

Marcello Novaes de Amorim

**Integrando *Crowdsourcing* e Computação
Humana para Tarefas Complexas de Anotação
de Vídeos**

Vitória, ES

2019

Marcello Novaes de Amorim

Integrando *Crowdsourcing* e Computação Humana para Tarefas Complexas de Anotação de Vídeos

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Informática da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do Grau de Doutor em Ciência da Computação.

Universidade Federal do Espírito Santo – UFES

Centro Tecnológico

Programa de Pós-Graduação em Informática

Orientador: Prof. Dr. Celso Alberto Saibel Santos

Coorientador: Prof. Dr. Orivaldo de Lira Tavares

Vitória, ES

2019

Ficha catalográfica disponibilizada pelo Sistema Integrado de
Bibliotecas - SIBI/UFES e elaborada pelo autor

N935i Novaes de Amorim, Marcello, 1979-
Integrando Crowdsourcing e Computação Humana para
Tarefas Complexas de Anotação de Vídeos / Marcello Novaes de
Amorim. - 2019.
129 f.

Orientador: Celso Alberto Saibel Santos.
Coorientador: Orivaldo de Lira Tavares.
Tese (Doutorado em Informática) - Universidade Federal
do Espírito Santo, Centro Tecnológico.

1. crowdsourcing. 2. computação humana. 3. agregação. 4.
anotação de vídeo. I. Alberto Saibel Santos, Celso. II. de Lira
Tavares, Orivaldo. III. Universidade Federal do Espírito Santo.
Centro Tecnológico. IV. Título.

CDU: 004

Agradecimentos

Gostaria de agradecer a todos que me incentivaram e me apoiaram ao longo desses anos do doutorado. Agradeço aos meus orientadores Celso e Orivaldo por acreditarem em mim e me ajudarem a delimitar e refinar o trabalho desde a concepção até a conclusão.

Também quero lembrar dos meus colegas de laboratório pela ajuda, pelas conversas, e pelas discussões sobre o projeto, que contribuíram significativamente para identificar pontos de melhoria e questões relevantes que foram essenciais para o sucesso do trabalho.

Quero agradecer também à minha mãe que, mesmo não entendendo exatamente a motivação de fazer o doutorado, sempre me apoiou de todas as formas que pôde.

Finalmente, agradeço especialmente a minha esposa Dani. Obrigado por toda a sua paciência, toda a sua força e por me encorajar desde a primeira vez que mencionei voltar aos estudos.

Também quero deixar registrado o agradecimento à CAPES, pelo suporte financeiro através do projeto PGPTA de referência 59/2014.

"La venganza nunca es buena, mata el alma y la envenena"
(Don Ramón Valdés)

Resumo

A anotação de vídeo é uma atividade que visa suplementar este tipo de objeto multimídia com conteúdo ou informações adicionais sobre seu contexto, natureza, conteúdo, qualidade e outros aspectos. Estas anotações são a base para construção de uma variedade de aplicações multimídia com propósitos diversos que variam de entretenimento até segurança. Existem métodos automáticos para a anotação de vídeo, porém, em alguns casos, estes métodos podem exigir condições e recursos específicos, os quais nem sempre serão encontrados em cenários reais. A anotação manual é uma estratégia que utiliza a inteligência e a força de trabalho de pessoas no processo de produção, sendo uma alternativa para os casos em que os métodos automáticos não podem ser aplicados. Contudo, a anotação manual de vídeo pode ser um processo custoso, pois conforme o tamanho do conteúdo a ser anotado aumenta e o tipo de anotação requerida torna-se mais complexo, a carga de trabalho para produzir anotações também aumenta. *Crowdsourcing* aparece como uma estratégia de solução viável para este tipo de problema, pois se baseia na terceirização das tarefas para uma multidão de trabalhadores, que executam partes específicas do trabalho de forma distribuída. Todavia, conforme aumenta a complexidade das anotações de mídia requeridas, passa a ser necessário utilizar mão de obra especializada, ou disposta a executar tarefas maiores, mais difíceis e mais demoradas. Isso dificulta a utilização do *crowdsourcing*, uma vez que os especialistas demandam maior remuneração, e o seu recrutamento tende a ser uma atividade difícil. Para contornar este problema, surgiram estratégias baseadas na decomposição do problema principal em um conjunto de sub-tarefas mais simples e adequadas para os processos *crowdsourcing*. Estas tarefas menores são organizadas em um *workflow*, de forma que o processo de execução possa ser formalizado e controlado. Na literatura existem algumas propostas diferentes para a construção deste tipo de *workflow*, porém cada uma delas apresenta limitações que ainda precisam ser superadas. Nesse sentido, esta tese visa apresentar um novo *framework* que possibilita a utilização de *crowdsourcing* para criação de aplicações que demandam tarefas complexas de anotação de vídeo. O *framework* desenvolvido considera todo o processo desde a definição do problema e a decomposição das tarefas, até a construção, execução e gerência do *workflow*. Este *framework*, chamado *CrowdWaterfall*, contempla os pontos fortes das propostas atuais, incorporando novos conceitos, técnicas e recursos para superar algumas das suas limitações.

Palavras-chave: crowdsourcing. computação humana. agregação. anotação de vídeo.

Abstract

Video annotation is an activity that aims to supplement this type of multimedia object with additional content or information about its context, nature, content, quality and other aspects. These annotations are the basis for building a variety of multimedia applications for various purposes ranging from entertainment to security. There are automatic methods for video annotation. However, these methods require specific conditions and features that are not even found in actual application scenarios. Manual annotation is a strategy that uses the intelligence and workforce of people in the annotation process and is an alternative to cases where automatic methods cannot be applied. However, manual video annotation can be a costly process because as the content to be annotated increases, so does the workload for annotating. Crowdsourcing appears as a viable solution strategy in this context because it relies on outsourcing the tasks to a multitude of workers, who perform specific parts of the work in a distributed way. However, as the complexity of required media annoyances increases, it becomes necessary to employ skilled labor, or willing to perform larger, more complicated, and more time-consuming tasks. This makes it challenging to use crowdsourcing, as experts demand higher pay, and recruiting tends to be a difficult activity. In order to overcome this problem, strategies based on the decomposition of the main problem into a set of simpler subtasks suitable for crowdsourcing processes have emerged. These smaller tasks are organized in a workflow so that the execution process can be formalized and controlled. In the literature, there are some different proposals for the construction of this type of workflow, but each of them presents limitations that still have to be overcome. In this sense, this thesis aims to present a new framework that allows the use of crowdsourcing to create applications that require complex video annotation tasks. The developed framework considers the whole process from the definition of the problem and the decomposition of the tasks, until the construction, execution, and management of the workflow. This framework, called CrowdWaterfall, contemplates the strengths of current proposals, incorporating new concepts, techniques, and resources to overcome some of its limitations.

Keywords: crowdsourcing. human computation. aggregation. video annotation.

Lista de ilustrações

Figura 1 – Sequência de passos do processo utilizado no <i>CrowdWaterfall</i>	44
Figura 2 – Processo <i>crowdsourcing</i> baseado no cascadeamento de <i>microtasks</i> . . .	47
Figura 3 – Estrutura de um subprocesso <i>task-set</i> referente à uma <i>microtask</i> . . .	48
Figura 4 – Componentes do sistema <i>CrowdNoter</i> e suas interfaces de comunicação. .	62
Figura 5 – Ferramenta para a execução da primeira tarefa do estudo de caso 1. .	72
Figura 6 – Ferramenta para a execução da segunda tarefa do estudo de caso 1. .	73
Figura 7 – Ferramenta para a execução da terceira tarefa do estudo de caso 1. .	73
Figura 8 – Ferramenta para a execução da ultima tarefa do estudo de caso 1. . .	74
Figura 9 – <i>Workflow</i> para o processo <i>crowdsourcing</i> de enriquecimento de vídeo. .	76
Figura 10 – <i>Workflow</i> do processo <i>crowdsourcing</i> para segmentação de sinais de LIBRAS.	79
Figura 11 – Ferramenta para a execução da primeira tarefa do estudo de caso 2. .	80
Figura 12 – Ferramenta para a execução da segunda tarefa do estudo de caso 2. .	81
Figura 13 – <i>Workflow</i> para anotação de efeitos multissensoriais.	85
Figura 14 – Ferramenta para a execução da primeira tarefa do estudo de caso 3. .	86
Figura 15 – Ferramenta para a execução da segunda tarefa do estudo de caso 3. .	87

Lista de tabelas

Tabela 1 – Tarefas Simples X Tarefas Complexas.	6
Tabela 2 – Microtasks X Macrotasks.	6
Tabela 3 – Resumo da pesquisa.	10
Tabela 4 – Características observadas no trabalho de Mendes, Amorim e Santos (2017).	21
Tabela 5 – Características observadas no trabalho de Stepanov et al. (2018). . .	22
Tabela 6 – Características observadas no trabalho de Pérez, Suokas e Bednarik (2018).	23
Tabela 7 – Características observadas no trabalho de Hendrickx et al. (2018). . .	24
Tabela 8 – Características observadas no trabalho de Gadgil et al. (2014)	25
Tabela 9 – Características observadas no trabalho de Cross et al. (2014a).	26
Tabela 10 – Características observadas no trabalho de Deshpande et al. (2014). .	27
Tabela 11 – Características observadas no trabalho de Biel e Gatica-Perez (2013). .	28
Tabela 12 – Atividades e responsabilidades na fase de preparação.	50
Tabela 13 – Atividades e responsabilidades na fase de execução.	52
Tabela 14 – Atividades e responsabilidades na fase de conclusão.	54
Tabela 15 – Interfaces do sistema <i>CrowdNoter</i>	66
Tabela 16 – Tipos de anotação atualmente disponíveis no <i>CrowdNoter</i>	67
Tabela 17 – Métodos de Agregação atualmente disponíveis no <i>CrowdNoter</i>	68
Tabela 18 – Estudos de Caso.	69
Tabela 19 – Anotações no Roteiro X Opinião da <i>Crowd</i>	75
Tabela 20 – instantes iniciais (em s) dos sinais nos vídeos V_1 a V_4	82
Tabela 21 – Contribuições coletadas e agregadas em cada tarefa.	87

Lista de abreviaturas e siglas

CS	Crowdsourcing
HC	Human Computation (Computação Humana)
WoC	Wisdom of Crowds (Sabedoria das Multidões)
HIT	Human Intelligence Task (Tarefa de Inteligência Humana)
PCS	Plataforma de Crowdsourcing
GPC	Gerente do Processo Crowdsourcing
LIBRAS	Língua Brasileira de Sinais

LISTA DE TERMOS

Crowd (Multidão)	Conjunto de <i>workers</i> que disponíveis para realizar as <i>tasks</i> .
Microtask (Microtarefa)	Task que requer pouco tempo, esforço e capacitação.
Owner (Proprietário)	Iniciador e responsável por um processo <i>crowdsourcing</i> .
Task (Tarefa)	HIT, tarefa de computação humana que compõe os processos <i>crowdsourcing</i> .
Worker (Trabalhador)	Colaborador que realiza as <i>tasks</i> em um processo <i>crowdsourcing</i> .

Sumário

1	INTRODUÇÃO	1
1.1	Motivação	3
1.2	Objetivo	6
1.3	Escopo	8
1.4	Questão de Pesquisa	8
1.5	Roteiro de Atividades	9
1.6	Resumo da Pesquisa	10
1.7	Estrutura da Tese	10
2	CONCEITOS FUNDAMENTAIS	11
2.1	Inteligência Coletiva	11
2.2	Computação Humana	12
2.3	Elementos Essenciais dos Processos <i>Crowdsourcing</i>	13
2.4	Discussão sobre os Conceitos Fundamentais	16
3	ESTADO DA ARTE	17
3.1	Estado da Arte do uso de Crowdsourcing na Anotação de Vídeo	17
3.1.1	Trabalhos Relacionados	21
3.1.2	Considerações, Limitações e Desafios no Uso do <i>Crowdsourcing</i> na Anotação de Vídeo	30
3.2	Estado da Arte no Uso de Workflows em Processos Crowdsourcing	32
3.2.1	Estratégias para Construção de Workflows Crowdsourcing	32
3.2.2	Workflows na Anotação Crowdsourcing de Vídeo	34
3.2.3	Considerações, Limitações e Desafios na Definição e Uso de <i>Workflows</i> nos Processos <i>Crowdsourcing</i>	39
3.3	Discussão sobre o Estado da Arte	41
4	FRAMEWORK CROWDSOURCING PARA SUPORTE A TAREFAS COMPLEXAS DE ANOTAÇÃO DE VÍDEO	43
4.1	Estágio de Planejamento	45
4.2	Estágio de Execução	46
4.2.1	<i>Task-Sets</i> : subprocessos crowdsourcing	48

4.2.2	Atores Envolvidos no Processo	49
4.2.3	Fase de Preparação	50
4.2.4	Fase de Implementação	52
4.2.5	Fase de Finalização	53
4.3	Estágio de Distribuição	54
4.4	Agregação Temporal para Anotação Crowdsourcing de Vídeo . . .	55
4.4.1	Método para Agregação de Instantes	58
4.4.2	Método para Agregação de Intervalos	60
4.5	O Sistema <i>CrowdNoter</i>	62
4.5.1	Módulo de Persistência	63
4.5.2	Módulo Servidor	64
4.5.3	Módulo Cliente	65
4.5.4	Interfaces de Comunicação	65
4.5.5	Ferramentas de Anotação	65
4.5.6	Métodos de Agregação	65
4.5.7	Discussão sobre o <i>Framework</i>	67
5	ESTUDOS DE CASO	69
5.1	Aspectos Operacionais dos Estudos de Caso	70
5.2	Estudo de Caso 1: Enriquecimento de Vídeo	71
5.2.1	Preparação do Estudo de Caso 1	71
5.2.2	Objetivo do Estudo de Caso 1	72
5.2.3	Workflow do Estudo de Caso 1	72
5.2.4	Produto Estudo de Caso 1	75
5.3	Estudo de Caso 2: Segmentação de Sinais LIBRAS de Vídeos . . .	77
5.3.1	Objetivo do Estudo de Caso 2	77
5.3.2	Preparação do Estudo de Caso 2	78
5.3.3	Workflow do Estudo de Caso 2	79
5.3.4	Produto do Estudo de Caso 2	81
5.4	Estudo de Caso 3: Anotação de Efeitos Multissensoriais em Vídeos	84
5.4.1	Preparação do Estudo de Caso 3	84
5.4.2	Objetivo do Estudo de Caso 3	84
5.4.3	Workflow do Estudo de Caso 3	85
5.4.4	Produto do Estudo de Caso 3	87
5.5	Discussão sobre os Resultados	88

6	CONCLUSÃO E TRABALHOS FUTUROS	91
6.1	Revisitando a Questão de Pesquisa	91
6.2	Revisitando os Objetivos	92
6.2.1	Objetivo Geral	92
6.2.2	Objetivo Específico 1	93
6.2.3	Objetivo Específico 2	93
6.3	Contribuições	94
6.3.1	Publicações	95
6.4	Limitações	97
6.5	Trabalhos Futuros	97
	REFERÊNCIAS	99

1 Introdução

A anotação de vídeo é um processo que visa complementar este tipo de objeto de multimídia com metadados que descrevam seu conteúdo, contexto, características, qualidade, avaliação dos usuários, que os classifiquem em categorias, ou que associem os vídeos a outros artefatos de mídia (WANG et al., 2009). Esta informação suplementar pode ser usada para facilitar o trabalho de usuários e sistemas que manipulam itens anotados. A anotação serve para destacar pontos de interesse e adicionar informações a conteúdos apresentados (WANG; HUA, 2011; MIHALCEA; CSOMAI, 2007), facilitando a distribuição, indexação, resumo, navegação e a composição de conteúdos multimídia.

Os sistemas que requerem vídeos anotados apresentam diferentes graus de complexidade e sofisticação, variando de simples bases indexadas para busca até sistemas sofisticados de rastreamento de objetos ao longo de vídeos. Isso é algo natural, uma vez que a anotação de mídia é um recurso utilizado para fornecer informação ao sistema, que pode utilizar estas informações para diferentes tipos de processamento. Esta observação é fundamental para delimitar claramente que o escopo do projeto que é o processo de geração das anotações, e não a sua utilização posterior.

Contudo, o vídeo é um tipo de mídia muito expressivo, capaz de carregar uma carga semântica de alta densidade uma vez que vídeos curtos são capazes de transmitir grandes quantidades de informação. Esta característica faz com que as atividades necessárias para a anotação de vídeos tendem a ser custosas, tanto em relação ao tempo necessário quanto pela natureza das tarefas necessárias para realizar a anotação. Este custo não é tão perceptível nos casos em que o volume de vídeo a ser anotado não é significativo, porém se torna evidente conforme este volume aumenta. A demanda por anotação de grandes *datasets* de vídeo motivou a busca por métodos automáticos de anotação.

A automação de atividades dessa natureza traz diferentes benefícios, que vão desde a possibilidade de automatizar o todo o processo de anotação até reduzir os custos com pessoas capacitadas para realizar as tarefas necessárias. Conforme a complexidade das tarefas de anotação aumentam, o grau de capacitação das pessoas necessárias para realizá-las também tende a aumentar, o que também aumenta o custo da anotação de vídeo assim como a dificuldade em conseguir encontrar pessoas com os requisitos necessários.

Os métodos automáticos são uma opção interessante para resolver o problema, porém eles não podem ser aplicados em todos os cenários. [Pal et al. \(2015\)](#) apresentou um estudo de caso que exemplifica como utilizar métodos automáticos para anotação de vídeo. O método apresentado identifica pontos de segmentação em um vídeo com base na similaridade entre os quadros. Nos casos em que se pode descrever objetivamente o que deve ser anotado, é possível utilizar métodos automáticos baseados em similaridade ou reconhecimento de padrões, porém nos problemas abordados nesta tese as anotações estão relacionadas com critérios subjetivos que não podem ser claramente descritos.

Existem casos em que não é possível descrever exatamente a sequência de passos necessárias, ou o que deve ser observado no vídeo durante as atividades de anotação, inviabilizando a criação de algoritmos para automatizar o processo. Mesmo os sistemas baseados em Aprendizado de Máquina (do inglês *Machine Learning* ou *ML*), que são conhecidos por sua capacidade de aprender por meio de exemplos, enfrentam dificuldade para serem aplicados em alguns cenários uma vez que não existem dados de treinamento para todos os casos. Além disso, quando existe uma mudança de contexto, é necessário atualizar os modelos de *ML* para que eles possam funcionar corretamente, e isso restringe a aplicação deste tipo de sistema a alguns domínios para os quais existam dados de treinamento.

O estudo realizado por [Abbas, Ibrahim e Jaffar \(2017\)](#) apontou desafios que ainda precisam ser superados pelos algoritmos de aprendizado profundo (do inglês *Deep Learning* ou *DL*) no contexto da análise de cenas em vídeos. Um dos desafios descritos é a dificuldade para modelar computacionalmente a maneira como o ser humano é capaz de lidar com situações do mundo real como oclusão de objetos, mudanças de cenários, e mesmo transformações geométricas. Estes são exemplos de situações que tornam algumas atividades triviais para os humanos, mas difíceis para as máquinas. Outro problema apresentado foi a dificuldade na obtenção de *dataset* de treinamento adequados para determinados cenários de aplicação. A escassez destas bases de exemplos é um fator limitante para o *ML*. Os fatores limitantes identificados no estudo estão presentes nos problemas de computação humana abordados nesta tese, de forma que os métodos baseados em *DL* encontram dificuldades para resolvê-los.

O estudo da literatura revelou que, até o momento, as tarefas de anotação de vídeo que requerem características humanas como empatia, sensibilidade, subjetividade e intuição ainda requerem trabalho de pessoas, e conforme a complexidade da tarefa, ou o grau de refinamento das anotações aumenta, é necessário empregar pessoas cada vez mais especializadas. Vale ressaltar que o trabalho de especialistas tende a ser caro, e

existe dificuldade de encontrar pessoal qualificado em diversas atividades como no caso dos intérpretes de LIBRAS (Língua Brasileira de Sinais).

Crowdsourcing é uma estratégia que se destaca neste cenário, uma vez que consiste em terceirizar a execução das tarefas para uma multidão (*crowd*) de trabalhadores (*workers*), que podem inclusive ser anônimos e não especializados. As contribuições dos *workers* são processadas e agregadas com o uso de diferentes algoritmos com base estatística para gerar resultados semelhantes, equivalentes, complementares, ou com aceitação parecida aos que seriam gerados por especialistas. Isso pode ser observado no artigo (AMORIM et al., 2017), no qual foi utilizado *crowdsourcing* para geração de apresentações multimídia com base na anotação de vídeo.

A utilização de *crowdsourcing* neste tipo de aplicação é fundamentada em outros dois conceitos: computação humana e sabedoria das multidões. A computação humana (do inglês *Human Computation* ou *HC*) é um paradigma introduzido por Ahn (2005), e consiste em identificar em um processo quais são as tarefas que realmente demandam inteligência humana para serem realizadas. Por sua vez, o conceito de Sabedoria das Multidões (do inglês *Wisdom of Crowds* ou *WoC*), introduzido por Galton (1907), define o modo como as contribuições dos *workers* podem ser agregadas para produzir resultados que representem o senso comum, de forma semelhante a que foi abordada, em um nível mais conceitual, por Lévy e Costa (1993) ao escrever sobre o desenvolvimento do conhecimento nos sistemas colaborativos.

