar se haviam ou não no passado contribuições extremamente rápidas.

- **Detector de texto sem sentido:** dentre as tarefas concluídas, nenhuma delas foi realizada com texto incompreensível. O sistema não foi programado para salvar tentativas negadas pela ferramenta, por isso não se sabe se de fato ele chegou a bloquear algum texto ou se simplesmente nenhum usuário inseriu texto incompreensível. Também não se sabe se houveram “falsos-positivos”, onde um texto válido foi rejeitado incorretamente.
- **Estatísticas de contribuições:** possui importante aplicação na gestão dos usuários pois permite informar ao gestor do projeto se os dados tiveram diversidade de pessoas contribuindo, quem são os maiores colaboradores e se as contribuições foram contínuas, sazonais ou situacionais. No experimento em questão, as contribuições foram situacionais (localizadas em um espaço curto de tempo) e diversificadas (diferentes voluntários participando).

5.3 Avaliação das melhorias para Motivação

A motivação com gamificação parte do pressuposto que as pessoas gostam de recompensas virtuais e indicativos gráficos de que etapas foram concluídas. Porém, isso nem sempre é suficiente. Se o voluntário não se identificar com a causa e a área de pesquisa ou se a tarefa for extremamente chata de ser realizada, pode ser necessário motivá-lo de outra forma, como exemplo usando recompensas financeiras. O serviço Amazon Mechanical Turk (<https://www.mturk.com/>) é o melhor exemplo disso. Ele provê uma ferramenta de crowdsourcing onde o gestor paga um determinado valor para cada tarefa executada.

Sem utilizar de artifícios financeiros, a proposta neste trabalho foi de implementar as seguintes funcionalidades para motivação:

- **Ranqueamento:** foi a funcionalidade mais bem recebida pelos voluntários. O feedback foi positivo (média da avaliação foi 4,6 de 5) e foi possível acompanhar ao longo dos dias de experimento que as posições de liderança eram alternadas, indicando que havia uma competição entre os participantes.
- **Recompensas:** os voluntários também responderam de forma positiva (nota média 4,1 de 5) que as recompensas, como pontos e medalhas, geraram satisfação pessoal.

- **Autopromoção:** as ferramentas de autopromoção, como fotos de avatar e links para as redes sociais, tiveram o pior desempenho. Além de terem sido mal avaliadas pelo grupo com gamificação (notas médias 2,4 e 2,6 de 5) em relação ao interesse em visitar o perfil de outros usuários. Além disso, nenhum dos participantes preencheu o seu perfil pessoal.
- **Barra de progresso:** com uma avaliação média razoável (nota 3,9 de 5), entende-se que houve interesse dos participantes pelo acompanhamento do progresso para obtenção de medalhas. Cabe destacar que, no grupo com gamificação, único que exibia esse recurso, após duas semanas de experimento, 2 entre os 12 participantes se empenharam e conseguiram obter medalhas de ouro.

5.4 Trabalhos Futuros

5.4.1 Lei de Benford

A Lei de Benford trata de uma observação estatística da frequência dos primeiros dígitos em conjuntos reais de dados numéricos extraídos de sistemas financeiros, eleitorais, de contabilidade, etc. A lei define que em muitas coleções de números encontradas na vida real, os números possuem uma tendência maior de serem iniciados por um algarismo pequeno (BENFORD, 1938). As estatísticas mostram que, geralmente, os números iniciam com o algarismo 1 em 30% dos casos e iniciam com o algarismo 9 em 5% dos casos.

Essa lei é usada, principalmente, para detectar fraudes. Ela possui aplicações para análise de contas de campanhas eleitorais (CHO et al., 2007) e já foi usada diversas vezes por ferramentas de análise para identificar suspeitas de fraude. O argumento é que, quando uma pessoa fraudar um sistema, ela insere números pseudoaleatórios e estes números acabam fugindo da estatística esperada.