1.1 Motivação

A anotação de vídeo é uma atividade que, em uma abordagem centralizada, demanda tempo e esforço consideráveis de um anotador. Mesmo grupos de trabalho enfrentam problemas para anotar conteúdos de vídeo mais extensos. Neste cenário, as abordagens distribuídas são uma alternativa para contornar o alto esforço exigido para as anotações manuais de mídia (SANTOS; SANTOS; TAVARES, 2007).

Em relação à dinâmica de trabalho os processos distribuídos para anotação de vídeo podem ser cooperativos ou colaborativos. Nas dinâmicas colaborativas, os anotadores trabalham (laboram) em conjunto para resolver um problema principal enquanto, em uma abordagem cooperativa, cada indivíduo resolve uma parte de um problema principal, operando em conjunto para produzir um resultado final, em uma estratégia do tipo *Dividir para Conquistar* (MISANCHUK; ANDERSON, 2001). Ambas as dinâmicas de

trabalho podem ser utilizadas para anotação de mídia, apresentando diferentes vantagens, desvantagens, e requisitos para serem aplicadas.

Para ilustrar a diferença entre as dinâmicas colaborativas e cooperativas para anotação de vídeo, pode-se utilizar um exemplo no qual é necessário identificar os atores que aparecem em vídeo. Em um processo colaborativo, cada participante anotaria todo o vídeo, avaliando, discutindo e complementando as anotações feitas pelos outros participantes. Em um processo cooperativo, cada participante receberia um segmento do vídeo, e o anotaria de forma independente. Neste exemplo, é possível assumir que as anotações do processo colaborativo tendem a ser mais refinadas e completas, uma vez que as anotações finais são geradas a partir da discussão e consenso entre os participantes. Por outro lado, este processo demanda anotadores dispostos a anotar toda a extensão do vídeo, enquanto no processo cooperativo cada anotador recebe uma parte do conteúdo.

O modelo *3C*, introduzido por [Ellis, Gibbs e Rein \(1991\)](#), definiu três aspectos básicos que devem ser considerados ao modelar sistemas colaborativos: comunicação, coordenação e colaboração. Nas dinâmicas colaborativas os anotadores trabalham sobre os mesmos conteúdos, o que demanda comunicação frequente entre eles e uma coordenação ativa para mediar conflitos e chegar a acordos. Nas dinâmicas cooperativas os participantes trabalham de forma independente sobre conteúdos diferentes (ou partes diferentes do conteúdo), de forma que a comunicação entre eles é um aspecto menos relevante, e a coordenação se concentra em distribuir as tarefas e agregar as contribuições individuais em um resultado final.

O uso de estratégias colaborativas para anotação de vídeo é relatado em trabalhos como ([MANZATO, 2011](#)) e ([PIMENTEL et al., 2010](#)). Estes trabalhos utilizaram grupos de colaboradores para produzir anotações semânticas que demandavam características humanas como subjetividade. [Manzato \(2011\)](#) apresentou uma arquitetura para indexação semântica de vídeos a partir de anotações colaborativas. Esta arquitetura permite a personalização da indexação multimídia, preservando as características individuais das indexações geradas, mesmo em uma dinâmica colaborativa. [Pimentel et al. \(2010\)](#) também utilizaram a anotação colaborativa de vídeo, dessa vez para evidenciar pontos de interesse no conteúdo de vídeos e programas de TV interativa, com base em tarefas de anotação.

De acordo com Kozar (2010), uma vantagem do trabalho colaborativo é o refinamento das contribuições por conta das discussões que ocorrem durante o processo. Por outro lado, existe uma dificuldade em gerenciar essas discussões conforme o número de participantes aumenta. No trabalho cooperativo as contribuições são independentes e precisam ser agregadas para gerar um resultado final, porém esta característica também permite paralelizar o processo de anotação.

Por conta da possibilidade de paralelizar o processo de anotação, e por demandar menos dos participantes ao permitir que anotem segmentos de mídia menores, foi decidido que a estratégia adotada nesta pesquisa utilizará uma dinâmica de trabalho cooperativa. A abordagem escolhida para operacionalizar o processo foi o *crowdsourcing*, que define uma estratégia eficiente para terceirizar, para uma *crowd* de *workers*, tarefas que requerem inteligência humana, agregando as contribuições em um resultado final, o que permite anotar grandes quantidades de objetos de vídeo (ZHAO; HOEK, 2015).

Todavia, a geração de anotações complexas de vídeo tende a demandar tarefas grandes ou complexas, nas quais é necessário anotar diversos aspectos de um objeto de mídia, assim como relacionar e contextualizar as anotações. Este tipo de tarefa requer o trabalho de especialistas, ou pessoas que estejam dispostas a receber treinamento ou dedicar mais tempo e esforço na execução da tarefa. Dessa forma, é importante definir uma estratégia para remodelar os problemas que requerem tarefas complexas de anotação de vídeo, para que possam ser resolvidos por meio de um conjunto de tarefas simples e fáceis, que possam ser distribuídas e realizadas por uma multidão de colaboradores não necessariamente especializados.

No contexto deste trabalho, *Anotação Simples de Vídeo* é aquela que demanda pouco trabalho e dedicação, sendo facilmente aplicável em cenários nos quais não se pode assumir que os anotadores possuem treinamento ou especialização no domínio em questão. As anotações simples se usualmente se referem a anotações de aspectos específicos e pontuais do vídeo.

A *Anotação Complexa de Vídeo* é aquela que, usualmente, demanda a anotação de diferentes aspectos do vídeo, sendo comum que seja solicitado que também se anote relacionamentos entre os itens anotados. Este tipo de anotação costuma demandar mais esforço e dedicação, e é comum que seja necessário empregar anotadores com treinamento ou conhecimento específico.

As características que diferenciam objetivamente as tarefas simples das complexas podem ser observadas na tabela 1.

Tabela 1 – Tarefas Simples X Tarefas Complexas.

	Simples	Complexa
Independência	As anotações são geradas de forma independente e atômicas.	As anotações relacionam os itens, eventos e aspectos anotados.
Aspectos	As anotações são feitas sobre um único item, evento ou aspecto do vídeo.	São anotados múltiplos itens, aspectos ou eventos do vídeo.

No campo do *crowdsourcing* é possível traçar um paralelo entre as tarefas simples e complexas, e as *microtasks* e *macrotasks*. Embora sejam conceitos distintos, é possível compreender, com base nas informações apresentadas na tabela 2, que as tarefas simples podem ser modeladas como *microtasks* enquanto é mais viável modelar tarefas complexas como *macrotasks*.

Tabela 2 – Microtasks X Macrotasks.

	Microtask	Macrotask
Tempo	Pode ser realizada rapidamente, sem interrupções.	Requer um tempo maior para ser realizada.
Esforço	Demanda pouco esforço, e pode ser realizada com poucas interações.	Demanda esforço considerável e múltiplas interações.
Qualificação	Pode ser realizada por pessoas sem treinamento ou qualificação específica.	Requer determinadas habilidades, treinamento ou características específicas.

1.2 Objetivo

[Mattsson \(1996\)](#) define os *frameworks* como arquiteturas compostas por um conjunto de artefatos reutilizáveis, e [Fayad, Schmidt e Johnson \(1999\)](#) complementam informando que eles permitem a implementação de uma família de sistemas de forma padronizada e facilitada. Os *frameworks* englobam os métodos, técnicas, processos e recursos necessários para fornecer um arcabouço adequado para uma classe de soluções.

Esta tese visa apresentar uma maneira de aplicar o *crowdsourcing* e a *HC* nas tarefas complexas de anotação de vídeo que requeiram inteligência humana. A maneira encontrada foi desenvolver um *framework* que permite remodelar essas tarefas complexas como um *workflow* de tarefas pequenas e simples. Essas tarefas simples podem então ser realizadas, de forma fácil e rápida, por pessoas não especializadas e sem a necessidade de treinamento ou capacitação consideráveis. *Workflow* é o termo em inglês para fluxo

de trabalho, e está relacionado com a ordem em que um conjunto de tarefas deve ser executado para automatizar um processo.

Assim, o objetivo geral desta tese é desenvolver um *framework* que inclui o processo *crowdsourcing* baseado na decomposição e a remodelagem da tarefa complexa original como um *workflow* de tarefas simples. Este *framework* inclui a formalização e descrição dos papéis, fases e atividades do processo, os métodos de agregação específicos para pontos e intervalos temporais, e os recursos de software para suportar projetos *crowdsourcing* que demandam tarefas complexas de anotação de vídeo. O *framework* desenvolvido recebeu o nome de *CrowdWaterfall*.

Para fins de organização, a formalização do processo e os recursos computacionais serão listados como objetivos específicos: (*OB1*) definir as atividades, as fases e os papéis envolvidos no processo que operacionaliza o *framework* proposto, assim como apresentar um *workflow* formal para ele, e (*OB2*) desenvolver um ambiente de *software* que forneça o suporte computacional para a execução do *workflow* associado ao processo *crowdsourcing*.

O objetivo específico *OB1* visa formalizar um processo que represente os papéis, fases e atividades necessários para operacionalizar o *framework* proposto, assim como apresentar um modelo de *workflow* que seja adequado para guiar este processo. Dessa forma, espera-se apresentar um processo formalizado e um modelo de *workflow* que seja capaz de representar aplicações *crowdsourcing* de vídeo que demandem tarefas complexas de anotação, a partir de uma composição de *microtasks*, possibilitando a especificação e representação em alto nível, o controle do *workflow*, e gerência do processo *crowdsourcing*.

O objetivo específico *OB2* consiste em desenvolver os recursos de *software* e os algoritmos necessários para filtrar e agregar as contribuições ao longo do processo proposto, assim como desenvolver os recursos computacionais necessários para a executá-lo. Os recursos de *software* incluem o ambiente capaz de suportar a execução do processo *crowdsourcing*, incluindo as ferramentas que permitem a realização das tarefas e o módulo responsável pelo monitoramento e gerência do *workflow*. O ambiente de *software* desenvolvido recebeu o nome de *CrowdNoter*, e incorpora os recursos computacionais necessários para operacionalizar o *CrowdWaterfall*.

Entre os algoritmos, é importante destacar os que são responsáveis pela agregação temporal de pontos e intervalos (seção 4.4), pois eles ajudam a generalizar o processo de agregação das anotações de vídeos coletadas por meio de *crowdsourcing*.

1.3 Escopo

Existem diversos tipos de aplicações que estão relacionadas com a anotação de vídeo e isso torna importante definir o escopo desta tese. Os cenários abordados nesta tese envolvem aplicações de vídeo que requerem anotações ainda dependentes da inteligência humana para serem realizadas. Isso exclui os casos nos quais seja possível definir algoritmos para produzir as anotações, assim como os casos para os quais já existem soluções baseadas em *machine learning* ou alguma outra baseada em inteligência artificial.

Outra delimitação importante diz respeito à complexidade da atividade que deve ser realizada sobre os vídeos. Uma vez que a execução de tarefas simples de anotação de vídeo é algo corriqueiro no domínio do *crowdsourcing*, serão abordados somente os casos em que o processo de anotação demande tarefas mais complexas, as quais podem ser quebradas em tarefas simples de processamento do conteúdo do vídeo, sendo estas últimas executadas por colaboradores não especializados pertencentes a uma *crowd*.

Os casos excluídos do escopo também poderiam ser abordados por meio do *framework* proposto, todavia, o foco desta tese está nos problemas que ainda precisam de uma estratégia formal e padronizada de solução.

Sobre os resultados esperados, os esforços serão concentrados em apresentar um *framework* que permita decompor e remodelar as tarefas complexas de anotação de vídeo para torná-las adequadas para o uso de *crowdsourcing*, e executá-las de forma a gerar o produto final desejado. Além disso espera-se apresentar um processo que seja adequado para operacionalizar o *framework*, com os papéis, fases e atividade necessárias, e um modelo de *workflow* que represente este processo.

1.4 Questão de Pesquisa

Para ajudar a entender até que ponto os objetivos podem ser alcançados, existem uma questão de pesquisa que será respondida a partir da análise dos problemas tratados nos estudos de caso.

- QP: Como as anotações de vídeo obtidas de um processo *crowdsourcing*, baseado em um *workflow* de tarefas pequenas e simples, executadas por pessoas que não precisam ser especialistas, se comparam aos resultados gerados por tarefas grandes ou complexas executadas por especialistas?

1.5 Roteiro de Atividades

Os objetivos desta tese foram alcançados a partir da realização de cinco atividades principais: (i) estudo de trabalhos relacionados, (ii) definição da estratégia de decomposição e remodelagem da tarefa complexa, (iii) formalização do processo, (iv) realização dos estudos de caso, e (v) avaliação dos resultados.

O estudo de trabalhos relacionados contempla dois temas que, apesar de distintos, são complementares no contexto deste projeto. O primeiro engloba os trabalhos que utilizam *crowdsourcing* para anotação de vídeos, e o segundo diz respeito aos trabalhos que utilizam os propões *workflows* para processos *crowdsourcing*. O objetivo dessa atividade é identificar o estado atual de desenvolvimento dos trabalhos, quais são as limitações presentes, e descobrir de que forma os resultados gerados nesta tese podem contribuir para o avanço no estado da arte.

Uma vez que o estudo de trabalhos relacionados revelou as formas utilizadas atualmente para modelar e executar atividades complexas sobre vídeos como um processo *crowdsourcing* baseado em *microtasks*, com o auxílio de *workflows* para formalização e controle, a próxima atividade foi entender, definir e formalizar os passos necessários para remodelar a tarefa complexa de anotação de vídeo como um *workflow* de tarefas *crowdsourcing* simples. O próximo passo foi descrever todos os papéis envolvidos e suas respectivas atividades. Uma vez que as atividades foram formalizadas, foi definido o *workflow* capaz de guiar todo o processo, assim como permitir a geração do sistema capaz de operacionalizá-lo.

Conforme abordado por [Teegavarapu e Summers \(2008\)](#), [Freitas e Jabbour \(2011\)](#) e [Santos \(2014\)](#), os estudos de caso fornecem uma maneira eficiente de validar métodos e ferramentas voltadas para aplicações práticas. Sendo assim, a próxima atividade é realizar uma série de estudos de casos para validar o *framework*, e entender as características dos produtos gerados, e identificar necessidades de adequação e melhoria.

Cada um dos estudos de caso contribuiu de forma diferente, e complementar, para a avaliação dos resultados obtidos. Nos estudos de caso em que existiam informações sobre os resultados gerados por especialistas, foi realizada uma comparação entre estas informações e os resultados obtidos pelo processo *crowdsourcing*. Nos outros casos foi avaliada a percepção dos usuários em relação aos artefatos gerados, assim como também foi feita a avaliação desses artefatos por especialistas.

1.6 Resumo da Pesquisa

Com base nos objetivos, hipóteses e a questão de pesquisa, foi formulada a tabela 3 com o resumo da pesquisa que foi realizada para guiar esta tese.

Tabela 3 – Resumo da pesquisa.

Tema	Integrando <i>Crowdsourcing</i> e Computação Humana para Execução de Tarefas Complexas em Vídeos.
Objetivo Geral	Apresentar um <i>framework</i> que inclui o processo <i>crowdsourcing</i> que contempla desde a decomposição e a remodelagem da tarefa complexa original como um <i>workflow</i> de tarefas simples, a formalização e descrição dos papéis, fases e atividades do processo, os métodos de agregação específicos para pontos e intervalos temporais, e os recursos de software para suportar projetos <i>crowdsourcing</i> que demandam tarefas complexas de anotação de vídeo.
Objetivos Específicos	OB1: Definir as atividades, as fases e os papéis envolvidos no processo que operacionaliza o <i>framework</i> proposto, assim como apresentar um <i>workflow</i> formal para ele.
	OB2: Desenvolver um ambiente de <i>software</i> que forneça o suporte computacional para a execução do <i>workflow</i> associado ao processo crowdsourcing.
Questão de Pesquisa	QP: Como as anotações de vídeo obtidas de um processo <i>crowdsourcing</i> , baseado em um <i>workflow</i> de tarefas pequenas e simples, executadas por pessoas que não precisam ser especialistas, se comparam aos resultados gerados por tarefas grandes ou complexas executadas por especialistas?

1.7 Estrutura da Tese

O restante deste trabalho está organizado da seguinte forma: no capítulo 2 serão apresentados os conceitos fundamentais utilizados nesta tese, no capítulo 3 serão analisados os trabalhos relacionados e será feita uma discussão sobre o estado da arte, no capítulo 4 será apresentado o *framework* proposto nesta tese, no capítulo 5 serão apresentados os estudos de caso realizados, e no capítulo 6 será apresentada a conclusão e feita uma discussão sobre as contribuições, limitações e trabalhos futuros.

2 Conceitos Fundamentais

Tendo sido definido o objetivo principal é importante introduzir alguns conceitos fundamentais relacionados com o domínio da pesquisa. Neste capítulo serão apresentados os conceitos de base abordados nesta tese, que foram identificados com base em trabalhos que utilizaram processos *crowdsourcing*, assim como *surveys* e revisões sistemáticas que analisaram vários trabalhos neste mesmo tema.

2.1 Inteligência Coletiva

A Inteligência Coletiva, ou Sabedoria das Multidões (do inglês *Wisdom of Crowds* ou *WoC*), está relacionada a como a correta agregação das contribuições de uma *crowd* heterogênea, grande o suficiente, pode produzir resultados que tendem a ser tão bons quanto os fornecidos por especialistas (GALTON, 1907). Surowiecki (2005) listou quatro requisitos importantes para obter bons resultados de uma multidão.

O primeiro destes requisitos é a **Diversidade**. Cada pessoa possui seu próprio conjunto de experiências, conhecimento prévio e conceitos pessoais. A diversidade promove o aumento na variância das contribuições, diminuindo a chance de se obter grandes quantidades de contribuições replicadas que são reflexo de grupos de *workers* com perfil semelhante.

A **Independência** é o segundo requisito, uma vez que as pessoas formam suas opiniões de forma independente. A independência diminui a chance de direcionamento nas respostas. Não havendo hierarquia entre os *workers* se reduz a chance de se obter contribuições tendenciosas.

O terceiro requisito é a **Descentralização**. As pessoas se baseiam em seus próprios conhecimentos. Uma vez que os *workers* não devem interagir ou se comunicar a respeito da execução das tarefas, é reduzida a chance de que as contribuições de alguns *workers* sejam imitados e suas contribuições replicadas por outros *workers*.

Agregação é o quarto requisito listado, e significa que existe um mecanismo para transformar opiniões pessoais em uma decisão coletiva. O método de agregação utilizado para agregar as contribuições deve ser adequado a cada tarefa. Existem diversas abordagens que podem ser aplicadas de acordo com a natureza das contribuições.

Adicionalmente é importante considerar o processo de **Filtragem** das contribuições, eliminando dados inconsistentes e que poderiam contaminar o resultado da agregação. O método e critérios de filtragem também devem ser selecionados de acordo com a natureza das contribuições e do tipo de resultado a ser gerado. Esses aspectos são requisitos importantes para obter bons resultados pois caracterizam uma *crowd* heterogênea, e que é capaz de fornecer contribuições independente que podem ser agregadas com a finalidade de produzir resultados que representam a inteligência coletiva (LÉVY; COSTA, 1993).

As plataformas de *crowdsourcing* são ambientes que se adequam a esse conceito, especialmente porque é possível atingir *workers* de diferentes partes do mundo, com diferentes contextos e culturas, aumentando assim a diversidade. Um cenário comum nesses ambientes envolve *workers* recebendo tarefas e executando-as sem interagir uns com os outros, o que apoia independência e descentralização. Para cada tarefa solicitada aos *workers* devem ser aplicados critérios de filtragem adequados, assim como um método de agregação específico, que pode envolver a convergência de respostas, médias geométricas, distribuição normal, fusão de contribuição, bem como outros tipos de processamento.

2.2 Computação Humana

Computação Humana (do inglês *Human Computation* ou *HC*), é um paradigma que permite abordar os problemas a partir deste ponto de vista, identificando quais tarefas podem ser automatizadas e quais requerem trabalho humano. Um dos benefícios na utilização do paradigma *HC* é que ele permite aos colaboradores humanos se concentrarem apenas nas tarefas que realmente requerem sua atenção e inteligência. Este tipo de tarefa é chamado Tarefa de Inteligência Humana (do inglês *Human Intelligence Task* ou *HIT*) e corresponde a algo que os humanos podem resolver facilmente, mas que pode ser difícil para ser resolvido por um algoritmo ou uma máquina (LAW; AHN, 2011).

Existem tarefas que são triviais para as pessoas, mas extremamente difíceis mesmo para os softwares e máquinas mais poderosos e sofisticados. Esse tipo de tarefa é caracterizado por requerer criatividade, sensibilidade emocional, empatia, sarcasmo e nível variável de abstração que, até o momento, são características inerentes dos seres humanos. Alguns exemplos desse tipo de tarefa envolvem reconhecimento de objetos ou personagens nas cenas (especialmente quando há oclusão ou envolve análise subjetiva), autoria de conteúdo, análise de emoções e tantas outras atividades que inerentemente requerem que a inteligência humana seja utilizada.

É certo que métodos automáticos estejam sendo pensados para realizar tarefas atualmente realizadas por computação humana, todavia isso demanda condições específicas como conjuntos bem definidos de regras, conteúdo bem estruturado e planejado, ou ainda vastos *datasets* de treinamento (LECUN; BENGIO; HINTON, 2015). Estas condições fazem com que os métodos automáticos não possam ser ainda aplicados em alguns cenários, incluindo os que envolvem conteúdo gerado espontaneamente por usuários que representam uma grande parte do conteúdo gerado atualmente.

A *HC* pode melhorar o desempenho de um processo uma vez que ajuda a modelar *HITs* de forma que possam ser distribuídas para diferentes *workers* para que possam ser executadas em paralelo (ROHWER, 2010). Uma *HIT* pode envolver atividades criativas, como autoria, observação simples, como na identificação de eventos em vídeos, ou empatia, como a detecção de emoções em imagens de expressões faciais.

2.3 Elementos Essenciais dos Processos Crowdsourcing

Crowdsourcing é uma abordagem que lida eficientemente com um grande número de *workers* e contribuições, sendo adequada para operacionalizar processos baseados em HC (AHN, 2005). Essa abordagem geralmente fornece resultados de boa qualidade derivados das contribuições do público e pode distribuir, coletar, validar e combinar grandes quantidades de resultados de tarefas (MO et al., 2013). Considerando que essa abordagem é projetada para lidar com um número considerável de *workers* e contribuições para tarefas que exigem inteligência humana (HOWE, 2006), crowdsourcing é uma forma eficiente para aplicar o paradigma HC em uma escala maciça (AHN, 2011). Segundo Hosseini et al. (2014), *crowdsourcing* é suportado por quatro pilares: *Crowdsourced Task*, *Owner*, *Crowd* e *Crowdsourcing Platform*.

Crowdsourced Task é a *HIT* projetada, de acordo com o paradigma HC, para adquirir contribuições dos *workers*. As instâncias de tarefas são apresentadas aos *workers* que devem executá-las (DIFALLAH et al., 2015). Em termos de complexidade, um *HIT* pode variar de tarefas simples (*microtasks*) a tarefas complexa (*macrotasks*). ***Macrotasks*** exigem mais esforço dos *workers*, muitas vezes exigindo que sejam especialistas na área ou que tenham conhecimento prévio sobre assuntos relacionados à tarefa. No contexto da anotação de mídia, esse tipo de tarefa é adequado para anotações complexas porque pressupõe que o *worker* é qualificado e dedicará o tempo e o esforço necessários para concluí-la. Contudo, tarefas complexas geralmente exigem sistemas de anotação sofisticados e limitam o grupo de *workers* capazes de executá-las (HAAS et al., 2015).

As **Microtasks** geralmente são projetadas para não demandarem tempo, esforço ou capacitação significativos, podendo ser realizadas por *workers* menos qualificados e sem treinamento de forma rápida e fácil. Considerando um cenário de anotação de mídia, tarefas deste tipo são aplicadas para gerar anotações simples, que incluem um ou alguns itens a serem anotados em cada tarefa. As *microtasks* são amplamente utilizadas em projetos de *crowdsourcing* e são apoiadas por plataformas comerciais bem estabelecidas, como a *Amazon Mechanical Turk*¹, a *Figure Eight*² e a *Microworkers*³. Seguindo essas definições, uma *microtask* deve ser modelada de forma a ser:

- **Pequena:** um *worker* deve concluir uma tarefa com poucas interações.
- **Rápida:** um *worker* deve concluir uma tarefa em pouco tempo.
- **Fácil:** quanto mais fácil a tarefa, menos qualificados os *workers* podem ser.

Owner (Crowdsourcer) é o proprietário do projeto, que pode ser um indivíduo ou instituição que possui uma equipe de gerenciamento de *crowdsourcing* composta por especialistas na área que especificam os requisitos técnicos, bem como o pessoal qualificado para especificar e formalizar as tarefas. O *owner* é responsável por iniciar o processo de *crowdsourcing*, definindo qual tarefa deve ser concluída e como ela deve ser apresentada aos funcionários como trabalhos (HOSSEINI et al., 2014).

Crowd é a força de trabalho que move o processo, uma vez que é composto de todos os *workers* que executam as tarefas necessárias para gerar o resultado. Cada *worker* realiza suas tarefas de forma independente para que as obras possam ser executadas em paralelo, de acordo com o paradigma HC (ROHWER, 2010).

Crowdsourcing Platform é um sistema computacional responsável por gerenciar todo o processo, servindo como ponto de entrada para que o *owner* disponibilize as tarefas e que os *workers* as executem. Um ambiente de *crowdsourcing* é necessário, pois as tarefas devem ser disponibilizadas para um número potencialmente grande de *workers*. A plataforma de *crowdsourcing* é um elemento-chave de apoio à cooperação em grande escala. O uso de uma plataforma comercial de *crowdsourcing* traz benefícios, como não ter que se preocupar com o gerenciamento de problemas relacionados a *workers*, *tasks* e contribuições. Além disso, essas plataformas fornecem interfaces para o recrutamento de *workers* e a coleta de contribuições, além de facilitar os pagamentos dos *workers*.

¹ Amazon Mechanical Turk - <https://www.mturk.com>

² Figure Eight - <https://www.figure-eight.com>

³ Microworkers - <http://ttv.microworkers.com>

Em sua dissertação, Neto (2018) classificou esses quatro pilares como elementos essenciais para que os processos *crowdsourcing* possam ser realizados, assim como identificou outros cinco elementos essenciais, que também são apresentados no trabalho de Neto e Santos (2018): Mecanismo de Recompensa, Controle de Atividades, *Workflow* do Projeto, Recursos, e Produto.

O **Mecanismo de Recompensa** é de suma importância para os processos *crowdsourcing* uma vez que os *workers* são humanos e precisam ser motivados a realizarem corretamente as tarefas. A recompensa financeira é uma maneira eficiente de motivar e manter a *crowd*, embora haja outros fatores de enfrentamento, como realização pessoal e gamificação. Um exemplo clássico na utilização de gamificação para se obter anotações de mídia da *crowd* é o ESP Game (AHN, 2006), projeto no qual Luis von Ahn utilizou o conceito de *Games with a Purpose* (GWAP) para anotar um enorme *dataset* de imagens.

O **Controle de Atividades** está relacionado com a gerência do processo *crowdsourcing*. Enquanto a Plataforma Crowdsourcing tem seu foco em intermediar o recrutamento dos *workers*, distribuir as tarefas, e entregar as recompensas, o Controle de Atividades é responsável por preparar os dados de entrada para as tarefas a serem enviadas aos *workers*, agregar as contribuições, compilar os resultados, gerar o resultado final desejado, assim como realizar a verificação da sua integridade e qualidade.

Outro aspecto importante a se considerar é o **Workflow do Projeto** *crowdsourcing*. Neto e Santos (2018) definiram que é essencial projetar um *workflow* do processo de *crowdsourcing*. Esse *workflow* estabelece o fluxo e o controle entre as tarefas, os métodos de agregação e os mecanismos de gerência e verificação da qualidade dos resultados. Por meio destes *workflows* é possível acompanhar o processo e definir critérios para determinar o encerramento das tarefas, e acompanhar a geração do resultado ao longo do processo.

O conjunto de **Recursos** necessários para alimentar o processo *crowdsourcing* também é um elemento essencial. No contexto desta pesquisa os recursos estão relacionados com os Itens de Anotação, que são os objetos de vídeo que devem ser anotados ao longo das tarefas *crowdsourcing*. Além disso, os recursos também englobam as ferramentas necessárias para realizar cada uma das tarefas.

O **Produto** é o resultado final do processo *crowdsourcing*. Ele deve ser definido no início do projeto, e todo o *workflow* deve ser construído com o objetivo de gerar este produto. Neste ponto vale ressaltar que um processo *crowdsourcing* pode gerar diversos tipos de resultado, que podem variar desde arquivos de metadados até objetos de mídia.

2.4 Discussão sobre os Conceitos Fundamentais

Conforme apresentado neste capítulo, existe uma relação direta entre **Inteligência Coletiva**, **Computação Humana**, e **Crowdsourcing**. Nesta pesquisa, a computação humana é utilizada na execução das tarefas que requerem formas de processamento que não podem ser automatizadas, o *crowdsourcing* é utilizado para distribuir as tarefas e coletar as contribuições dos *workers*, e os conceitos relacionados com inteligência coletiva são utilizados na agregação das contribuições de forma a produzir resultados similares ou complementares aos que seriam produzidos por especialistas.

Foram abordados os elementos essenciais dos processos *crowdsourcing*, e estes elementos serão levados em conta tanto na análise dos trabalhos relacionados quanto na concepção do *framework* proposto. Todavia, nesta tese serão discutidos aspectos mais operacionais dos processos *crowdsourcing* de anotação de vídeo, de forma que não serão enfatizadas questões relacionadas com motivação, recompensa e retenção da *crowd*.

No próximo capítulo será feito o levantamento do estado da arte do uso de *crowdsourcing* na anotação de vídeo, assim como o do uso de *workflows* nos processos *crowdsourcing*. Ao final serão discutidos os pontos que ainda estão em aberto, e será apresentada a proposta de incremento no estado da arte.

3 Estado da Arte

Neste capítulo serão apresentados, analisados e discutidos trabalhos relacionados com o uso do *crowdsourcing* na execução de tarefas que demandam anotação de vídeos, com a proposta de métodos para a criação de *workflows* para processos *crowdsourcing*, e com o uso de *workflows* em processos *crowdsourcing*. Ao final haverá uma discussão sobre os pontos que ainda precisam ser cobertos, e apresentada a proposta de incremento no estado da arte.

3.1 Estado da Arte do uso de Crowdsourcing na Anotação de Vídeo

A anotação de vídeo é abordada com frequência com métodos *crowdsourcing*, e diversos trabalhos bem sucedidos já foram relatados em artigos acadêmicos e científicos. Neste capítulo são analisados trabalhos relevantes neste campo, com o objetivo de compreender suas características e o que eles possuem em comum. Além dos artigos dedicados a apresentar experimentos *crowdsourcing* para anotação de vídeo, também foram analisados *surveys*, como os publicados por [Li et al. \(2017\)](#) e [Chittilappilly, Chen e Amer-Yahia \(2016\)](#), que analisaram aspectos importantes de diversos trabalhos relacionados com *crowdsourcing*. Com base nestes *surveys*, e no estudo dos trabalhos apresentados nesta seção, foram definidas as características que deveriam ser analisadas nos trabalhos que utilizaram *crowdsourcing* para anotação de mídia, e estes aspectos foram:

- **Tipo de Interação:** Forma com que o *worker* interage com a ferramenta de anotação e, conseqüentemente, o tipo de dado fornecido na contribuição.
- **Complexidade da Tarefa:** *microtask* ou *macrotask*.
- **Método de Agregação:** recursos de confiabilidade e estratégias de agregação.
- **Qualificação do Worker:** leigos ou especialistas.

O **Tipo de Interação** com o *worker* diz respeito à forma como o ele executa a tarefa, e representa um aspecto importante da concepção da tarefa e *design* da ferramenta

de anotação. Pode variar desde o envio de um arquivo que suplemente o vídeo até a inserção de marcas sobre a linha do tempo de um vídeo para delimitar instantes ou intervalos de eventos. A **Complexidade da Tarefa** foi discutida na seção 2.3, e diz respeito ao tempo, esforço e habilidades necessários para executar uma tarefa. O **Método de Agregação** e a **Qualificação do Worker** serão discutidos a seguir.

O **Método de Agregação** adequado é fundamental para a produção do resultado esperado em um processo *crowdsourcing*. Segundo Galton (1907), Ahn (2005), e Howe (2006), a contribuição de um *worker* não é necessariamente confiável e nem possui a mesma qualidade da contribuição de um especialista, porém, a tendência das contribuições da *crowd* aponta para um resultado equivalente ao fornecido por um especialista. A explicação para isso é que as contribuições da *crowd* tendem a seguir uma distribuição normal (HUBERMAN; ROMERO; WU, 2009) na qual a faixa predominante de valores indica a inteligência coletiva (LÉVY; COSTA, 1993). A normalização dilui as características individuais das contribuições de forma que, conforme a quantidade de contribuições aumenta, a sua agregação revela a opinião da *crowd* e não dos *workers* que a compõe.

Existem diferentes estratégias que podem ser utilizadas para agregar as contribuições, e escolher o método de agregação correto é essencial para gerar o resultado consolidado. Via de regra, os métodos de agregação funcionam normalizando as contribuições, agrupando-as, e verificando a convergência. Porém, de acordo com a natureza do problema a ser resolvido, e das contribuições coletadas, é necessário usar técnicas específicas tanto para o agrupamento das contribuições quanto para o cálculo do resultado.

De acordo com Hung et al. (2013) existem três aspectos a serem considerado na escolha do método de agregação.

O primeiro aspecto é o **Momento da Agregação**. Os métodos *offline* permite o processamento em ciclos iterativos, ajustando os parâmetros do método a cada passo do algoritmo, gerando resultados mais acurados. Todavia, os métodos de agregação *online* são mais leves e podem ser aplicados em tempo real possibilitando o monitoramento da agregação ao longo do processo de coleta de contribuições.

O segundo aspecto diz respeito à **Identidade do Worker**. Existem métodos de agregação que levam em conta a identidade dos *workers*, utilizando sua reputação e experiência para ajustar o peso de suas contribuições no processo de agregação.

O terceiro aspecto é a **Dimensão do Problema**. Alguns métodos admitem apenas contribuições binárias ou previstas em um conjunto restrito, enquanto outros admitem maior variância.

O processo de agregação pode ser *online* ou *offline*. A agregação *online* utiliza algoritmos leves e pode ser aplicada em tempo real, o que permite o monitoramento da convergência do resultado, permitindo que se interrompa a coleta de contribuições assim que se obtiver o resultado esperado. Por sua vez, a agregação *offline* utiliza algoritmos iterativos que podem ter seus parâmetros ajustados a cada passo gerando resultados mais acurados, porém não são adequados para aplicações nas quais se deseja monitorar em tempo real os resultados obtidos.

Segundo Hung et al. (2013), os métodos de agregação *online* mais utilizados em processos *crowdsourcing* são: *Majority Decision* (MD), *HoneyPot* (HP) e *Expert Label Injected Crowd Estimation* (ELICE). Os métodos de agregação *offline* mais comuns são: *Expectation Maximization*, *Supervised Learning from Multiple Experts*, *Generative Model of Labels, Abilities, and Difficulties*, e *Iterative Learning*.

Analisando os problemas de anotação de mídia em relação a estes aspectos foi possível restringir os métodos adequados para o método proposto.

Os métodos de agregação utilizados serão do tipo *online* uma vez que se deseja otimizar a coleta de contribuições, o que demanda avaliação em tempo real dos resultados obtidos a partir das contribuições coletadas o momento.

Uma vez que o método prevê a utilização de diferentes *crowds*, podendo receber contribuições de ambientes comerciais de *crowdsourcing*, contribuições voluntárias via *open call*, ou mesmo grupos internos em ambientes acadêmicos, os métodos de agregação devem permitir contribuições de *workers* anônimos, apesar de serem aceitas contribuições de *workers* identificados.

Por fim, os problemas que demandam atividades complexas de anotação de vídeo muitas vezes admitem contribuições que apresentam grande variância. Como exemplo é possível citar a anotação de conteúdo com respostas abertas, ou ainda anotações temporais em vídeos com uma variância que aumenta junto com o comprimento da mídia. Com base nesta análise, os métodos adequados para nosso cenário são o *MD* e o *HP*.

O *MD* é um método simples que agrega cada objeto de forma independente. Nele é determinado, para cada item de entrada, a resposta que aparece em maior proporção entre as contribuições para este item. Formalmente, dado um objeto $\mathbf{O}_i$, entre as k contribuições recebidas para $\mathbf{O}_i$, é contabilizado o número de contribuições para cada possível *label* $\mathbf{l}_z$. A probabilidade $P(X_i = \mathbf{l}_z)$ de um *label* $\mathbf{l}_z$ é a proporção do número de ocorrências dele sobre k , sendo que $\sum_{\mathbf{l}_z \in L} P(X_i = \mathbf{l}_z) = 1$ (KUNCHEVA et al., 2003).

O *HP* é similar ao *MD*, porém as contribuições fornecidas por *workers* maliciosos são filtradas em uma etapa de pré-processamento. Esta filtragem é realizada por meio da comparação entre as contribuições obtidas e um conjunto de valores de referência. Isso é feito inserindo aleatoriamente nas tarefas perguntas para as quais já se conhece as respostas, e classificando como maliciosos os *workers* que erram estas questões de controle. No entanto, esta abordagem tem algumas desvantagens. Nem sempre existem valores de referência que podem ser utilizados e, quando existem, frequentemente são definidos de forma subjetiva e arbitrária pelo *owner*. Dessa forma, *workers* confiáveis podem ser erroneamente identificados como *spammers* se as questões de controle não forem bem formuladas (LE et al., 2010).

Nos casos em que seja possível determinar a reputação dos *workers* pode-se ainda optar pelo método de agregação *ELICE*, que é uma extensão do *HP* que utiliza dois parâmetros adicionais para atribuir pesos às contribuições: a reputação do *worker* e a dificuldade da tarefa. Dessa forma, a sua aplicação depende de informações que permitam determinar a reputação do *worker*. Estas informações podem ser fornecidas pelo ambiente de *crowdsourcing*, calculada com base no perfil do *worker*, ou extraídas do seu histórico de contribuições, logo é necessário que estas informações estejam disponíveis.

A **Qualificação do Worker** também é algo importante a se considerar, uma vez que os especialistas tendem a fornecer respostas melhores do que as dos leigos. Todavia, além da qualificação dos *workers* é necessário fazer algumas considerações a respeito da confiabilidade da *crowd*. Quando se utiliza uma abordagem *crowdsourcing*, na qual a quantidade de *workers* e contribuições tende a ser muito grande, é importante definir estratégias para garantir a confiabilidade das contribuições. Além da validação com base em valores de referência, existem ainda outras formas de detectar contribuições maliciosas, como a apresentada por Rainer e Timmerer (2014) que consistia em verificar se o tempo que o *worker* levou para executar a tarefa é compatível com a média, assim como se pode verificar se o *worker* visualizou todos os vídeos antes de anotá-los. No caso da anotação de vídeos, pode-se ainda utilizar a técnica apresentada por Anegekuh e Sun (2014), verificando se o *worker* assistiu todo o segmento antes de anotá-lo.

Segundo Khattak e Salleb-Aouissi (2011) a dificuldade da tarefa é determinada pela proporção entre os erros e acertos para as questões de controle utilizadas para a filtragens dos *spammers*.

Yu et al. (2012) realizou um estudo sobre o comportamento da *crowd* e classificou os *workers* em quatro categorias: *Honest Workers* (*HW*), *Moderated Honest Workers* (*MH*), *Moderated Malicious Workers* (*MM*), e *Malicious Workers* (*MW*). Os *HW* são os *workers* honestos, eles fornecem contribuições confiáveis 90% das vezes. Os *MH* são os *workers* moderadamente honestos, que fornecem contribuições confiáveis 70% das vezes. Os *MM* são os *workers* moderadamente maliciosos, que produzem contribuições confiáveis 30% das vezes. Finalmente, os *MW* são os *workers* maliciosos, que fornecem contribuições confiáveis apenas 10% das vezes. Dessa forma, é importante utilizar técnicas como a *Gold Standard* para determinar quem são os *workers* honestos e quem são os maliciosos, para que se possa otimizar a coleta de contribuições.

Surowiecki (2005) propôs que uma boa *crowd*, deve ser heterogênea e não coordenada, de forma que os *workers* maliciosos não podem conspirar para produzir um resultado ruim. Todavia, uma grande quantidade de *workers* maliciosos pode fazer com que sejam necessárias mais contribuições para que os resultados possam convergir.

A seguir serão apresentados trabalhos relacionados com o uso de *crowdsourcing* na anotação de vídeo, e eles serão analisados de acordo com os quatro aspectos discutidos.

3.1.1 Trabalhos Relacionados

Mendes, Amorim e Santos (2017) apresentaram uma estratégia *crowdsourcing* para o sincronismo de vídeos. Neste trabalho os *workers* recebiam pares de segmentos de vídeo que variavam de 5 a 300 segundos, e a contribuição consistia em procurar por um ponto de sincronização entre os segmentos. O objetivo do processo era preencher uma matriz que relacionava todos os vídeos, de forma que cada célula representava o atraso de um de vídeos em relação a um outro.

Tabela 4 – Características observadas no trabalho de Mendes, Amorim e Santos (2017).

Item	Tipo	Observação
Tipo de Interação	Anotação Fechada	Ao clicar em um botão era registrado um ponto de sincronia.
Complexidade da Tarefa	Macrotask	Procurar um ponto de sincronia entre um par de vídeos.
Qualificação do Worker	Leigo	Não era exigida especialização.
Agregação	Majority Decision	Convergência entre os pontos de sincronismo reportados.

Nessa abordagem híbrida era medido o grau de convergência das contribuições para cada par de vídeos, e assim o valor mais prevalente para cada par era calculado em uma abordagem baseada em *MD*. Adicionalmente, conforme a matriz era preenchida, as lacunas eram preenchidas por transitividade sempre que possível, ou seja, ao se determinar o atraso entre o par de vídeos $[A, B]$ e também para o par $[B, C]$, era calculado a distância entre o par de vídeos $[A, C]$.