Esta estratégia poderia ser utilizada em crowd science para identificar ou levantar suspeitas em casos onde voluntários inserem números de medição falsos, fora da estatística esperada, apenas para “gamificar” ou prejudicar a plataforma. A ideia seria criar um controle estatístico entre contribuições de usuários diferentes para identificar quais usuários estão inserindo dados destoantes e que merecem mais atenção por parte do pesquisador para garantir a sua correção.

5.4.2 Ferramentas de comunicação e divulgação

O objetivo de ter usado gamificação neste trabalho foi para aumentar a motivação dos voluntários e, assim, incentivá-los a colaborar mais. Outra forma de aumentar a motivação é utilizando ferramentas de comunicação para uma maior integração e interação entre os usuários. Quando uma comunidade é atraente ao voluntário, ele passa a visitá-la com mais frequência para que, além de colaborar, ele possa conversar (via chats ou fóruns) e para acompanhar as contribuições e progressos de outros voluntários.

Autopromoção é um elemento bastante comum em jogos, onde o jogador coleta insígnias, medalhas, recordes de pontuação para poder exibi-las em suas redes sociais. Neste trabalho, tentou-se aplicar esta ideia ao permitir que o voluntário criasse um perfil público, adicionando texto, dados pessoais e links para as suas próprias redes sociais. Porém, como observado nos resultados do experimento, esta implementação não teve o uso esperado. Portanto, como trabalho futuro sugere-se a necessidade de estudar melhor o problema e melhorar a implementação, seja na forma visual ou em funcionalidades, para que as ferramentas de comunicação e divulgação (comunidades, fóruns e redes sociais) possam ajudar a aumentar a motivação e a participação dos voluntários.

Atualmente, no Fast Science, não existe integração de logins com contas de redes sociais nem é possível postar resultados e conquistas nas mídias sociais dos voluntários. Se esse tipo de ferramenta fosse implementada, seria possível divulgar e promover melhor os projetos para captação de novos voluntários.

Referências Bibliográficas

- ALABRI, A. e HUNTER, J., “Enhancing the Quality and Trust of Citizen Science Data”, *IEEE Sixth International Conference on e-Science*, 2010.
- ANDERSON, A., HUTTENLOCHER, D., KLEINBERG, J. e LESKOVEC, J., “Discovering Value from Community Activity on Focused Question Answering Sites: A Case Study of Stack Overflow”, *KDD’12*, August 12–16, 2012, Beijing, China, 2012.
- ANTELIO, M., ESTEVES, M. G. P., SCHNEIDER, D. e SOUZA, J. M., “Qualitocracy: A Data Quality Collaborative Framework Applied to Citizen Science”, *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2012.
- BENFORD, F., “The Law of Anomalous Numbers”, *Proceedings of the American Philosophical Society*, Vol. 78, No. 4, 1938.
- BORNFELD, B. e RAFAELI, S., “Gamifying with badges: A big data natural experiment on Stack Exchange”, *First Monday*, Volume 22, Number 6 - 5, June 2017.
- BOWSER, A., HANSEN, D. e PREECE, J., “Gamifying citizen science: Lessons and future directions”, *Designing gamification: Creating gameful and playful experiences*, workshop at CHI, 2013.
- CAILLOIS, R. e BARASH, M., “Man, play, and games”, *University of Illinois Press*, 2001.
- CHO, W. L. T. e GAINES, B. J., “Braking the (Benford) Law: Statistical Fraud Detection in Campaign Finance”, *The American Statistician*, Vol. 61, issue 3, 2007.
- COOK, G., “How crowdsourcing is changing science”, *The Boston Globe*, Boston, November 11, 2011.
- DETERDING, S., DIXON, D., KHALED, R. e NACKE, L., “From Game Design Elements to Gamefulness: Defining ‘Gamification’”, *Mindtrek*, 2011.
- DRESCH, A., LACERDA, D. P. e JÚNIOR, J. A. V. A. “Design science research: método de pesquisa para avanço da ciência e tecnologia”, *Bookman Editora*, 2015.