Além da abordagem híbrida, este trabalho também apresenta um simulador de *crowd* que foi utilizado para validar a parte automática da abordagem híbrida, assim como a estrutura criada para armazenar as contribuições.

Em um trabalho recente [Stepanov et al. \(2018\)](#) se propôs a realizar o mapeamento de diálogos do idioma Espanhol para o Grego. Para isso utilizaram uma *crowd* formada apenas de especialistas para produzir anotações semânticas sobre um *dataset* composto por diálogos sobre os quais já se conhecia a tradução do espanhol para o grego.

A estratégia adotada foi recrutar um time de especialistas, e distribuir entre eles tarefas que consistiam em anotar 50 sentenças, distribuídas em cinco páginas.

Os *workers* deveriam relacionar os termos contidos nas sentenças com os conceitos da ontologia LUNA ([STEPANOV et al., 2014](#)). Ao selecionar um termo, o *worker* podia acessar um *menu* do tipo *DropDown* a partir do qual ele selecionava o conceito da ontologia relacionado com o termo.

Os argumentos apresentados pelos autores para a escolha de usar especialistas foi que, por conta da complexidade da tarefa, a utilização de uma *crowd* genérica (*in-the-wild crowdsourcing*) levaria a um baixo grau de convergência nos resultados.

Tabela 5 – Características observadas no trabalho de [Stepanov et al. \(2018\)](#).

Item	Tipo	Observação
Tipo de Interação	Agrupamento	Associação de termos com conceitos de uma ontologia.
Complexidade da Tarefa	Macrotask	Associar os termos de 50 sentenças a conceitos de uma ontologia.
Qualificação do Worker	Especialista	exigiam proficiência em grego e espanhol.
Agregação	Majority Decision	Coefficiente de Cohen para escalas nominais.

O trabalho de Pérez, Suokas e Bednarik (2018) utilizou uma abordagem *crowdsourcing* para anotar um *dataset* de 792 imagens de olhos humanos com o objetivo de obter um conjunto de treinamento para um sistema automático de detecção de pupilas.

Foi utilizada uma *crowd* não especializada para realizaram duas campanhas independentes e comparar os resultados obtidos. Na primeira solicitaram que os *workers* marcassem um único ponto no centro da pupila, e na segunda solicitaram que eles marcassem vários pontos delimitando o contorno da pupila. Embora a segunda tarefa demande mais interações do *worker*, ela ainda é uma *microtask*.

A cada execução das tarefas o *worker* recebia uma das imagens do *dataset* por meio de uma ferramenta simples, a partir da qual ele podia marcar os pontos solicitados a cada *microtask* e enviar suas contribuições. Cada *worker* podia contribuir diversas vezes, podendo anotar várias das imagens do *dataset*.

Visando garantir a consistência das contribuições, o resultado obtido em um experimento anterior com métodos automáticos foi utilizado como *groundtruth*. Uma vez filtradas as contribuições foi utilizada uma estratégia de agregação *Majority Decision* baseada na posição geométrica dos pontos marcados sobre as imagens.

De acordo com os autores, a coleta de múltiplos ponto para delimitar o contorno da pupila apresentou melhores resultados do que a coleta de pontos únicos no centro da pupila. Eles acreditam que, em parte, isso se deva a capacidade motora limitada de alguns *workers*, o que os impediu de marcar os pontos com precisão. Na estratégia de delimitar o contorno da pupila os erros acabavam se anulando, pois, a delimitação era circular, de forma que imprecisões na vertical ou na horizontal se anulavam.

Tabela 6 – Características observadas no trabalho de Pérez, Suokas e Bednarik (2018).

Item	Tipo	Observação
Tipo da Interação	Anotação Fechada	Coordenadas (X,Y) referentes a pontos sobre uma imagem.
Complexidade da Tarefa	Microtask	Clicar com o mouse sobre imagens de olhos humanos para identificar a pupila.
Qualificação do Worker	Leigo	não foi exigida nenhuma habilidade específica
Agregação	Majority Decision	Média geométrica dos pontos marcados sobre as imagens.

Existem na *Wikipedia* páginas para um mesmo conceito em diferentes idiomas, estas páginas não são traduções de um texto base, mas são construídas por colaboradores em cada idioma. O trabalho de [Hendrickx et al. \(2018\)](#) utilizou uma abordagem *crowdsourcing* para relacionar conceitos extraídos de sentença alinhadas em diferentes idiomas com as páginas da *Wikipedia* relacionadas com ele, nos diferentes idiomas disponíveis.

As sentenças alinhadas são legendas de cursos online, disponíveis em plataformas como a *Cursera*. Essas legendas foram previamente traduzidas do inglês em outros experimentos *crowdsourcing*, de forma que já estavam alinhadas com outros 11 idiomas.

O processo foi realizado com apenas uma tarefa, que consistia em apresentar ao *worker* um par de sentenças no qual uma estava escrita em inglês, assim como uma pequena descrição do contexto no qual as sentenças se encaixam. O *worker* deveria selecionar nas duas sentenças os termos relacionados com um conceito relevante (ponto de interesse) que ele identificasse, e em seguida ele deveria fornecer os links para a página em inglês sobre o conceito, assim como a página no outro idioma. Trata-se de uma tarefa complexa que requer do *worker* tempo e conhecimento dos dois idiomas apresentados.

Tabela 7 – Características observadas no trabalho de [Hendrickx et al. \(2018\)](#).

Item	Tipo	Observação
Tipo da Interação	Anotação Aberta	Termos equivalentes em ambos os textos, e hyperlinks para a Wikipedia nos 2 idiomas.
Complexidade da Tarefa	Macrotask	Analisar pares de sentenças em dois idiomas diferente, identificar conceitos, relacionar termos, pesquisar na Wikipedia, e fornecer hyperlinks para páginas da Wikipedia em dois diferentes idiomas.
Qualificação do Worker	Especialista	O worker precisava dominar tanto o idioma inglês quanto um segundo idioma.
Agregação	Honey Pot	Média geométrica dos pontos marcados sobre as imagens.

Para melhorar a confiabilidade das contribuições foi criado um conjunto de valores de referência com 30 questões de controle para cada possível par de idiomas. Dessa forma foi possível aplicar um método de agregação *HP*, e garantir que cada sentença recebesse

três contribuições válidas. O método de agregação agrupou as contribuições por sentença, e foram mantidas apenas as sentenças sobre as quais houve consenso de ao menos dois entre as três contribuições.

Os autores relataram que aproximadamente 30% das contribuições foram maliciosas, não podendo ser aproveitadas. Eles acreditam que, em parte, isso se deva à complexidade da tarefa, que é grande para um processo *crowdsourcing*. Tarefas grandes e complexas tendem a desmotivar os *workers* que passam a se comportar como *spammers*, visando apenas finalizar a tarefa e receber a recompensa.

O trabalho de [Gadgil et al. \(2014\)](#) utilizou uma abordagem *crowdsourcing* para anotar eventos específicos em vídeos de câmeras de segurança. Neste trabalho foi utilizada uma *crowd* não especializada. Metade dos *workers* recebeu treinamento antes de executarem a tarefa, e o resultado obtido deles foi comparado com o obtido dos que não foram treinados. O treinamento consistia em uma simulação da tarefa, acompanhada de uma trilha de áudio que narrava o que deveria ser feito, como em um tutorial interativo. Adicionalmente foi criado um *groundtruth* a partir de um grupo de especialistas que anotaram uma base de vídeos executando a mesma *microtask* apresentada aos *workers*.

Esta *microtask* consistia em apresentar ao *worker* um segmento de vídeo a partir de uma ferramenta *Web* composta por um *player* que o permitia navegar pelo vídeo, e uma sequência de eventos pré-definidos que ele deveria procurar no segmento. Estes eventos poderiam ser *Assalto*, *Brigas*, *Invasão* e outras ocorrências relacionadas com segurança. Para cada possível evento havia uma barra na ferramenta que permitia ao *worker* delimitar o intervalo da ocorrência, acaso tenha ocorrido. Cada intervalo anotado deveria ser acompanhado de um comentário sobre o evento detectado.

Tabela 8 – Características observadas no trabalho de [Gadgil et al. \(2014\)](#)

Item	Tipo	Observação
Tipo da Interação	Anotação Fechada	Instantes inicial e final de cada evento detectado.
Complexidade da Tarefa	Microtask	Anotar intervalos nos quais ocorreram eventos.
Qualificação do Worker	Leigo	Apenas alguns dos workers receberam treinamento.
Agregação	Moderação Manual	Os administradores avaliavam as contribuições válidas.

Para medir o resultado foram utilizados dos indicativos: (i) *Precision* (relação entre o número de eventos detectados e o total de anotações coletadas), e (ii) *Recall* (relação entre o total de eventos detectados e o total de eventos no *groundtruth*). Embora não conste no artigo, a agregação parece ter sido realizada manualmente pelos administradores do processo, que recebiam alertas a cada ocorrência reportada pela *crowd*.

Os autores reportaram que para os eventos relacionados com *Bagagem Abandonada* houve um desempenho melhor dos *workers* que receberam treinamento, porém para os eventos relacionados com *Infrações de Trânsito* os *workers* sem treinamento produziram um melhor resultado. A conclusão deles foi a de que o *treinamento* ajudou a reduzir os *Falsos Negativos* nas contribuições, mas não puderam afirmar que melhorou a qualidade das contribuições.

Cross et al. (2014a) apresentou o sistema *VidWiki*, que utilizou uma abordagem *crowdsourcing* para enriquecer vídeos educacionais. As ações realizadas pelos *workers* no estudo de caso incluíam substituir textos escritos no quadro pelo professor por caixas de texto digitados, assim como propostas de correção, ajustar tamanho, fonte, cor e posição. O *VidWiki* também permite a inclusão de imagens, fórmulas matemáticas e gráficos. Também era possível ajustar o intervalo de tempo em que cada item ficaria visível.

Além de gerar uma apresentação significativamente mais legível e visualmente agradável, a digitalização dos textos escritos pelo professor no quadro permite inclusive a tradução dos textos para outros idiomas, ou a utilização de tecnologias inclusivas como sistemas de gesticulação para *LIBRAS*.

Tabela 9 – Características observadas no trabalho de Cross et al. (2014a).

Item	Tipo	Observação
Tipo da Interação	Anotação Aberta	Sugestões de edição ou inserção de conteúdo sobre o vídeo.
Mídia Anotada	Vídeo	Um segmento de vídeo obtido por câmeras de segurança.
Complexidade da Tarefa	Macrotask	Sugerir edições ou inserções de conteúdo em vídeos educativos.
Qualificação do Worker	Especialista	Habilidade com editores de texto e imagem, e conhecimento da linguagem LaTeX.
Agregação	Moderação Manual	Os administradores avaliavam e corrigem as contribuições válidas.

Entretanto, cada proposta de edição ou inserção de conteúdo era feita de forma atômica em cada tarefa realizada, o que consiste em uma *macrotask* que exige certo grau de especialização dos *workers*. Entre as habilidades requeridas da *crowd* estão a familiaridade com programas de edição de texto e imagem, assim como conhecimentos da linguagem LaTeX.

Por conta da natureza complexa das anotações coletadas a cada execução da tarefa, foi necessário utilizar moderação humana no lugar de um método automático de agregação, assim como foi feito em (GADGIL et al., 2014). A moderação seguiu o modelo tradicional aplicado nos sistemas de *Wiki*, com moderadores especialistas sendo responsáveis por excluir contribuições inadequadas e ajustar o que for necessário nas contribuições que forem aceitas. Adicionalmente, foi necessário construir um sistema elaborado de edição para permitir que os *workers* executassem a tarefa.

Deshpande et al. (2014) teve como objetivo a legendagem (transcrição) de vídeo-aulas que consistiam na filmagem de aulas presenciais. O intuito de legendar estas aulas foi torná-las acessíveis, em especiais para os estudantes surdos e os que não têm o Inglês como primeiro idioma. Com os argumentos de que o reconhecimento automático de voz ainda apresenta acurácia variável de acordo com as condições, e que o trabalho de tradução manual é lento e cara, os autores propuseram uma estratégia *crowdsourcing* para a legendagem dos vídeos.

Tabela 10 – Características observadas no trabalho de Deshpande et al. (2014).

Item	Tipo	Observação
Tipo da Interação	Anotação Fechada	Seleção de blocos no texto que demandavam edição, e atualização do estado dos blocos.
Complexidade da Tarefa	Macrotask	Sugerir edições para trechos de uma legenda, e sugerir novas revisões caso ache necessário.
Qualificação do Worker	Leigo	Bastava que fosse fluente no idioma original do vídeo.
Agregação	Majority Decision	Os workers chegavam ao consenso de que os blocos da legenda estavam adequados e prontos.

O processo utilizado começa com a geração de uma legenda inicial por meio de um sistema de reconhecimento automático de voz (*ASR*). Para melhorar o desempenho do *ASR* foi utilizada a técnica de *parroting*, que consiste em repetir as falas originais de forma mais clara e pausada, aumentando assim a acurácia do *ASR*.

Neste estudo de caso o desempenho do *ASR* subiu de 68.11% para 95.83% utilizando esta técnica. Em seguida, o texto obtido foi particionado em blocos formados por cinco sentenças, e estes blocos eram disponibilizados para os *workers*.

Por meio de um editor que contava com um sistema de gerenciamento dos blocos de texto, o *worker* podia selecionar blocos para editar. A tarefa então consistia em analisar o texto, editá-lo caso julgasse necessário, e em seguida marcá-lo como *concluído* ou como *precisa de outra revisão* caso esteja inseguro sobre ela. Conforme os *workers* concordavam que um bloco estava *concluído* ele era considerado convergido, de forma que o método de agregação pode ser considerado *Majority Decision*, apesar de não ser automático.

Ao final do experimento os autores aplicaram um questionário para estudantes de diferentes turmas, e avaliaram que eles consideraram que a inclusão das legendas geradas no vídeo realmente foi capaz de enriquecê-lo de forma a melhorar sua eficácia em transmitir o conteúdo.

As anotações de mídia coletadas da *crowd* podem ser feitas sobre diferentes aspectos. Os *workers* podem fazer anotações sobre o conteúdo da mídia, a qualidade, e até mesmo sobre os objetos e personagens presentes nela. O *Youtube Lens* (BIEL; GATICA-PEREZ, 2013) foi um trabalho que teve como objetivo determinar a personalidade de 442 diferentes *Vloggers* a partir de observações feitas pela *crowd* a respeito da impressão que tiveram ao assistir seus vídeo. Um dataset com 2269 vídeos entre 1 e 6 minutos foi segmentado em intervalos entre 50 e 70 segundos.

A tarefa dos *workers* consistia em receber um destes segmentos e preencher um formulário no qual podia avaliar, com notas entre 1 a 5, características do *vlogger*, como sarcasmo, simpatia e ansiedade. Na etapa de agregação dos resultados os formulários foram agrupados por cada autor, e foi calculado o valor médio de cada uma de suas características, segundo a *crowd*. Os autores do trabalho obtiveram resultados semelhantes aos perfis de personalidade que já possuíam sobre os *vloggers*, e que foram utilizados como *groundtruth*.

Tabela 11 – Características observadas no trabalho de Biel e Gatica-Perez (2013).

Item	Tipo	Observação
Tipo da Interação	Avaliação	Nota de 1 a 5 para as questões.
Complexidade da Tarefa	Microtask	Preencher um formulário.
Qualificação do Worker	Leigo	Não foram exigidas.
Método de Agregação	Majority Decision	Média das contribuições.

A lista de trabalhos que utilizam *crowdsourcing* para anotação de vídeo é extensa e, além dos mais relevantes apresentados anteriormente, outros trabalhos também foram analisados de acordo com os parâmetros definidos nesta seção. Um ponto interessante a observar é a variedade de aplicações para este tipo de estratégia, entre elas: sincronização de vídeos (WU et al., 2014; AMORIM; SEGUNDO; SANTOS, 2016), detecção de eventos (SULSER; GIANGRECO; SCHULDT, 2014), anotação de objetos de cena (PINTO; VIANA, 2013), identificação de ações e atividades (RIEK; O'CONNOR; ROBINSON, 2011; DESELL et al., 2015), avaliação de qualidade (FREIBURG; KAMPS; SNOEK, 2011; HAN; LEE, 2014), geo-referenciamento (CHEN et al., 2015; GOTTLIEB et al., 2012), relevância social (HURON SAMUEL, 2013; BERTINI et al., 2013), e legendagem (KACORRI; SHINKAWA; SAITO, 2014).

Entre estes trabalhos foi possível observar a utilização de *microtasks* em (BERTINI et al., 2013; RIEK; O'CONNOR; ROBINSON, 2011; AMORIM; SEGUNDO; SANTOS, 2016; WU et al., 2014; PINTO; VIANA, 2013; SULSER; GIANGRECO; SCHULDT, 2014). De forma complementar, foram utilizadas *macrotasks* em (VONDRICK CARL, 2013; PARK SUNGHYUN, 2014; DESELL et al., 2015; GOTTLIEB et al., 2012; HURON SAMUEL, 2013).

A seguir serão feitas considerações e discutidas limitações e desafios do uso de *crowdsourcing* na anotação de vídeo.

3.1.2 Considerações, Limitações e Desafios no Uso do *Crowdsourcing* na Anotação de Vídeo

Nesta seção foram analisados trabalhos que relataram o uso de estratégias *crowdsourcing* na geração de anotações de mídia, para diferentes aplicações. O estudo destes e outros trabalhos, assim como de *surveys* como (CHITILAPPILLY; CHEN; AMERYAHIA, 2016) e (LI et al., 2017), revelou algumas relações importantes, e mostrou que outras aparente óbvias não são necessariamente verdadeiras.

Por exemplo, uso de *microtasks* na geração de anotações simples e de *macrotasks* para gerar anotações mais complexas parecem inicialmente escolhas óbvias, mas os estudos revelaram que esta associação não é uma regra. Existem casos, como o apresentado em (MENDES; AMORIM; SANTOS, 2017), no qual é necessário aplicar *macrotasks* para gerar anotações simples como uma matriz de sincronização entre vídeos, assim como existem casos, como o apresentado em (BIEL; GATICA-PEREZ, 2013), no qual foi possível utilizar *microtasks* para produzir anotações de natureza mais complexas, como o perfil de personalidade de pessoas. Foi observado que a complexidade das anotações geradas está mais relacionada com a maneira como as contribuições são agregadas, de forma que as anotações são construídas durante o processo de agregação e não diretamente coletadas dos *workers*.

Seguindo esta mesma linha também foi observado que a qualificação dos *workers* está mais relacionada com o tipo de interação e da contribuição coletada do que com a complexidade da tarefa. Existem os casos como o (CROSS et al., 2014b) no qual realmente são necessários especialistas para realizar uma *macrotask*, porém outros trabalhos como (MENDES; AMORIM; SANTOS, 2017) mostram que existem tarefas mais complexas que podem ser executadas por leigos.

A necessidade do uso de especialistas está mais relacionada com o tipo de interação do *worker* na execução da tarefa. Os especialistas são usualmente recrutados em dois cenários: (i) para fornecerem anotações abertas, e (ii) para fornecerem anotações relacionadas e contextualizadas. O primeiro caso está relacionado com a coleta de contribuições abertas, nas quais os *workers* tem um grau de liberdade maior para formular suas respostas. Neste caso o uso de especialistas diminui a carga de moderação posterior ou dos métodos de confiabilidade e agregação. O segundo caso diz respeito à necessidade de relacionar o conteúdo anotado com conhecimentos específicos, ou mesmo com outras partes do conteúdo. Anotações deste tipo podem ser geradas diretamente por especialistas, sendo uma alternativa para as estratégias que utilizam processos em

cascata para produzir este tipo de anotação ao longo de uma sequência de tarefas.

Uma outra observação importante feitas ao analisar esses trabalhos foi a de que os trabalhos utilizaram estratégias semelhantes para anotar diferentes tipos de mídia, como texto, vídeo e imagens. Uma vez que as anotações são tratadas, no contexto do processo *crowdsourcing*, como *metadados*, é possível aplicar formas semelhantes de coleta das contribuições, de distribuição das tarefas, e mesmo utilizar métodos de agregação similares sobre tipos de mídia diferentes. Esta observação é importante para o processo uma vez que a abstração do tipo de mídia a ser anotada ajuda na generalização do processo de anotação de mídia.

Ainda sobre os métodos de agregação, foi possível constatar nos trabalhos analisados o apontamento feito em (CHITILAPPILLY; CHEN; AMER-YAHIA, 2016) sobre a grande utilização dos métodos baseados em *Majority Decision*. Mesmo nos trabalhos que não explicitaram claramente o método de agregação, foi possível observar que os resultados foram construídos com base nas contribuições mais prevalentes ou opiniões mais frequentes, o que caracteriza uma estratégia *Majority Decision*. Em alguns trabalhos, como (CROSS et al., 2014b) e (GADGIL et al., 2014), foram usados especialistas para moderação manual no lugar de métodos de agregação. Todavia, os trabalhos relataram que a carga sobre os especialistas foi menor do que em uma abordagem centralizada.

Na próxima seção será levantado, analisado e discutido o estado da arte no uso de *workflows* em processos crowdsourcing encontrados na literatura.

3.2 Estado da Arte no Uso de Workflows em Processos Crowdsourcing

Para entender melhor como as atividades, papéis e responsabilidades em processos *crowdsourcing* podem ser formalizados na forma de *workflows*, foram analisados tanto trabalhos que utilizaram *workflows* para guiar processos *crowdsourcing* de anotação quanto trabalhos que apresentaram propostas para a criação ou formalização de *workflows* para processos *crowdsourcing*. O objetivo dessa análise é entender como estes trabalhos constroem e gerenciam seus *workflows*, e descobrir quais recursos poderiam ser propostos e introduzidos para facilitar esta prática e torná-la mais prática e eficiente.

3.2.1 Estratégias para Construção de Workflows Crowdsourcing

Após analisar os trabalhos que utilizaram *workflows* bem definidos em processos *crowdsourcing* para anotação, serão apresentados e analisados, em ordem cronológica, os trabalhos que apresentaram as propostas para construção de *workflows crowdsourcing*, a fim de descobrir seus pontos fortes e a suas limitações. Dessa forma será possível selecionar os modelos e recursos que serão utilizados como base, e determinar quais problemas ainda precisam ser resolvidos e os recursos que ainda precisam ser criados para alcançar os requisitos estabelecidos para o *framework*.

Um dos grandes desafios do *crowdsourcing* é definir tarefas que sejam compatíveis com os processos baseados em *microtasks*, que utilizam as plataformas comerciais como o *AMT* e o *Microworkers* para terceirizar o trabalho para uma *crowd* não especializada. O *survey* de [Chittilappilly, Chen e Amer-Yahia \(2016\)](#) revelou que, frequentemente, os trabalhos parte de uma tarefa principal, grande e complexa, e realizam a sua decomposição em *microtasks*. Os trabalhos que realizam esta decomposição utilizam, basicamente, três diferentes abordagens de decomposição de tarefa.

A primeira abordagem é a **Sequencial**. Nesta abordagem as *microtasks* seguem uma ordem de dependência, de forma que a sua execução segue um fluxo em cascata. Na execução em cascata a saída gerada por uma *microtask* é utilizada como entrada para a seguinte. Por conta disso a decomposição da tarefa principal em *microtasks* é feita por meio do mapeamento entre os artefatos iniciais e o produto final desejado, sendo que cada uma das tarefas é modelada para refinar e incrementar o resultado até certo ponto. O trabalho de [Bernstein et al. \(2010\)](#) é um forte exemplo desta abordagem, estabelecendo um padrão para este tipo de cascadeamento de tarefas.

A segunda abordagem é a **Paralela**, ela difere da sequencial em um aspecto básico: uma *microtask* não depende do resultado agregado da anterior. Assim, mesmo que uma *microtask* seja realizada sobre as contribuições coletadas por outra, não é necessário esperar a finalização da primeira para que a segunda seja iniciada. O trabalho de [Noronha et al. \(2011\)](#) apresenta um exemplo desta estratégia, no qual uma *microtask* é utilizada para anotar características nutricionais de alimentos identificados em fotografias por outra *microtask*. Neste exemplo a segunda tarefa coleta anotações baseadas nas contribuições da primeira, sem que seja necessária a agregação do resultado anterior para fornecer uma entrada para ela.

A terceira abordagem é conhecida como **Dividir para Conquistar**. Esta estratégia é utilizada quando a tarefa principal não admite decomposição em sua estrutura, de forma que a decomposição é feita sobre os dados a serem processados pela *crowd*. Em outras palavras, a tarefa principal é distribuída para os membros da *crowd*, todavia as contribuições são feitas apenas sobre um subconjunto ou segmento da entrada original. [Negri et al. \(2011\)](#) apresenta uma aplicação desta estratégia na construção de corpus bi-linguais. A tarefa principal deste trabalho é anotar e associar sentenças em idiomas diferentes, mas na abordagem *crowdsourcing* cada *worker* recebe apenas um pequeno conjunto de sentenças para anotar.

As unidades básicas para a construção dos *workflows* para os processos *crowdsourcing* são as *microtasks*. A partir delas é possível criar diferentes modelos de *workflow*, com diferentes atividades e eventos de transição ([CHILTON et al., 2013](#)). Todavia, entre os trabalhos analisados é possível observar duas estratégias básicas para construir estes *workflows*: (i) *Map-Reduce* e (ii) *Cascading*.

O trabalho que introduziu os *workflows* para processos *crowdsourcing* baseados em *microtasks* foi o *TurKit* ([LITTLE et al., 2010](#)), que apresentou um *framework* capaz de interpretar algoritmos escritos em uma linguagem procedural em um paradigma imperativo. Dessa forma é possível escrever verdadeiros algoritmos de computação humana, associando as funções de entrada de dados à *microtasks* que são executadas pela *crowd* a partir do *Amazon Mechanical Turk*. Uma vez que é possível especificar algoritmos nos quais o retorno de uma função (*microtask*) pode ser processado e utilizado como entrada para uma outra função, este trabalho também introduziu a estratégia *Cascading* para construção de *workflows crowdsourcing*. Atualmente o *Turkit* encontra-se descontinuado.

Apesar de ser adequado para especificar *workflows* para diversas aplicações *crowdsourcing* o *Turkit* apresenta algumas sérias limitações.

A primeira limitação é a necessidade de habilidades de programação para especificar os processos *crowdsourcing*. De modo geral este não é o perfil médio do *owner*. Outro fator limitador na utilização é o baixo nível de abstração, uma vez que a modelagem do processo é feita por meio da escrita de um algoritmo em vez de ferramentas visuais de *design*. Existe também uma forte dependência da plataforma, no caso a *Amazon Mechanical Turk*, e o *Turkit* também não apresenta meios para realizar a gerência interativa do *workflow* nem monitoramento do processo.

A estratégia *Map-Reduce* para a construção de *workflows crowdsourcing* foi introduzida por Kittur et al. (2011) com o *framework CrowdForge*. Este *framework* permite a construção dos *workflows* utilizando três operações: (i) *Partition*, (ii) *Map* e (iii) *Reduce*. A operação *Partition* é responsável por dividir uma tarefa em sub-tarefas; a operação *Map* é responsável por criar instâncias das sub-tasks para que sejam executadas pelos *workers*; e a operação *Reduce* é responsável por combinar as contribuições dos *workers*, sucessivamente.

Apesar de permitir, de forma intuitiva, a construção de *workflows* para processos *crowdsourcing*, a própria estrutura destes *workflows* traz algumas limitações. Entre as limitações é importante citar que estes *workflows* não suportam processos recursivos nem iterativos, não sendo possível retomar uma tarefa finalizada. O modelo *Map-Reduce* também não é adequado para o cascadeamento de tarefas, não sendo possível alimentar uma tarefa com o resultado da anterior. Adicionalmente, os *workflows* gerados a partir de *forks* e *joins* não são compatíveis com as atividades de monitoramento de agregação e gerência do *workflow*.

Uma vez que foram analisadas as propostas atuais para construção de *workflows* para processos *crowdsourcing*, a seguir serão apresentados e discutidos trabalhos que utilizaram *workflows crowdsourcing* na anotação de vídeos.

3.2.2 Workflows na Anotação Crowdsourcing de Vídeo

A seguir serão apresentados e analisados trabalhos que utilizam *workflows*, claramente representados, em processos *crowdsourcing* de anotação, e ao final desta seção será feita uma discussão sobre o que foi observado. Vale observar que as limitações apresentadas para cada trabalho são aquelas declaradas pelos respectivos autores em seus artigos.

O *Soylent* (BERNSTEIN et al., 2010) é um sistema que se propõe a fornecer três atividades independentes relacionadas à editoração de textos: (i) encurtar parágrafos mantendo o conteúdo semântico; (ii) corrigir erros de grafia e gramaticais; e (iii) corrigir problemas de formatação.

Este trabalho apresenta uma contribuição muito importante para os processos *crowdsourcing* que é o *pattern Find-Fix-Verify*, utilizado por outros trabalhos relevantes como (KIM; SIGAL; XING, 2014), (KULKARNI; CAN; HARTMANN, 2012), (AMBATI; VOGEL; CARBONELL, 2012), e (CHEN et al., 2017). Cada padrão *Find-Fix-Verify* equivale a um *task-set*, que é um conjunto de atividades que compõe uma tarefa, e pode ser composto de *microtasks* e métodos automáticos. No caso do *Soylent* cada uma das três foi modelada como um destes *task-sets*.

Nas três tarefas, a atividade *Find* representa a coleta de anotações da *crowd* sobre um segmento do conteúdo. O método de agregação agrupa as contribuições para cada item anotado, e as mescla em um resultado parcial para cada item. Na atividade *Fix*, o *worker* recebe um dos resultados obtidos na tarefa anterior e sua contribuição consiste em editá-la, com o objetivo de corrigir problemas que podem ter sido gerados no processo de agregação. O método de agregação da segunda tarefa agrupa as sugestões de edição para cada item, e elimina as contribuições inconsistentes. Por fim, a atividade *Verify* consiste em apresentar ao *worker* as sugestões de anotação para um determinado item para que ele possa votar no que acredita ser mais adequado. O método de agregação utiliza a estratégia *Majority Decision* para determinar as anotações escolhidas pela *crowd*.

Este é um trabalho importante tanto por ter introduzido a utilização de *workflows* em cascata nos processos *crowdsourcing*, quanto por demonstrar que é possível aplicar um padrão de *task-set* para se modelar as diferentes tarefas que compõem um *workflow crowdsourcing*.

Ambati, Vogel e Carbonell (2012) utilizou um *workflow* de três *microtasks*, no qual cada tarefa demandou um grau de expertise diferente da *crowd*. As tarefas foram executadas em sequência em uma estratégia de cascadeamento na qual o resultado foi refinado sucessivamente a cada fase.

Na primeira tarefa foi realizada uma tradução léxica, uma palavra por vez. Nesta *microtask* o *worker* recebia palavras e fornecia uma tradução para ela, e cada *worker* podia realizar esta tarefa sucessivas vezes com palavras diferentes. Para minimizar problemas de contexto, era apresentada ao *worker* a sentença na qual a estava localizada. A agregação foi feita com base na convergência das contribuições, determinando a tradução mais

recorrente para cada palavra. Nesta primeira tarefa foram admitidos *workers* com noções básicas dos dois idiomas, o fonte e o destino.

Para executar a segunda tarefa foram admitidos apenas *workers* com bom conhecimento dos dois idiomas, ou seja, especialistas. Os *workers* recebiam uma sentença no idioma fonte, e deveriam sugerir uma tradução completa para a sentença. O resultado da primeira *microtask* foi utilizado para auxiliar no trabalho dos *workers*, apresentando as traduções individuais de cada palavra, e era solicitado que se utilizasse essas traduções para gerar a sentença. Dessa forma, foi possível verificar se as sentenças sugeridas eram válidas, assim como permitiu a comparação das contribuições na fase de agregação. A saída desta tarefa foi uma lista de sugestões para tradução de cada sentença.

A terceira tarefa consistiu em apresentar ao *worker* a lista de sugestões de tradução para uma determinada sentença, e solicitar que ele escrevesse uma nova sentença com suas próprias palavras. Nesta tarefa os *workers* não precisavam ser bilíngues, porém precisavam ser fluentes no idioma destino.

Os pontos de destaque para este trabalho foram (i) utilizar *workers* com qualificação diferente em cada *microtask*, e introduzir no *workflow* uma tarefa com o objetivo de fornecer recursos para auxiliar os *workers* na próxima tarefa. Neste trabalho foi utilizado o *pattern Find-Fix-Verify* (BERNSTEIN et al., 2010) para a criação do *workflow*, todavia, não foram utilizados recursos computacionais para gerenciar o processo nem o fluxo de dados entre as *microtasks*.

Heilbron e Niebles (2014) teve como objetivo gerar um *dataset* anotado de vídeos disponíveis na *Web*, com anotações sobre atividades cotidianas contidas neles. Para isso foi definido um *workflow* de três *microtasks*. O *dataset* obtido seria utilizado para treinar um sistema automático capaz de recuperar vídeos que contêm cenas de atividades cotidianas das pessoas.

Na primeira tarefa consistiu em obter *strings* de busca os vídeos, e para obter estas *strings* foi requisitado aos *workers* que sugerissem nomes de atividades cotidianas, como "passear com o cachorro" ou "fazer a barba". Apesar de não haver descrição do método de agregação utilizado, foi relatado a aplicação de atividades de pós-processamento após esta tarefa, que consistiram em expandir a lista obtida por meio de *wordnets* e listas de sinônimos, e posteriormente a utilização destas *strings* de busca para recuperar vídeos do *Youtube*. Assim, a saída da primeira tarefa foi um conjunto de vídeos rotulados pela *string* de busca.

A segunda tarefa teve como objetivo filtrar o conjunto de vídeos obtido anteriormente. O *worker* recebia um vídeo por meio de uma interface que continha um formulário. Por meio deste formulário o *worker* podia marcar se o vídeo realmente era relacionado com o rótulo, assim como especificar algumas características do vídeo. Para melhorar a confiabilidade das contribuições foi adotada uma filtragem por *Gold Standard*, e o método de agregação considerou o percentual de concordância entre as respostas da *crowd*.

A terceira atividade visou determinar os segmentos de vídeo nos quais a ação cotidiana acontecia, por exemplo, identificar o trecho do vídeo no qual uma pessoa está escovando os dentes. Para isso, o *worker* recebia um vídeo por meio de uma interface que o permitia selecionar segmentos, assim ele podia anotar o trecho do vídeo que poderia ser segmentado para isolar a ação. O algoritmo de agregação para esta tarefa se baseou na interseção entre as contribuições para determinar estes segmentos.

Vale observar que o resultado produzido, apesar de complexo para ser gerado por pessoas sem treinamento, pode ser realizado pelo *crowd* por conta da estratégia de cascadeamento de *microtasks*. Neste trabalho, o *workflow* foi construído utilizando um padrão *Find-Verify-Expand*, baseado no *pattern* introduzido em (BERNSTEIN et al., 2010). Entretanto a gerência e comunicação do *workflow* foram realizadas manualmente pelo *owner*, sem a utilização de recursos computacionais que apoiassem estas atividades.

O *ToolScape* (KIM; SIGAL; XING, 2014) é um sistema *crowdsourcing* para enriquecimento de vídeo baseado em um *workflow* composto por três *microtasks* em cascata. Este sistema se propões a enriquecer vídeos tutoriais do tipo *how-to* adicionando a possibilidade de navegar entre os instantes em que cada instrução tem início.

Na primeira tarefa do *workflow*, o *worker* recebe um segmento do vídeo por meio de uma ferramenta que conta com um *player* e um botão para que ele possa marcar o início de cada instrução do vídeo tutorial. A cada novo ponto marcado, é solicitado ao *worker* que insira um rótulo que descreva a instrução identificada. O método de agregação para esta *microtask* agrupa temporalmente as contribuições, de forma que cada agrupamento está relacionado com uma instrução. Assim, a agregação gera uma lista com o instante inicial de cada instrução no vídeo tutorial, e as sugestões de rótulo para cada uma delas.

Na segunda tarefa o *worker* recebe o vídeo posicionado no instante inicial de uma das instruções identificadas anteriormente, assim como a lista de sugestões de rótulos para identificá-la. O trabalho do *worker* consiste em selecionar o rótulo que acredita ser mais adequado para descrever a instrução. O método de agregação identifica o rótulo

mais voltado para cada instrução, e fornece para a próxima tarefa a lista com o instante inicial de cada instrução e o rótulo que melhor o descreve, segundo a *crowd*. Antes de iniciar a terceira *microtask*, o sistema realiza um pré-processamento que consiste em extrair *thumbnails* de *frames* contidas no intervalo de cada instrução. Assim, o *worker* recebe o vídeo posicionado no início de uma instrução, o rótulo que a descreve, e a lista de *thumbnail* que o sistema extraiu.

A terceira tarefa consistiu em selecionar a *thumbnail* que melhor ilustra a instrução. O método de agregação novamente utiliza a estratégia *Majority Decision* para determinar a *thumbnail* escolhida pela *crowd* para cada instrução.

O resultado deste processo é um vídeo enriquecido com recursos de navegação aprimorados, que permitem localizar facilmente cada uma das instruções contidas em um vídeo tutorial.

Ford (2016) introduziu o primeiro *dataset* para *sequential vision-to-language*, e para demonstrar como este conjunto de imagens poderia ser utilizado para fins de *visual storytelling*, foi conduzido um experimento *crowdsourcing* que contava com um *workflow* de duas tarefas. Estas tarefas foram aplicadas em cascata, de forma que a saída da primeira serviu como entrada para a segunda.

A primeira tarefa consistia em apresentar ao *worker* um álbum de fotos, entre as quais ele deveria selecionar e sequenciar algumas de maneira que contassem uma história. Para otimizar a coleta de contribuições, alguns álbuns foram descartados conforme os *workers* relataram não ser possível construir uma sequência.

Na segunda tarefa, o *worker* recebia uma das sequências geradas anteriormente, e deveria escrever um texto que contasse a história representada nesta sequência de imagens. Para tentar garantir boas contribuições, foi estipulado que os textos deveriam conter ao menos 15 palavras.

Neste trabalho ficou muito evidente a estratégia de cascadeamento das *microtasks*, com a primeira tarefa sendo utilizada para alimentar a segunda, na qual o resultado foi realmente produzido. Todavia, não foi utilizado nenhum método formal para a criação do *workflow* e não foi utilizado nenhum recurso para gerenciá-lo. Todo o processo gerenciado manualmente, incluindo o fluxo de dados uma vez que o *owner* alimentou manualmente a segunda *microtask* com o resultado da primeira. Uma vez que se trata de um processo curto, com apenas duas *microtasks*, isso pode não representar um grande problema, mas um sistema que permitisse a gerência do *workflow* e a comunicação entre as *microtasks* poderia facilitar a execução do processo e o fluxo de dados.

A seguir será apresentada uma discussão sobre as limitações e desafios na utilização de *workflows* nos processos *crowdsourcing*, o que engloba as aplicações de anotação de vídeo, que são o foco desta tese.

3.2.3 Considerações, Limitações e Desafios na Definição e Uso de *Workflows* nos Processos *Crowdsourcing*

Nos últimos anos, tanto o trabalhos de [Little et al. \(2010\)](#) quanto o de [Kittur et al. \(2011\)](#) serviram de base para novas propostas para a construção de *workflows* para processos *crowdsourcing*.

O *Turkomatic* ([KULKARNI; CAN; HARTMANN, 2012](#)) é um trabalho que teve como objetivo adicionar ao *CrowdForge* as funcionalidades de visualização e edição dos *workflows*. Todavia, as ferramentas para criação e edição são baseadas em formulários e indentação, não são ferramentas visuais *WYSIWYG*.

Adicionalmente, este novo sistema suporta que os processos sejam especificados de forma algorítmica, de maneira semelhante ao que é feito no *Turkit*. Porém, estes algoritmos são capazes apenas de explicitar os *workflows* gerados pelo *CrowdForge*, ou seja, são apenas *scripts* sequenciais, sem estruturas condicionais nem de repetição. Assim, apesar de agregar ao *CrowdForge* recursos muito úteis de visualização e edição, este trabalho ainda mantém todas as limitações do *CrowdForge*.

O *CrowdWeaver* ([KITTUR et al., 2012](#)) é um outro trabalho que propôs uma evolução do *CrowdForge*, que resolve algumas das limitações e adiciona recursos interessantes. O ponto forte do *CrowdWeaver* é a possibilidade de se criar os *workflows* a partir de uma ferramenta visual *WYSIWYG*, o que permite aos *owners* sem habilidades de programação criarem os *workflows*. Adicionalmente, este sistema permite que o *owner* combine tarefas de computação humana (*HITs*) com tarefas automáticas, criando *workflows* híbridos para os processos *crowdsourcing*.

Houve ainda uma tentativa de superar a limitação do *CrowdForge* relativa a não permitir a criação de *workflows* iterativos. Embora o *CrowdWeaver* também não suporte *loops*, estruturas de decisão nem fluxos alternativos, foi oferecida uma funcionalidade que permite clonar uma *microtask* e adicionar esta cópia ao *workflow*. Este sistema foi construído com base no *CrowdFlower*, de forma que as *microtasks* precisam ser modeladas com base nos recursos oferecidos por esta plataforma. Em contrapartida, é possível utilizar os *templates* de *microtask* disponíveis no *CrowdFlower*. Todavia, apesar

de ter representado uma evolução sobre o trabalho anterior, esse sistema apresenta diversas limitações.

O *CrowdWeaver* não permite a criação de *workflows* iterativos, nem que se retome a execução de uma *microtask* que já foi finalizada. Também não oferece recursos para a gerência do *workflow* ou controlar a execução das *microtasks*, nem suporta monitoramento em tempo real dos resultados agregados a cada tarefa. O *CrowdWeaver* também não oferece suporte para métodos avançados de agregação, que realizam processamento adicional das contribuições. As operações de agregação se limitam aos algoritmos básicos como os de divisão, votação, casamentos de padrão e combinação.

O *Retool* (CHEN et al., 2017) apresenta uma nova abordagem para a construção de *workflows* para processos *crowdsourcing*. No lugar de utilizar a estratégia *Map-Reduce* para modelar os *workflows*, estes são construídos a partir da combinação dos quatro tipos básicos de *workflow*.