- ESTEVEZ, M. G. P., “‘Fast Science’ – Uma abordagem para a concepção e execução de projetos científicos com a participação de multidões”, tese de doutorado, *Universidade Federal do Rio de Janeiro*, 2016.
- EVELEIGH, A., JENNETT, C., LYNN, S. e COX, A. L., “‘I want to be a Captain! I want to be a Captain!’: Gamification in the Old Weather Citizen Science Project”, *Proceedings of the First International Conference on Gameful Design, Research, and Applications*, Pages 79-82, October 2013.
- FRANZONI, C. e SAUERMAN, H., “Crowd science: The organization of scientific research in open collaborative projects”, *Research Policy*, vol. 43, no. 1, pp. 1–20, 2014.
- GAMBLE, M. e GOBLE, C., “Quality, Trust, and Utility of Scientific Data on the Web: Towards a Joint Model”, *Proceedings of the 3rd International Web Science Conference*, Article No. 15, June, 2011.
- GHARIBI, R. e MALEKZADEH, M., “Gamified Incentives: A Badge Recommendation Model to Improve User Engagement in Social Networking Websites”, *International Journal of Advanced Computer Science and Applications*, Vol. 8, No. 5, 2017.
- GOLBECK, J., “Trust and nuanced profile similarity in online social networks”, *ACM Trans. Web*, 3(4): p. 1-33, 2009.
- GROH, F., “Gamification: State of the Art Definition and Utilization”, *Proceedings of the 4th Seminar on Research Trends in Media Informatics*, 2012.
- HOWE, J., “The rise of crowdsourcing.”, *Wired Magazine*, junho de 2006. Disponível em: <http://www.wired.com/wired/archive/14.06/crowds.html>. Acesso em: 11 nov. 2018.
- HUDSON-SMITH, A., BATTY, M., CROOKS, A. e MILTON, R., “Mapping for the Masses: Accessing Web 2.0 Through Crowdsourcing”, *Social Science Computer Review*, vol. 27, no. 4, 524-538, November 2009.
- JIN, Y., YANG, X., KULA, R. G., CHOI, E., INOUE, K. e IIDA, H., “Quick Trigger on Stack Overflow: A Study of Gamification-influenced Member Tendencies”, *12th Working Conference on Mining Software Repositories*, 2015.
- KELLING, G., L. e WILSON, J. Q., “Broken Windows: The police and neighborhood safety”, 1982.
- KOHAVI, R. e LONGBOTHAM, R., “Online Controlled Experiments and A/B Tests”, *Encyclopedia of Machine Learning and Data Mining*, April, 2015.