Com base em ferramentas *Wer*, o *owner* pode criar, publicar, e liberar para execução as *microtasks* que são utilizadas para compor o *workflow*. Entretanto, não existem ferramentas para a visualização dos *workflows*, como há no *Turkomatic* e no *CrowdWeaver*.

Neste sistema também foi empregado o conceito de *task-set*, utilizado anteriormente no *Soylent* (BERNSTEIN et al., 2010) para especificar tarefas *crowdsourcing* para editoração de textos. Estes *task-sets* consistem em conjuntos de atividades, que são aplicados em cascata, e que produzem o resultado desejado para uma determinada tarefa. Este trabalho apresenta soluções para diversas limitações existentes até então, porém ainda existem alguns pontos a serem resolvidos. O *Soylent* não suporta métodos de agregação mais sofisticados e algoritmos de pós-processamento ao fim das tarefas, além de não apresentar meios para a gerência do *workflow*, permitindo o controle de fluxo entre as tarefas. Também faltam meios para monitorar o estado atual do resultado agregado, assim como para o controle interativo de transição entre as tarefas.

Na próxima seção será feita uma discussão sobre o estado da arte com base nos trabalhos analisados neste capítulo, assim como será definido o escopo do incremento proposto para o estado da arte.

3.3 Discussão sobre o Estado da Arte

A análise dos trabalhos citados neste capítulo para geração de *workflows* para processos *crowdsourcing*, assim como outros trabalhos que utilizam este tipo de *workflow* para anotação de mídia, levaram a um conjunto de recursos e características desejados no *framework* proposto. Apesar de diversas questões já terem sido resolvidas nos outros trabalhos, ainda faltam partes importantes para que se possa alcançar os objetivos apresentados nesta tese.

O estudo dos trabalhos que utilizaram *crowdsourcing* para anotação de mídia apontou que este tipo de processo poderia ser aprimorado com recursos que para o gerenciamento do *workflow*, permitindo a realizar transições entre as tarefas ativas. Adicionalmente seria de grande valia a criação de recursos que permitissem o acompanhamento, em tempo real, do estado atual do resultado agregado com o intuito de determinar o ponto de parada da coleta de dados, ou a necessidade de retornar a uma tarefa anterior.

Também se concluiu que para os trabalhos relacionados com anotação de vídeo, que é o foco desta tese, seria bastante útil um modelo geral para o *workflow*, no qual cada tarefa fosse modelada como um *task-set*. Cada *task-set* seria composto pelas atividades de coleta de contribuições, agregação e pós processamento necessários para cada tarefa.

Com base nessa análise, os pontos de incremento no estado da arte que estão sendo propostos nesta tese estão relacionados com os itens apresentados a seguir.

O primeiro ponto consiste em realizar a **Formalização do Processo**, definindo os atores e as atividades de responsabilidade de cada um deles a cada fase do processo.

O segundo ponto é a **Formalização o Workflow**, com a aplicação de um modelo que seja adequado para representar processos *crowdsourcing* baseados no cascadeamento de tarefas, sendo que o foco neste trabalho é abordar os processos relacionados com atividades complexas sobre vídeos.

O terceiro ponto de incremento no estado da arte é fornecer recursos para a **Gerência de Workflow**, permitindo o controle de fluxo entre as *microtasks*, assim como a retomada de tarefas já finalizadas e navegação interativa entre as tarefas do *workflow*.

O quarto ponto é desenvolver meios para realizar o **Controle das Transições**, de forma que, a qualquer momento, seja definido qual a tarefa ativa, possibilitando que o início e finalização das tarefas sejam realizados de forma interativa, ou automatizados por critérios de decisão.

O quinto ponto é oferecer recursos que permitam que o resultado do processo seja construído incrementalmente por meio do **Cascadeamento de Tarefas**, de forma que a agregação do resultado parcial fornecido por uma delas seja utilizada como entrada para a próxima.

O sexto ponto consiste em suportar a **Edição do Workflow em tempo de execução**, adicionando novas tarefas sem a necessidade de reiniciar o processo.

O sétimo ponto é oferecer suporte aos métodos de agregação *online*, fornecendo meios para que seja possível realizar o **Monitoramento em Tempo Real** da agregação dos resultados parciais.

O oitavo ponto é apresentar um **Modelo de Task-Set** que possibilite que para cada tarefa seja modelado um subprocesso *crowdsourcing* que possa ser executado e gerenciado de forma padrão.

Uma outra questão relacionada com o estado da arte, e coberta nesta tese, é que enquanto os trabalhos analisados formalizam apenas a modelagem das *microtasks* a etapa *crowdsourcing* do processo, o *workflow* proposto também inclui as fases de preparação e conclusão, assim como define papéis responsáveis por cada uma das atividades.

No próximo capítulo será apresentado o *framework* desenvolvido com o objetivo de contribuir com o avanço no estado da arte, de acordo com os pontos expostos neste capítulo. Será apresentada a formalização do processo e do *workflow*, assim como o ambiente que adiciona suporte computacional ao *framework*.

4 *Framework Crowdsourcing* para Suporte a Tarefas Complexas de Anotação de Vídeo

O objetivo principal desta tese é produzir um *framework* adequado para modelar e executar processos *crowdsourcing* para tarefas complexas de anotação de vídeo, utilizando como abordagem o cascadeamento de *microtasks*.

Este *framework*, batizado de *CrowdWaterfall*, é baseado na decomposição do problema principal em um conjunto de *microtasks crowdsourcing*, e reúne um modelo de *workflow* para representar o processo no qual as *microtasks* são executadas em cascata, e pelo ambiente de *software* desenvolvido para prover o suporte computacional necessário.

Conforme apresentado na seção 1.2, Mattsson (1996) e Fayad, Schmidt e Johnson (1999) definem *framework* como uma arquitetura que visa resolver um classe de problemas, e é composta pelos métodos, técnicas, processos e recursos necessários para alcançar seu objetivo. Sendo assim, o *CrowdWaterfall* se encaixa perfeitamente nessa definição de *framework*.

O ambiente computacional de suporte à proposta foi chamado de *CrowdNoter*, e oferece todos os recursos necessários para gerenciar o processo *crowdsourcing* e coletar as contribuições de múltiplas fontes, incluindo plataformas comerciais de *crowdsourcing* ou plataformas colaborativas, assim como processar e agregar as contribuições. Adicionalmente o *CrowdNoter* oferece uma biblioteca de ferramentas que podem ser utilizadas para coletar as contribuições por meio da execução da *microtasks*. Este sistema será apresentado com detalhes na seção 4.5.

A premissa do *framework* proposto é a de que é possível decompor a solução do problema principal em um conjunto de *microtasks crowdsourcing*, que são equivalentes a funções realizadas por meio de computação humana. Em seguida, estas *microtasks* são utilizadas para compor um *workflow* que é equivalente a um algoritmo para resolver o problema principal.

O processo que integra o *CrowdWaterfall* é dividido nos três estágios ilustrados na figura 1: (i) planejamento, (ii) execução, e (iii) distribuição.

- (i) No estágio de planejamento é feita a formalização do problema, sua decomposição, a geração do *workflow* e os recursos a serem utilizados são definidos e alocados.
- (ii) O estágio de execução representa o período no qual ocorre a execução efetiva do processo *crowdsourcing*. Este estágio é dividido nas mesmas três fases identificadas nos trabalhos analisados no *survey* (CHITILAPPILLY; CHEN; AMER-YAHIA, 2016): preparação, implementação, e finalização.
- (iii) O último estágio é o de distribuição, e nele o produto final é disponibilizado de acordo com as definições feitas no estágio de planejamento.

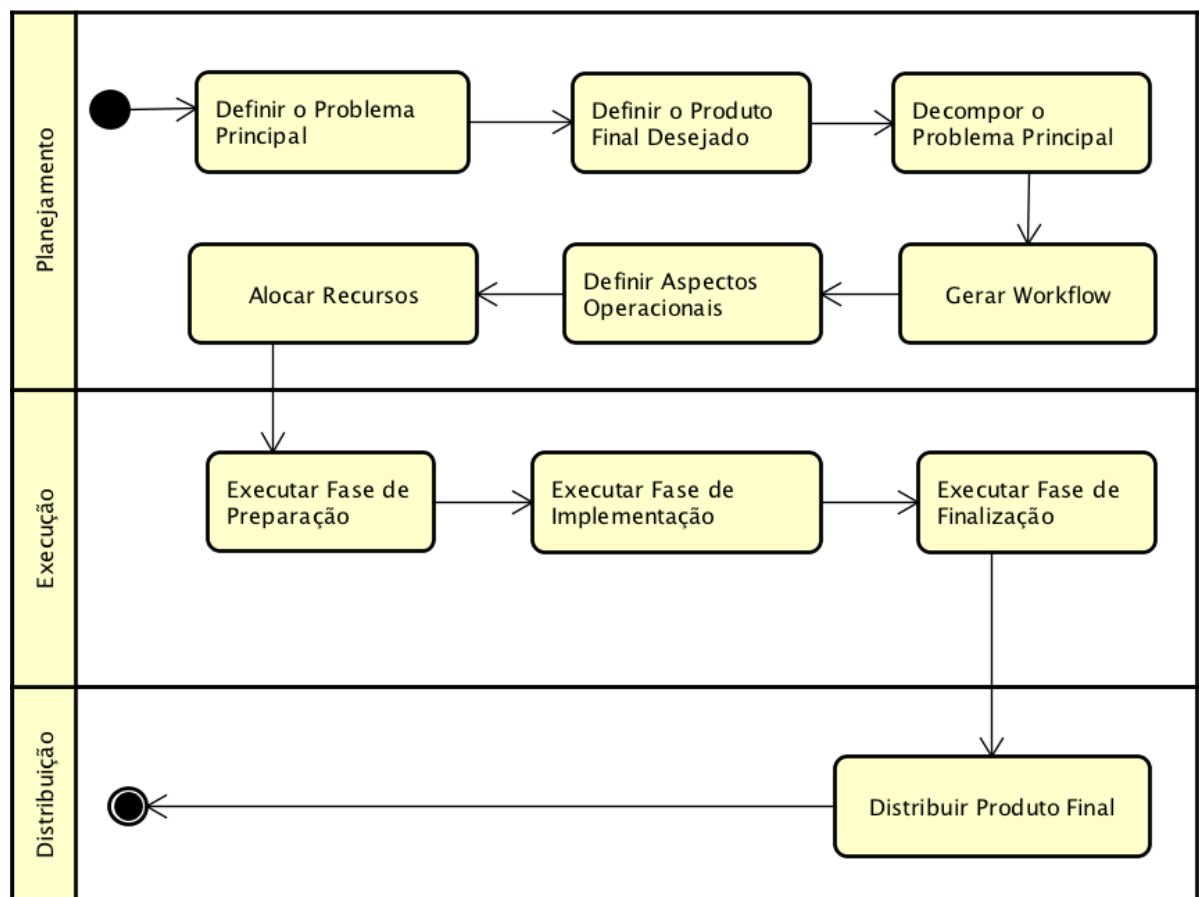


Figura 1 – Sequência de passos do processo utilizado no *CrowdWaterfall*.

4.1 Estágio de Planejamento

É no estágio de planejamento que se define o problema a ser resolvido, as *microtasks* que serão executadas pela *crowd*, o *workflow* que guiará o processo *crowdsourcing*, a forma como o resultado de cada *microtask* será agregado, e os recursos computacionais que serão empregados. A estratégia adotada é analisar a tarefa complexa de anotação de vídeo, identificando como se pode criar uma composição de tarefas simples de anotação capaz de produzir um resultado equivalente. A ideia básica é refinar e aprimorar gradativamente, ao longo das tarefas, as anotações geradas pelos *workers* até que seja gerado o resultado final.

O resultado desejado é analisado em relação às informações necessárias para gerá-lo, e então é estabelecida uma ordem de dependência para as informações necessárias. Uma vez que as informações necessárias estejam definidas, assim como a ordem de dependência entre elas, são definidas *microtasks* para se obter cada uma destas informações. Esta ordem de dependência é importante pois o resultado parcial gerado por cada *microtask* de anotação simples é utilizada como entrada para a próxima tarefa. As atividades que compõe o estágio de planejamento, que são descritas a seguir.

A primeira atividade do estágio de planejamento é **Definir o Problema Principal** a ser resolvido. Aqui é definido o cenário de aplicação, o tipo da mídia que será anotada, e os aspectos que devem ser anotados no conteúdo de mídia.

Antes de qualquer outra coisa é necessário **Definir o Produto Desejado**, ou seja, é necessário determinar quais são as anotações que se deseja obter do processo. Este passo é essencial pois as *microtasks* serão definidas e o *workflow* criado com o objetivo de gerar o produto desejado.

Uma vez que sejam definidas as anotações que são o alvo do processo, assim como os aspectos que compõem cada uma delas, chega o momento de **Decompor o Problema Principal**, levando em conta a ordem de dependência de cada um destes aspectos e de que forma eles podem ser anotados separadamente. Lembrando que no *CrowdWaterfall* cada tarefa de anotação simples aborda apenas um aspecto a ser anotado.

O objetivo deste passo é segmentar uma tarefa complexa de anotação de vídeo, que usualmente demandaria trabalho de especialistas, definindo um conjunto de *microtasks* de anotação simples que possam ser executadas por uma *crowd* heterogênea formada por *workers* com diferentes graus de conhecimento em relação ao domínio do problema.

O *CrowdWaterfall* sugere que a decomposição seja realizada de acordo com a estratégia *sequencial* apontada em (CHITILAPPILLY; CHEN; AMER-YAHIA, 2016). Nesta estratégia as *microtasks* devem ser projetadas para serem executadas sequencialmente, em cascata, de forma que os resultados sejam complementados, aprimorados e refinados até alcançar o produto final.

O próximo passo é **Gerar o Workflow**. Com as *microtasks* devidamente definidas, é necessário determinar as informações de entrada necessárias para cada uma delas, assim como a saída que pode ser produzida por cada uma delas. Dessa forma é possível determinar uma ordem de dependência entre as *microtasks*, e elas podem ser organizadas em um *workflow* que guiará o processo de anotação.

De acordo com este *workflow* as *microtasks* são executadas seguindo um fluxo em cascata, de maneira que a saída de uma *microtask* serve de entrada para a próxima *microtask* do *workflow*.

Uma vez que já se conhece o número e a natureza das *microtasks*, é necessário **Definir os Aspectos Operacionais** do processo. Estes aspectos incluem o orçamento disponível para o processo, a fonte das contribuições (ex. plataforma de *crowdsourcing*, redes sociais, etc), e mesmo o perfil exigido dos *workers* para as tarefas.

Com os recursos necessários para o processo, é necessário providenciá-los, ou seja, **Alocar os Recursos** que serão utilizados para recompensar os *workers* e taxas das plataformas de *crowdsourcing*.

4.2 Estágio de Execução

No estágio de execução, as anotações desejadas como resultado do processo são, efetivamente, produzidas a partir de um processo *crowdsourcing*. Nele é executado o *workflow*, gerado no estágio de planejamento, para processo de cascadeamento de *microtasks*. Este estágio segue o modelo observado do *survey* (CHITILAPPILLY; CHEN; AMER-YAHIA, 2016) para os processos *crowdsourcing* em geral. Este modelo apresenta três fases bem definidas, e que podem ser observadas na figura 2: (i) preparação, (ii) implementação, e (iii) finalização. Vale destacar que de acordo com o *CrowdWaterfall*, todas as atividades *crowdsourcing* se concentram neste estágio.

O modelo seguido no estágio de execução está em convergência com o que é apresentado no trabalho (LUZ; SILVA; NOVAIS, 2015), um *survey* que apresenta modelos e processos relevantes que são utilizados nas abordagens *crowdsourcing*, incluindo os cenários em que são necessárias tarefas complexas de anotação.

A fase de *preparação* é responsável por criar as condições para que as *microtasks* possam ser realizadas. Nesta fase o conteúdo a ser anotado é preparado para ser enviado aos *workers* e, se possível, é definido o *groundtruth* para o processo. Também nessa fase os recursos alocados no estágio de planejamento são direcionados para cada *microtask*, e as campanhas de coleta de contribuições são criadas na plataforma de *crowdsourcing*. A fase de *implementação* é quando as tarefas *crowdsourcing* são efetivamente realizadas. Nela, as *microtasks* são executadas de acordo com o *workflow* definido no estágio de planejamento. O subprocesso de execução para cada uma das *microtasks* é modelado de acordo com *task-set* padrão que será apresentado na seção 4.2.1. A fase de *finalização* é quando o produto final é gerado com base nas anotações obtidas na fase de implementação. Após a geração, o produto final é avaliado para ser disponibilizado para distribuição, ou o processo é retomado para que os resultados sejam refinados.

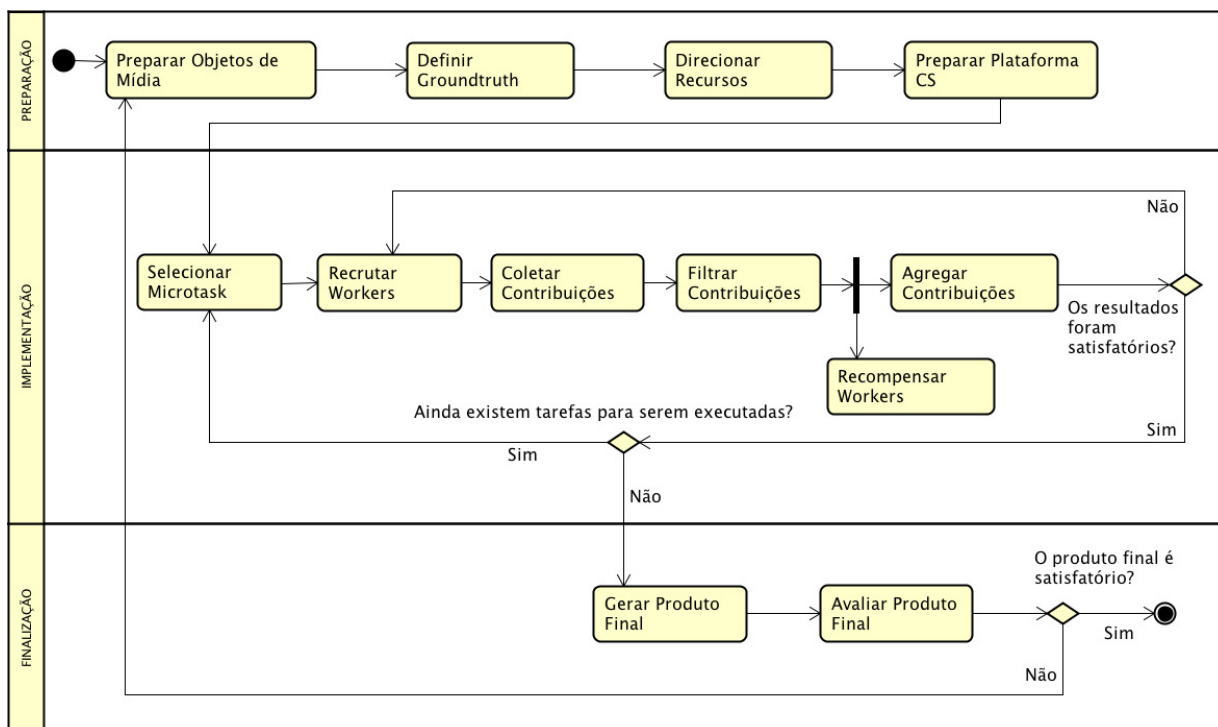


Figura 2 – Processo *crowdsourcing* baseado no cascadeamento de *microtasks*.

4.2.1 Task-Sets: subprocessos crowdsourcing

De acordo com o *CrowdWaterfall*, cada uma das *microtasks* que compõem o *workflow* são executadas em um *task-set*, que é um conjunto de atividades a partir do qual cada tarefa é modelada como um subprocesso independente. Todavia, os *task-sets* são executados sequencialmente para que a saída de um deles sirva como entrada para o próximo.

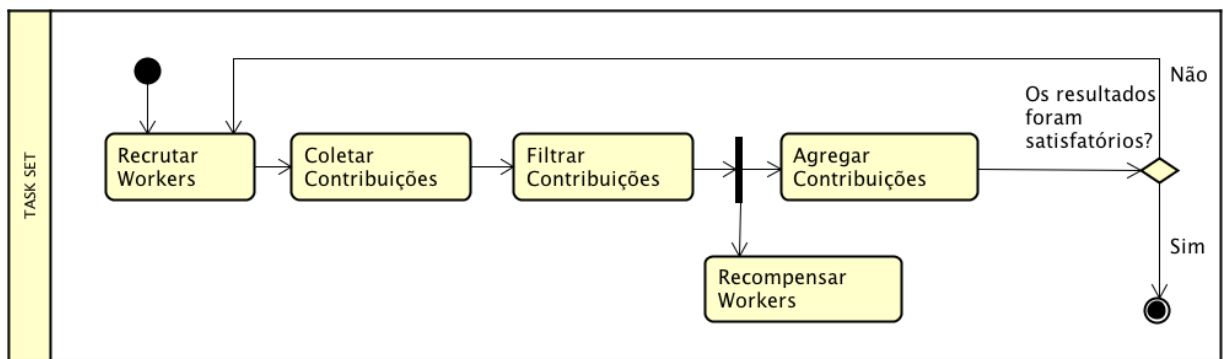


Figura 3 – Estrutura de um subprocesso *task-set* referente à uma *microtask*.

Este modelo de processo permite que cada *microtask* seja um subprocesso *crowdsourcing* completo de anotação de mídia, e que ao longo do cascadeamento as anotações sejam refinadas, complementadas e aprimoradas até chegar no resultado desejado.

Cada *task-set* segue o subprocesso padrão representado na figura 3, no qual os *workers* são selecionados de acordo com os requisitos mínimos da tarefa, como domínio do idioma utilizado no conteúdo a ser anotado ou mesmo o acesso à recursos técnicos como fones de ouvido ou velocidade mínima de internet. Na sequência são coletadas as contribuições da *crowd*, que são filtradas, por questões de confiabilidade, segundo critérios definidos no estágio de planejamento.

Neste ponto os *workers* podem ser recompensados, e o motivo desta atividade estar localizada após a filtragem é para dar a possibilidade de recompensar apenas os *workers* não maliciosos caso se deseje adotar esta política. Também após a filtragem é possível agregar as contribuições gerando um resultado parcial. Caso o resultado seja satisfatório o processo segue para o próximo *task-set*, e caso ainda não se tenha chegado ao resultado desejado para a tarefa atual, seleciona-se outros *workers* e continua-se a coleta de contribuições.

A modelagem de cada *task-set* como um subprocesso independente permite inclusive utilizar fontes de contribuição diferentes para cada *microtask* e ainda manter a integridade do processo. Por exemplo, pode-se utilizar um ambiente comercial de *crowdsourcing* para executar um *task-set* e um grupo interno com *workers* mais especializados para executar uma próxima tarefa.

4.2.2 Atores Envolvidos no Processo

O estudo da literatura revelou os atores envolvidos no processo, assim como as responsabilidades e atividades que cada um deles deve realizar. Esses atores são apresentados a seguir.

O **Owner** é o iniciador do processo, ele é responsável por prover os meios para que os *workers* sejam recompensados, fornecer os objetos de mídia que serão anotados, e configurar o ambiente de *crowdsourcing*. O *owner* pode ser um indivíduo, uma instituição, ou mesmo uma equipe multidisciplinar que inclui especialistas no domínio do projeto e profissionais capacitados para desenvolver os recursos computacionais necessários para o processo.

Apesar de muitas vezes o **Gerente do Processo Crowdsourcing (GPC)** e o *owner* serem a mesma pessoa, trata-se de um papel diferente, com atribuições diferentes. O *GPC* é responsável por gerenciar o processo *crowdsourcing*, é ele que monitora as contribuições coletadas até o momento e o estado atual do resultado agregado. Sendo assim é ele que dispara as ações de filtragem, agregação e determina se o processo deve seguir para o próximo *task-set* ou continuar na tarefa atual. Também é responsabilidade do *GPC* iniciar a geração do resultado final ao fim da última *microtask*.

A **Plataforma Crowdsourcing (Plataforma CS)** é utilizada no *CrowdWaterfall* de uma forma diferente do tradicional, em uma estratégia leve. Isso significa que se utiliza apenas a *crowd* cadastrada na plataforma, e não os outros recursos que ela oferece. Dessa forma, as atribuições da *Plataforma CS* são basicamente selecionar os *workers* que atendem aos requisitos mínimos e intermediar a recompensa deles. Vale reforçar que o papel *Plataforma CS* se refere somente às plataformas que oferecem as funcionalidades de recrutar e recompensar os *workers*, e não a qualquer sistema que implementa conceitos de *crowdsourcing*, logo, o *CrowdNoter*¹ não se encaixa neste papel.

¹ *CrowdNoter* é o sistema desenvolvido para apoiar o *CrowdWaterfall*.

O **Worker** é a força de trabalho que viabiliza os processos crowdsourcing. Sendo assim, sua atribuição é executar as tarefas, fornecendo contribuições que são agregadas para formar o resultado. Os *workers* podem ser membros registrados nas plataformas comerciais de *crowdsourcing*, integrantes de grupos de pesquisa, ou mesmo colaboradores voluntários atendendo a chamados divulgados em meios como as redes sociais.

4.2.3 Fase de Preparação

A fase de preparação é composta por quatro de atividades principais, que são executadas pelo *owner* do processo, e que estão relacionadas com a configuração do ambiente no qual será realizado o processo *crowdsourcing*. Estas atividades podem ser decompostas em sub-tarefas de acordo com a natureza da mídia a ser anotada ou do resultado a ser gerado, mas o *CrowdWaterfall* sugere que sejam realizadas nesta sequência. Conforme pode ser observado na figura 2, estas quatro atividades são: (i) preparar os objetos de mídia, (ii) definir *groundtruth*, (iii) direcionar recursos, e (iv) preparar a plataforma de *crowdsourcing*.

Vale reforçar que, conforme é exposto na tabela 12, as *microtasks* já foram definidas e o *workflow* já foi gerado no estágio de planejamento. Assim, o objetivo desta fase de preparação é configurar o ambiente de acordo com o que foi planejado.

Tabela 12 – Atividades e responsabilidades na fase de preparação.

Papel	Responsabilidades	Atividades
Owner	Preparar o ambiente para que possa ser realizado o processo <i>crowdsourcing</i> , o que inclui fornecer todos os recursos necessários, sejam tecnológicos ou financeiros.	Preparar os Objetos de Mídia
		Definir <i>Groundtruth</i> ou <i>Forma de Avaliação</i>
		Direcionar os Recursos
		Preparar a Plataforma de <i>Crowdsourcing</i>

A primeira atividade de responsabilidade desta fase é **Preparar os Objetos de Mídia**. O *CrowdWaterfall* sugere que os *workers* recebam, a cada tarefa, o mínimo possível de conteúdo para anotar. Dessa forma, é necessário que o *owner* registre no sistema os objetos de mídia a serem anotados, e realize o agrupamento e segmentação dos objetos de mídia para que possam ser distribuídos.

Os critérios utilizados para determinar o conteúdo distribuído em cada tarefa variam de acordo com o tipo de mídia e o tipo de anotação, mas deve-se buscar determinar o conteúdo mínimo que pode ser anotado sem que haja problemas de contexto que prejudiquem a anotação. Por exemplo, se o objetivo é anotar um vídeo, deve-se determinar os menores segmentos que ainda preservam o contexto do conteúdo, ou se o objetivo é anotar um *dataset* de imagens, deve-se evitar enviar mais de uma imagem para ser anotada a cada tarefa.

Contudo existem casos mais complexos, como em uma tarefa em que se deseja localizar um ponto de sincronia entre dois vídeos. Neste caso seria necessário determinar pares de segmentos de vídeo para serem distribuídos em cada tarefa, e estes pares de segmentos devem ser definidos de forma a aumentar a chance de existir um ponto de sincronização entre eles. Por isso esta fase é essencial para a produção de resultados satisfatórios e deve receber a devida atenção.

A próxima atividade é **Definir *Groundtruth* ou Forma de Avaliação**. Nos casos em que é possível definir um *groundtruth* para a checagem de confiabilidade das contribuições é neste ponto que se registra o *gold standard* que será utilizado para esta finalidade. O *gold standard* pode ser definido tanto manualmente quanto por meio de uma ferramenta de anotação. A existência de um *groundtruth* para uma tarefa torna possível a aplicação de um método de agregação *HoneyPot* no lugar de um método *Major Decision* simples.

Também é de responsabilidade do *owner*, na fase de preparação, **Direcionar os Recursos**. No estágio de planejamento foi definido o orçamento disponível para recompensar os *workers*, e os recursos necessários foram alocados. Nesta atividade os recursos alocados são direcionados para cada *microtask*. Para cada uma das campanhas criadas na plataforma de *crowdsourcing* devem ser depositados fundos suficientes para cobrir os custos referentes ao número máximo de contribuições autorizadas e os custos de operação cobrados pela plataforma. Nos cenários em que a recompensa não é financeira, esta atividade diz respeito a estabelecer meios para registrar os créditos de cada *worker* e viabilizar a comunicação da plataforma de *crowdsourcing* com o sistema de recompensa.

A última atividade desta fase é **Preparar a Plataforma de *Crowdsourcing***. Uma vez que as *microtasks* já foram definidas, assim como as ferramentas de anotação que serão utilizadas para executá-las, é necessário criar uma campanha para cada *microtask* na plataforma de *crowdsourcing*. Uma campanha é o processo de coleta de contribuições para uma *microtask*. Para criar uma campanha é necessário definir a quantidade máxima

autorizada de contribuições, assim como o valor que será pago aos *workers* por tarefa realizada. Também é durante a criação da campanha que se atribui às *microtasks* as suas respectivas ferramentas de anotação.

4.2.4 Fase de Implementação

Na fase de implementação ocorre o cascadeamento das *microtasks* por meio da execução sucessiva e iterativa dos *task-sets*. O cascadeamento de *microtasks* é o que viabiliza a geração de anotações, que usualmente demandariam tarefas complexas, a partir de contribuições simples obtidas de uma *crowd* heterogênea, sem a necessidade de conhecer o grau de expertise de cada *worker*. O *workflow* definido no estágio de planejamento guia esta fase do processo, mostrando a ordem em que as *microtasks* devem ser realizadas para que sejam fornecidas as entradas adequadas para cada um dos *task-sets*.

A coleta das contribuições da *crowd* ocorre apenas nesta fase, o que significa que todas as atividades realizadas pelos *workers* se concentram nesta fase. Todavia nela também existem atividades que são de responsabilidade do *GPC* e da *Plataforma CS*. As responsabilidades e as tarefas atribuídas a cada papel estão sumarizados na tabela 13.

Tabela 13 – Atividades e responsabilidades na fase de execução.

Papel	Responsabilidades	Atividades
Plataforma Crowdsourcing (Plataforma CS)	Intermediar as relações com os <i>workers</i> .	Recrutar os Workers
		Recompensar os Workers
Gerente do Processo Crowdsourcing (GPC)	Gerenciar o <i>workflow</i> do processo; Iniciar as filtragens e agregações; Controlar as transições entre tarefas.	Filtrar as Contribuições
		Agregar as Contribuições
		Gerar o Resultado Parcial
Worker	Fornecer as contribuições para o processo.	Executar as Microtasks

A primeira atividade desta fase é **Recrutar os Workers**. A *Plataforma CS* recruta os possíveis *workers* de acordo com as definições feitas pelo *owner* no estágio de planejamento, e divulga para eles as *microtasks* disponíveis.

A segunda atividade consiste em **Recompensar os Workers**. Após o consentimento do *GPC*, a *Plataforma CS* recompensa os *workers* pelas tarefas realizadas. Novamente, a recompensa é feita conforme foi estabelecido no estágio de planejamento.

A quarta atividade é **Filtrar as Contribuições**. O *GPC* dispara o processo de filtragem sobre as contribuições coletadas. Caso exista um *Gold Standard* para a tarefa em curso, ele é utilizado no processo de filtragem, viabilizando a utilização de um método de agregação *HoneyPot*.

A quinta atividade é **Agregar as Contribuições**. Uma vez que o *CrowdWaterfall* sugere a utilização de métodos de agregação *online*, o *GPC* pode acompanhar em tempo real o estado atual do resultado agregado. Dessa forma, ele pode decidir se é necessário continuar com a coleta de contribuições, ou se pode efetivar o processo de agregação e realizar a transição para a próxima tarefa.

A sexta atividade é **Gerar o Resultado Final**. Uma vez que todas as *microtasks* foram finalizadas, o *GPC* pode efetivar a agregação do resultado da última tarefa. Neste momento as anotações obtidas pelo processo são armazenadas na forma de metadados, e ficam disponíveis para serem finalizadas na fase de conclusão.

A sétima, e última, atividade desta fase é **Executar as Microtasks**. Os *workers* são responsáveis por realizar cada uma das tarefas previstas no *workflow* do processo de cascadeamento de *microtasks*. Apesar da contribuição individual não ter seu peso diluído no processo de agregação, todas as contribuições são importantes para a geração do resultado, independente do grau de expertise do *worker*.

4.2.5 Fase de Finalização

Nesta fase, os metadados resultantes do cascadeamento das *microtasks* são utilizados para gerar os artefatos de saída definidos pelo *owner* no estágio de planejamento. Uma vez que os artefatos de saída são gerados, o *owner* os avalia de acordo com os critérios definidos no estágio de planejamento. Caso estejam satisfatórios, o produto final é liberado e tem início o estágio de distribuição. Se o *owner* avaliar que o produto gerado não atende as especificações, o processo retorna para o estágio de planejamento, para que seja definido quais *microtasks* devem ser executadas novamente, e se os métodos de agregação devem ser ajustados. As responsabilidades atribuídas a cada papel nesta fase estão sumarizados na tabela 14.

Tabela 14 – Atividades e responsabilidades na fase de conclusão.

Papel	Responsabilidades	Atividades
Gerente do Processo Crowdsourcing (GPC)	Iniciar o mecanismo de geração do produto final.	Gerar o Produto Final
Owner	Avaliar os artefatos de saída que serão distribuídos.	Avaliar o Produto Final

A primeira atividade da fase de finalização é **Gerar o Produto Final**. Após ao final do cascadeamento das *microtasks*, as anotações estão armazenadas na forma de metadados. Desta forma o *GPC* pode iniciar o processo de geração do produto final. O produto final é definido no estágio de planejamento, e sua geração consiste em exportar as anotações para o formato desejado. Em alguns casos o produto final pode ser inclusive o próprio conjunto de metadados distribuído pelos meios definidos pelo *owner* no estágio de planejamento.

A outra atividade desta fase consiste em **Avaliar o Produto Final**. Uma vez que o produto final seja gerado, o *owner* deve avaliar se ele atende as especificações definidas no estágio de planejamento. Esta avaliação pode ser feita em relação ao conteúdo ou a forma do produto. Adicionalmente, é possível utilizar a própria *crowd* para avaliar este produto, sendo necessário prever esta tarefa no estágio de planejamento para que ela seja incluída no *workflow* do processo de cascadeamento de *microtasks*.

4.3 Estágio de Distribuição

O resultado final é gerado ao longo do processo de cascadeamento de *microtasks*, em que o resultado de cada tarefa é agregado, gerando a entrada da próxima tarefa. Assim, o processo de agregação da última *microtask* gera o resultado final. O resultado final do processo *crowdsourcing* é um conjunto de *metadados* que são utilizados para gerar artefatos adequados para distribuição. O estágio de distribuição é responsável por disponibilizar o produto final, podendo ser exibido em um *player*, ou exportado para diferentes formatos, uma vez que as anotações são armazenadas como metadados oferecendo esta flexibilidade.

Uma vez que o *framework* define que para cada *microtask* seja gerado um resultado a partir da agregação das contribuições, é possível ainda disponibilizar estes resultados parciais como artefatos secundários gerados ao longo do processo. Por exemplo, em um cenário no qual se deseja identificar quais animais aparecem em um vídeo e suas

características, pode ser usado um processo de duas tarefas. A primeira poderia resultar em uma lista de instantes em que aparecem animais no vídeo, e a segunda recebe esta lista e resulta na identificação dos animais.

4.4 Agregação Temporal para Anotação Crowdsourcing de Vídeo

O vídeo é um tipo mídia dinâmico, que apresenta um fluxo de informações audiovisuais dependentes do tempo, e esta característica faz com que a localização temporal das anotações de vídeo seja tão importante quanto o seu conteúdo. Assim, também é necessário aplicar técnicas adequadas para determinar a localização temporal de cada uma das anotações obtidas no processo *crowdsourcing*, assim como para determinar se há concordância da crowd sobre o que foi anotado ao longo da linha do tempo do vídeo.

O estudo de trabalhos relacionados revelou os tipos mais comuns de anotação de vídeo obtidas por *crowdsourcing*, assim como as principais formas de agregação dos resultados individuais das execuções das tarefas pelos *workers*. O conhecimento adquirido foi aplicado em nossos trabalhos anteriores, o que nos levou a um procedimento formado por três passos: (i) agrupamento, (ii) filtragem e (iii) agregação. Este procedimento é descrito neste artigo, que visa apresentar como ele pode ser utilizado na agregação temporal das anotações de vídeo obtidas em campanhas *crowdsourcing*.

A agregação temporal apresenta algumas especificidades que dizem respeito tanto à natureza contínua da linha do tempo, e como o fator humano pode influenciar no posicionamento temporal da anotação, quanto à diferença no processo de agregação de instantes e intervalos de tempo. Neste capítulo são apresentadas técnicas e algoritmos diferentes para lidar com instantes e intervalos de tempo, que serão demonstrados por meio do relato de uma série de estudos de caso realizados em nossos trabalhos anteriores.

As anotações de vídeo coletadas nas campanhas *crowdsourcing* refletem opiniões individuais dos *workers* sobre o conteúdo anotado. Para que se possa extrair a inteligência coletiva que representa o consenso, ou convergência, da *crowd* sobre a anotação dos vídeos é necessário realizar a agregação das contribuições individuais. Todavia, as contribuições devem ser agregadas de forma contextualizada. Não basta aplicar uma função de agregação sobre o conjunto de contribuições coletadas da *crowd*. É necessário que antes estas contribuições sejam agrupadas de forma a produzir conjuntos formados pelas contribuições relacionadas com um mesmo item anotado.

Por exemplo, quando se quer identificar ocorrências de um determinado eventos em um vídeo, não basta calcular a média dos instantes fornecidos nas suas contribuições, é necessário antes agrupar estes instantes de forma que as anotações feitas para cada ocorrência do evento sejam colocadas no mesmo conjunto. Assim, é possível calcular o instante que representa a percepção da *crowd* para cada ocorrência do evento.

Neste momento é importante definir o termo *item de anotação*, que se refere ao que foi solicitado que o *worker* anotasse. Um item de anotação pode ser um evento, um ponto de interesse, um objeto de cena, ou qualquer outra coisa que se deseja solicitar que os *workers* anotem no vídeo.

Em grande parte dos processos *crowdsourcing* não existe essa necessidade, pois é conhecido previamente o item sobre o qual as contribuições são coletadas, porém o vídeo é uma mídia dinâmica e apresenta conteúdo variando continuamente com o tempo. Sendo assim, é possível que os *workers* anotem, para um mesmo item, instantes próximo mas não idênticos. Por isso é importante definir métodos de agrupamento que sejam capazes de determinar quais anotações devem ser colocadas a cada conjunto associado a cada item de anotação.

A necessidade de agrupar as contribuições, de forma que cada conjunto será agregado em uma anotação resultante da agregação, traz ainda a necessidade de um tipo específico de filtragem.

A filtragem é uma atividade importante e frequentemente empregada nos processos *crowdsourcing*, sendo utilizada tanto para detecção de contribuições maliciosas quanto para exclusão de entradas inconsistentes, fora do padrão ou inadequadas. Contudo, existe um outro tipo de filtragem que pode ser importante para a agregação temporal, e que consiste em eliminar as contribuições que estão isoladas sobre a linha do tempo de um vídeo, ou seja, descartar as contribuições que ficariam isoladas na atividade de agrupamento.

Este tipo de filtragem não substitui a filtragem tradicional, baseada no conteúdo das contribuições, mas atua como um filtro complementar que visa manter apenas os conjuntos de contribuições que darão origem a anotações que refletem o consenso ou convergência da opinião da *crowd* sobre a existência de um item ou a relevância em anotá-lo. Reforçando que estes itens se referem ao que foi solicitado que os *workers* notassem, como eventos, pontos de interesse ou objetos de cena.

As contribuições coletadas dos *workers* nas atividades que requerem agregação temporal podem estar associadas a instantes ou a intervalos temporais. A diferenciação destes tipos é importantes pois existem técnicas e algoritmos específicos para agrupar, filtrar e agregar cada um deles. Sendo assim, o procedimento que está sendo proposto nesta tese para agregar temporalmente as anotações de vídeo coletadas da *crowd* segue três passos:

- **Passo 1 - Agrupamento:** as anotações são agrupadas de forma que cada conjunto obtido se refira a um item anotado.
- **Passo 2 - Filtragem:** as contribuições isoladas e grupos pouco representativos são eliminados, mantendo apenas os conjuntos que representem o consenso ou convergência da opinião da *crowd* sobre a existência, ou relevância, de um item anotado.
- **Passo 3 - Agregação:** cada agrupamento de contribuições, resultante dos passos anteriores, alimenta uma função de agregação que processa a posição temporal das contribuições para determinar o instante ou intervalo que deve estar associado a ele. Lembrando que, neste trabalho, nosso foco é especificamente a posição temporal das contribuições e não o seu conteúdo.

O agrupamento temporal das contribuições tem como objetivo determinar os itens anotados pelos *workers* ao longo da linha do tempo do vídeo, e associar a cada um destes itens, as contribuições referentes a ele. Uma vez que a linha do tempo dos vídeos é virtualmente contínua, é esperado que diferentes *workers* identifiquem um mesmo evento, ou outro item a ser anotado em posições próximas, mas não necessariamente idênticas, sobre a linha do tempo.

Sendo assim, é necessário aplicar técnicas de agrupamento para identificar as contribuições que se referem a um mesmo item de anotação. Para reforçar este conceito é necessário lembrar que o resultado que se espera da agregação temporal é a lista de porções do vídeo sobre os quais as anotações devem ser associadas. De acordo com as anotações desejadas, as contribuições coletadas dos *workers* podem estar associadas a instantes e intervalos. Esta distinção é importante pois para cada um destes tipos será utilizado um método de agrupamento diferente. Analogamente, a filtragem e os critérios de agregação também são específicos para instantes e para intervalos.