- LACERDA, D., DRESCH, A., PROENÇA, A. e JÚNIOR, J. A. V. A., “Design Science Research: a research method to production engineering”, *Gest. Prod.*, São Carlos, v. 20, n. 4, p. 741-761, 2013.
- MILACH, S., “Ciência cidadã em ambientes costeiros e marinhos: potencialidades para o gerenciamento costeiro integrado no brasil”, 146 p., dissertação de mestrado, *Universidade Federal Fluminense*, 2017.
- MORA, A., RIERA, D., GONZÁLEZ, C. e ARNEDO-MORENO, J., “A literature review of gamification design frameworks”, *7th International Conference on Games and Virtual Worlds for Serious Applications (VS-Games)*, September, 2015.
- NOV, O., ARAZY, O. e ANDERSON, D., “Crowdsourcing for science: understanding and enhancing SciSourcing contribution”, *ACM CSCW 2010 Workshop on the Changing Dynamics of Scientific Collaborations*, 2010.
- PEFFERS, K., TUUNANEN, T., ROTHENBERG, M. A. e CHATTERJEE, S., “A Design Science Research Methodology for Information Systems Research”, *Journal of Management Information Systems*, Vol. 24 Issue 3, Winter 2007-8, pp. 45-78, 2007.
- PEREIRA, C. V., “We4Fit: Promovendo mudanças comportamentais através de gamificação e persuasão”, *Universidade Federal do Rio de Janeiro*, 2014.
- POPPER, K. R., “The logic of scientific discovery”, *Hutchinson*, London, 1959.
- PRESTOPNIK, N. e CROWSTON, K., “Purposeful gaming and Socio-computational systems: A citizen science design case”, *Proc. Group 2012*, ACM Press, 2012.
- RODRIGUEZ, M. A., STEINBOCK, D. J. e WATKINS, J. H., “Smartocracy: Social Networks for Collective Decision Making”, *40th Annual Hawaii International Conference on System Sciences (HICSS'07)*, 2007.
- SANGER, L., “Why did Wikipedia become successful?”, abril de 2015. Disponível em: www.quora.com/Why-did-Wikipedia-become-successful. Acesso em: 11 nov. 2018.
- SHEPPARD, S. A. e TERVEEN, L., “Quality is a Verb: The operationalization of data quality in a citizen science community”, *WikiSym 2011 Conference Proceedings - 7th Annual International Symposium on Wikis and Open Collaboration*, 2011.
- STOYANOVA, M., TUPAROVA, D. e SAMARDZHIEV, K., “Gamification in 11th Grade Mathematics Lessons – One Possible Interactive Approach.”, In AUER,

- M.E., GURALNICK, D., UHOMOIBHI, J. (Editors), Interactive Collaborative Learning, *Proceedings of the 19th ICL Conference - Volume 2*, pp. 41-53, 2016.
- VON AHN, L., “Human Computation”, Doctoral Thesis, UMI Order Number: AAI3205378, *CMU*, 2005.
- WARD, J. S. e BARKER, A., “Undefined By Data: A Survey of Big Data Definitions”, *arXiv*, September 2013.
- WERBACH, K., e HUNTER, D., “For the Win: How Game Thinking Can Revolutionize Your Business”, *Wharton Digital Press*, p. 148, 2012.
- WERBACH, K., “(Re)Defining Gamification: A Process Approach.”, In: Spagnolli A., Chittaro L., Gamberini L. (eds) *Persuasive Technology, PERSUASIVE 2014*, Lecture Notes in Computer Science, vol 8462. Springer, Cham, 2014.
- WIGGINS, A., NEWMAN, G., STEVENSON, R. D. e CROWSTON, K., “Mechanisms for Data Quality and Validation in Citizen Science”, *Proceedings of the 2011 IEEE Seventh International Conference on e-Science Workshops*, Pages 14-19, 2011.
- YOUNG, J. R., “Crowd Science Reaches New Heights. The Rise of Crowd Science”, *Technology - The Chronicle of Higher Education*, 2010.
- ZICHERMANN, G. e CUNNINGHAM, C., “Gamification by design: implementing game mechanics in web and mobile apps.”, Sebastopol, CA: O’Reilly Media, Inc., pp. 35–67, 2011.

Anexo 1 – Formulário de Avaliação para o grupo com gamificação

Formulário de Avaliação

* Obrigatório

Termo de Consentimento

Os resultados obtidos a partir deste formulário de avaliação serão utilizados na dissertação do aluno de mestrado Diego Zanon, matriculado no Programa de Engenharia de Sistemas e Computação COPPE/UFRJ, sob orientação do professor Jano Moreira de Souza. Sua participação é voluntária e você pode desistir a qualquer momento. Os dados obtidos por meio desta avaliação serão mantidos em confidencialidade e os resultados serão posteriormente apresentados de forma agregada, de modo que um participante não seja associado a um dado específico. Obrigado por participar!

Você leu o texto acima e concorda em participar desta pesquisa? *

Sim Não

Nome

Avaliação da Experiência da Participação

As perguntas a seguir foram elaboradas para avaliar a percepção do participante em relação as funcionalidades oferecidas pela plataforma Fast Science.

1) A existência de um ranking geral destacando os usuários que mais contribuíram aumentou a minha motivação em participar

1 2 3 4 5

Discordo totalmente Concordo totalmente

2) Eu acredito que colaborei mais por existir um ranking e querer ficar mais bem colocado(a)

1 2 3 4 5

Discordo totalmente Concordo totalmente

3) Receber pontos e medalhas aumentou a minha satisfação pessoal em contribuir

1 2 3 4 5

Discordo totalmente Concordo totalmente

4) A presença da barra de progresso, indicando uma nova medalha, aumentou a minha motivação para contribuir mais

1 2 3 4 5

Discordo totalmente Concordo totalmente

5) Ao ver a lista dos usuários no ranking, eu tive interesse em acessar o perfil dos participantes

1 2 3 4 5

Discordo totalmente Concordo totalmente

6) O perfil público ajuda a conhecer os participantes e a acessar as suas redes sociais

1 2 3 4 5

Discordo totalmente Concordo totalmente

7) Receber notificações sobre a minha participação é um retorno importante para mim

1 2 3 4 5

Discordo totalmente Concordo totalmente

8) Ter a possibilidade de "Sinalizar a tarefa" e fazer um comentário ajuda a criar um canal de comunicação importante entre o participante e o pesquisador gestor do projeto

1 2 3 4 5

Discordo totalmente Concordo totalmente

Deseja fazer algum comentário?

Powered by



Anexo 2 – Formulário de Avaliação para o grupo sem gamificação

Formulário de Avaliação

* Obrigatório

Termo de Consentimento

Os resultados obtidos a partir deste formulário de avaliação serão utilizados na dissertação do aluno de mestrado Diego Zanon, matriculado no Programa de Engenharia de Sistemas e Computação COPPE/UFRJ, sob orientação do professor Jano Moreira de Souza. Sua participação é voluntária e você pode desistir a qualquer momento. Os dados obtidos por meio desta avaliação serão mantidos em confidencialidade e os resultados serão posteriormente apresentados de forma agregada, de modo que um participante não seja associado a um dado específico. Obrigado por participar!

Você leu o texto acima e concorda em participar desta pesquisa? *

Sim Não

Nome

Avaliação da Experiência da Participação

As perguntas a seguir foram elaboradas para avaliar a percepção do participante em relação as funcionalidades oferecidas pela plataforma Fast Science.

1) Eu acredito que a existência de um ranking geral, destacando os usuários que mais contribuíram, aumentaria a minha motivação em participar

1 2 3 4 5

Discordo totalmente Concordo totalmente

2) Eu colaboraria mais se existisse um ranking porque eu ia querer ficar bem colocado(a)

1 2 3 4 5

Discordo totalmente Concordo totalmente

3) Receber pontos e medalhas aumentariam a minha satisfação pessoal em contribuir

1 2 3 4 5

Discordo totalmente Concordo totalmente

4) A presença de uma barra de progresso, indicando uma nova medalha, aumentaria a minha motivação para contribuir mais

1 2 3 4 5

Discordo totalmente Concordo totalmente

5) Eu gostaria de visualizar a lista de participantes para acessar o perfil deles

1 2 3 4 5

Discordo totalmente Concordo totalmente

6) Um perfil público ajudaria a conhecer os participantes e a acessar as suas redes sociais

1 2 3 4 5

Discordo totalmente Concordo totalmente

7) Receber notificações sobre a minha participação seria um retorno importante para mim

1 2 3 4 5

Discordo totalmente Concordo totalmente

8) Ter a possibilidade de "Sinalizar a tarefa" e fazer um comentário ajuda a criar um canal de comunicação importante entre o participante e o pesquisador gestor do projeto

1 2 3 4 5

Discordo totalmente Concordo totalmente

Deseja fazer algum comentário?

Powered by