4.4.1 Método para Agregação de Instantes

O agrupamento de instantes utilizado nesta pesquisa se baseia na premissa de que, mesmo para itens próximos, duas marcações consecutivas para o mesmo item tendem a estar mais próximas que marcações feitas para itens de anotação consecutivos. Esta premissa se baseia no fato de que, ao identificar um item de anotação, os *workers* tentarão anotá-lo no instante em que perceberam, com margem de erro determinada por diferentes questões como *delay* do sistema de anotação, assim como o nível de reflexo e coordenação motora de cada anotador. Todavia, espera-se que essa margem de erro seja menor que a distância média entre os itens de anotação. Sendo assim, o algoritmo que propomos para realizar o agrupamento de instantes segue o Algoritmo 1. Este algoritmo se baseia na medida da distância média entre as anotações consecutivas, que será chamado de Δ , para em seguida agrupá-las utilizando o Δ como margem de tolerância.

Algoritmo 1: AGRUPAMENTO TEMPORAL DE INSTANTES.

```

input : List of Contributions ( $C$ ), tolerance factor ( $f$ )
output : List of contributions groupings ( $G$ )
1 begin
2    $S \leftarrow \text{Sort}(C)$ 
3    $Total \leftarrow 0$ 
4   for  $i \leftarrow 2 : \text{length}(S)$  do
5      $\delta \leftarrow \text{instant}(S[i]) - \text{instant}(S[i - 1])$ 
6      $Total \leftarrow Total + \delta$ 
7   end
8    $\Delta \leftarrow Total / (\text{length}(S) - 1)$ 
9    $G = []$ 
10   $g = []$ 
11  for  $i \leftarrow 2 : \text{length}(S)$  do
12     $\delta \leftarrow \text{instant}(S[i]) - \text{instant}(S[i - 1])$ 
13    if  $\delta \leq (\Delta \cdot f)$  then
14       $\text{add}(g, S[i - 1])$ 
15    end
16    else
17       $\text{add}(G, g)$ 
18       $g = []$ 
19    end
20  end
21 end
22 return  $G$ 

```

O algoritmo tem início com a ordenação da lista de contribuições coletadas dos *workers* (linha 2). O próximo passo é calcular o Δ , o que é feito somando a distância entre todos os pares consecutivos de contribuições (linhas 4 a 7), e em seguida dividindo este somatório pela quantidade de pares (linha 8). Uma vez que o valor do Δ é conhecido, o próximo passo é agrupar as contribuições. A lista G (linha 9) começa vazia, e ao final do procedimento cada uma de suas posições deve conter um dos agrupamentos de contribuições. Cada um dos agrupamentos é acumulado na lista g (linha 10).

Como feito anteriormente, é calculada a diferença entre cada par consecutivo de contribuições (linha 12), e as contribuições são adicionadas em g (linha 14) até que a distância entre uma contribuição e a próxima seja maior que o Δ (linha 16). Quando isso acontece, o agrupamento g é adicionado em uma das posições de G (linha 17), e é iniciado um novo agrupamento g (linha 18). Dessa forma, quando o procedimento chega ao final, a lista G está preenchida por todos os agrupamento de contribuição.

O parâmetro f , que é o fator de tolerância, pode ser manipulado para definir grupos mais restritos ou mais abrangentes, de acordo com cada caso. Além disso, é possível ainda aplicar este algoritmo recursivamente dentro dos grupos, criando subgrupos ou refinando os agrupamentos.

Uma vez que o Algoritmo 1 foi utilizado para agrupar as contribuições em relação aos instantes a que cada uma delas se refere, é necessário filtrar os agrupamentos. Como cada um dos agrupamentos irá resultar em uma anotação após a agregação, é necessário eliminar os agrupamentos que não representam a convergência ou o consenso dos *workers*. Para isso, são eliminados os grupos contendo apenas 1 item de anotação, que representam a opinião isolada de um *worker*, assim como os agrupamentos pequenos, que representam grupos pouco expressivos em relação ao total de contribuições da *crowd* associadas àquele item.

A tolerância sugerida nesta pesquisa para o filtro utilizado é igual ao tamanho médio dos agrupamentos. O algoritmo utilizado calcula a quantidade média de contribuições contidas em cada grupo e, com base nesta quantidade, elimina os agrupamentos pouco expressivos. O parâmetro f é o fator de tolerância, de forma que é possível adaptar o rigor do algoritmo na eliminação de agrupamentos.

Finalmente, a agregação é realizada calculando o ponto médio de cada um dos agrupamento, utilizando assim uma estratégia baseada em *Majority Decision*. Na figura 4 é possível observar o resultado deste método sobre as contribuições coletadas na primeira *microtask* do estudo de caso apresentado na seção 5.3.

4.4.2 Método para Agregação de Intervalos

O método de agregação segue um caminho diferente do aplicado na *microtask* anterior. O agrupamento das contribuições é feito com base na intersecção das contribuições, e em seguida a função de agregação determina a região de convergência em cada agrupamento. Este procedimento é realizado por meio do Algoritmo 2.

Esse algoritmo recebe como parâmetro uma lista de intervalos, e retorna um intervalo agregado obtido a partir deles. Para entender esta estratégia basta ter em mente que cada lista de intervalos é obtida por meio de contribuições sobre um segmento de vídeo, por meio de uma *microtask* que, de modo geral, consiste em ajustar os limites do segmento de acordo com o que é solicitado na descrição da tarefa.

O algoritmo tem início ordenando os segmentos ajustados enviados pelos *workers* (linha 2). Em seguida, os segmentos são agrupados utilizando como critério a ocorrência de intersecção, ou seja, em cada agrupamento existem apenas segmentos que possuem intersecção. Este procedimento pode ser observado nas linhas de 7 a 13.

A ordenação inicial facilita a detecção de intersecção por transitividade, o que significa que se um intervalo A tem intersecção com um intervalo B , e B possui intersecção com um intervalo C , então A , B e C serão colocados no mesmo agrupamento. Uma vez que o algoritmo armazena na lista G todos os agrupamentos, o próximo passo é realizar a agregação dentro de cada um deles. Isso pode ser observado nas linhas de 14 a 44. Nas linhas de 18 a 25 são definidos o início e o final do intervalo associado a cada um dos agrupamentos.

Em seguida, nas linhas de 27 a 34, é determinado, para cada agrupamento, a intersecção máxima entre as linhas que o compõe. Vale observar que as observações para verificar as intersecções são feitas em passos de 0.05s para efeito de padronização. Nas linhas de 36 a 40 são eliminados os trechos do segmento que possuem grau de intersecção inferior a metade do grau máximo para o agrupamento. O segmento agregado contendo o intervalo que representa a *Majority Decision* para o agrupamento. Na figura 5 é possível observar o resultado deste método sobre as contribuições coletadas na segunda *microtask* do estudo de caso apresentado na seção 5.3.

Algoritmo 2: AGRUPAMENTO TEMPORAL DE INTERVALOS.

```

input  : List of segments ( $S$ )
output : Aggregated segments ( $R$ )
1 begin
2    $S \leftarrow \text{Sort}(S)$ 
3    $G \leftarrow []$ 
4    $G[0] \leftarrow []$ 
5    $\text{add}(G[0], S[0])$ 
6    $\text{current} \leftarrow 0$ 
7   for  $i \leftarrow 1 : \text{length}(S)$  do
8     if  $S[i-1].\text{end} < S[i].\text{start}$  then
9        $\text{current}++$ 
10       $G[\text{current}] \leftarrow []$ 
11    end
12     $\text{add}(G[\text{current}], S[i])$ 
13  end
14   $R \leftarrow []$ 
15  for  $i \leftarrow 0 : \text{length}(G)$  do
16     $\text{start} \leftarrow G[i][0].\text{start}$ 
17     $\text{end} \leftarrow G[i][0].\text{end}$ 
18    for  $j \leftarrow 0 : \text{length}(G[i])$  do
19      if  $G[i][j].\text{start} < \text{start}$  then
20         $\text{start} \leftarrow G[i][j].\text{start}$ 
21      end
22      if  $G[i][j].\text{end} > \text{end}$  then
23         $\text{end} \leftarrow G[i][j].\text{end}$ 
24      end
25    end
26     $\text{counts} \leftarrow []$ 
27    for  $\text{step} \leftarrow \text{start}, \text{max} \leftarrow 0; \text{step} \leq \text{end}; \text{step} += 0.05$  do
28      for  $j \leftarrow 0, \text{count} \leftarrow 0; j < \text{length}(G[i]); j++$  do
29        if  $\text{count} > \text{max}$  then
30           $\text{max} \leftarrow \text{count}$ 
31        end
32         $\text{add}(\text{counts}, ("step" : \text{step}, "count" : \text{count}))$ 
33      end
34    end
35  end
36  for  $j \leftarrow 0; j < \text{length}(\text{counts}); j++$  do
37    if  $\text{max}/2 > \text{counts}[j].\text{count}$  then
38       $\text{del}(\text{counts}, j)$ 
39       $j--$ 
40    end
41  end
42   $\text{start} \leftarrow \text{counts}[0].\text{step}$ 
43   $\text{end} \leftarrow \text{counts}[\text{length}(\text{counts}) - 1].\text{step}$ 
44   $\text{add}(R, ("start" : \text{start}, "end" : \text{end}))$ 
45 end
46 return  $R$ 

```

4.5 O Sistema *CrowdNoter*

O sistema *CrowdNoter* foi construído para dar suporte ao processo de cascadeamento de *microtasks* definido para operacionalizar o *framework CrowdWaterfall*. O *CrowdNoter* oferece os recursos necessários para que o *owner*, o *GPC*, e os *workers* realizem todas as atividades que são de sua responsabilidade, além de oferecer os recursos de integração necessários para que as atividades realizadas pela *CS Platform* sejam incorporadas no *workflow* do processo.

Este sistema é capaz de controlar o estado do processo a partir de recursos para a gerência do *workflow*, além disso ele está preparado para receber contribuições de diferentes fontes, desde plataformas comerciais de *crowdsourcing*, até *websites* e redes sociais. Dessa forma, o *CrowdNoter* permite que todo o fluxo de informação do processo *crowdsourcing* ocorra em um ambiente controlado, utilizando plataformas comerciais, como o *AMT* e o *Microworkers*, apenas para recrutar *workers* e realizar pagamentos.

O *CrowdNoter* também permite que sejam utilizados provedores de mídia externos, como o *Youtube*. A estratégia utilizada para isso é armazenar e transmitir apenas metadados entre os componentes do sistema, assim como referências para recuperar os objetos de mídia hospedados nos provedores externos.

Conforme pode-se observar na Figura 4, o *CrowdNoter* foi construído sobre uma arquitetura baseada em componentes com três módulos principais *Servidor*, *Cliente* e *Persistência*.

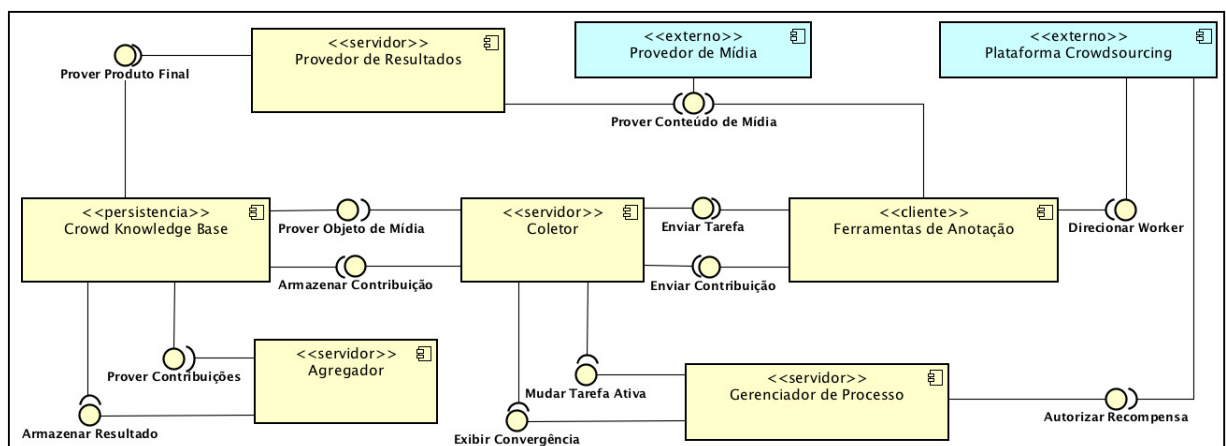


Figura 4 – Componentes do sistema *CrowdNoter* e suas interfaces de comunicação.

Todo o processamento das contribuições, filtrações, agregações e geração de resultados são realizados no módulo *Servidor*, oferecendo assim um diferencial importante uma vez que tanto as contribuições quanto os resultados produzidos ficam restritos ao contexto do *CrowdNoter*, não passando pela plataforma de *crowdsourcing*, nem mesmo durante as atividades de coleta das contribuições. Isto se deve ao uso do *Cliente*, que permite instalar em ambientes externos ao *CrowdNoter* interfaces leves para as ferramentas de anotação usadas para realizar as *microtasks*. Estas ferramentas são exibidas aos *workers* a partir de um navegador de Internet, e as contribuições são enviadas diretamente para o módulo *Cliente*.

O *CrowdNoter* também é responsável por distribuir as tarefas entre os *workers*, determinando os itens a serem anotados em cada tarefa, além de oferecer os métodos de agregação para gerar o resultado de cada *microtask* e os painéis de monitoramento que permitem ao *GPC* observar em tempo real o resultado parcial da tarefa atual.

4.5.1 Módulo de Persistência

O módulo de *Persistência* é responsável por armazenar e gerenciar os metadados referentes tanto às contribuições coletadas quanto aos resultados gerados. Este módulo possui um único componente, o *Crowd Knowledge Base*, que oferece uma camada de abstração, disponibilizando para o *CrowdNoter* todos os dados presentes no contexto a qualquer momento.

O módulo *Persistência* se comunica com o *Servidor* para três atividades diferentes. Ele se comunica com o componente *Coletor*, tanto para fornecer as informações necessárias para exibir as tarefas aos *workers*, quanto para armazenar as contribuições. Ele também se comunica com o componente *Agregador* fornecendo as contribuições coletadas, e armazenando os resultados gerados pelos métodos de agregação. No estágio de distribuição, este módulo se comunica com o componente *Provedor de Resultados* fornecendo o produto final e os metadados referentes aos resultados parciais, para que possam ser distribuídos.

De modo geral, todas as informações que trafegam no *workflow*, durante qualquer parte do processo, são armazenadas e disponibilizadas pelo componente *Crowd Knowledge Base*. Uma vez que estes dados estão sempre disponíveis, e este componente recebe as contribuições diretamente das ferramentas de anotação, não existe nenhuma necessidade de que os dados coletados sejam persistidos nas plataformas de *crowdsourcing*. Esta estratégia pode ser especialmente interessante ao se trabalhar com informações sensíveis sobre as quais se deseja garantir segurança e sigilo, uma vez que é uma prática muito

comum acumular as contribuições na plataforma comercial de *crowdsourcing* e recuperar estes dados apenas ao final do processo de coleta.

4.5.2 Módulo Servidor

O módulo do Servidor é responsável por distribuir as tarefas, gerenciar o *workflow* do processo de cascadeamento de *microtasks*, e controlar a aquisição, processamento e acesso aos dados. Este módulo é composto por quatro componentes: Gerenciador de Processo, Coletor, Agregador, e Provedor de Resultados. Este módulo também controla a comunicação entre o módulo *Cliente* e o módulo de *Persistência*, criando assim um nível adicional de segurança para as informações disponíveis no *CrowdNoter*. Os componentes que formam o módulo Servidor são apresentados a seguir.

O **Gerenciador de Processo** é o componente responsável pelo controle do processo *crowdsourcing*. Este módulo oferece as funcionalidades para que o *GPC* monitore a coleta das contribuições e o resultado gerado até o momento para a tarefa atual. Além disso, ele permite realizar a transição entre as *microtasks* tornando possível controlar a execução do *workflow*. O *Gerenciador de Processo* é responsável por informar ao *Coletor* que a *microtask* que está em execução.

O **Coletor** é o módulo responsável por capturar as contribuições da *crowd*. Este módulo envia para o módulo *Cliente* as informações necessárias para que este exiba para os *workers* a tarefa a ser executada, e recebe as contribuições diretamente das ferramentas de anotação. Além disso, esse componente é responsável por reunir as anotações fornecidas pelo *workers* e as envia para o módulo de *Persistência*.

O **Agregador** é o módulo responsável tanto por aplicar os filtros de confiabilidade sobre as contribuições quanto por processar as anotações válidas, gerando o resultado agregado de cada tarefa. O módulo *Agregador* utiliza, para cada *microtask*, o método de agregação que foi associado a ela no estágio de planejamento.

O **Provedor de Resultados** é o módulo responsável por tornar os resultados acessíveis, para que possam ser distribuídos no estágio de distribuição. Este módulo oferece serviços que entregam o produto final para outros sistemas. Dessa forma, o *Provedor de Conteúdo* pode tanto enviar os objetos de mídia suplementados com anotações, como pode disponibilizar apenas os metadados no formato definido no estágio de planejamento.

4.5.3 Módulo Cliente

O módulo Cliente gerencia as ferramentas de anotação, gerando a visualização da tarefa que será apresentada ao *worker* e provendo os recursos necessários para que ele possa executá-la por meio de um navegador de internet. Para cada *microtask*, o cliente utiliza a ferramenta de anotação associada a ela no estágio de planejamento e, a partir dos dados que recebe do *Coletor*, monta e exibe ao *worker* a tela da tarefa para que ele possa executá-la. Uma vez que o *worker* realize a tarefa, o *Coletor* pode então capturar a contribuição diretamente da ferramenta de anotação.

4.5.4 Interfaces de Comunicação

A comunicação entre os módulos do sistema *CrowdNoter* ocorre através das interfaces de comunicação representadas na Figura 4 e detalhadas a seguir na tabela 15.

4.5.5 Ferramentas de Anotação

O *CrowdNoter* fornece um conjunto inicial de ferramentas de anotação simples, que foram projetadas para que possam ser facilmente modificadas para atender novas demandas. Estas ferramentas foram desenvolvidas ao longo dos estudos de caso realizados, de forma que cobrem todos os tipos de anotação que foram necessários para a execução das *microtasks* realizadas em cada um deles.

Estas ferramentas foram desenvolvidas sob demanda para atender aos estudos de caso realizados, e atualmente já cobrem os principais tipos de interação apresentados no *survey* (LI et al., 2017).

4.5.6 Métodos de Agregação

Os métodos de agregação disponíveis no *CrowdNoter* são variações da estratégia *Majority Decision*, e estes métodos admitem a aplicação de técnicas de confiabilidade baseadas em valores de referência. A utilização de valores de referência incrementa o método de agregação, tornando-o um método *HoneyPot*. Conforme pode ser observado no *survey* (LI et al., 2017), os métodos de agregação baseados em *Majority Decision* predominam entre os trabalhos que utilizam processos *crowdsourcing*.

Tabela 15 – Interfaces do sistema *CrowdNoter*.

Interface	Descrição
Mudar Tarefa Ativa	O owner envia uma solicitação ao Servidor para definir a tarefa ativa no momento.
Exibir Convergência	O gerenciador exibe para o owner o estado de convergência atual da tarefa ativa.
Prover Objeto de Mídia	O Coletor recebe uma entrada do componente Persistência para gerar a visualização de cada tarefa a ser executada.
Prover Conteúdo de Mídia	O provedor de conteúdo, que pode ser o Youtube, Flickr ou outro serviço externo, disponibiliza o conteúdo requerido.
Enviar Tarefa	O Coletor envia um trabalho para um worker que visualiza a tarefa através do Cliente e a executa.
Direcionar Worker	A plataforma de crowdsourcing direciona um worker, por meio de um browser, para o endereço onde se localiza a ferramenta para execução da tarefa que deve ser realizada.
Enviar Contribuição	O Cliente envia ao coletor a anotação feita pelo worker.
Armazenar Contribuição	O Coletor envia contribuições dos workers para a persistência, que armazena no banco de dados.
Prover Contribuições	A Persistência envia para o Agregador todas as contribuições coletadas relacionadas a tarefa atual.
Armazenar Resultado	O Agregador armazena as entradas recebidas do processo de agregação. O resultado gerado pode ser fornecido como entrada para a próxima tarefa, ou ser usado para gerar o produto final caso seja a última tarefa do workflow.
Prover Produto Final	A Persistência fornece ao Provedor de Resultados o produto final do processo de cascadeamento de microtasks. Este produto é então distribuído, ficando acessível para outros sistemas ou módulos de visualização de mídia.

De modo geral, todos os métodos seguem uma estratégia de três passos, sendo que nos casos em que exista um conjunto de valores de referência é adicionado um passo inicial para descartar as contribuições dos *workers* maliciosos. Dessa forma, a utilização de valores de referência não altera o funcionamento do método, uma vez que é feita antes da agregação.

Na tabela 17 é listada a coleção de métodos de agregação disponíveis atualmente no *CrowdNoter*. Esses métodos foram desenvolvidos sob demanda para atender aos protótipos utilizados para realizar os estudos de caso.

Tabela 16 – Tipos de anotação atualmente disponíveis no *CrowdNoter*.

ID	Interação	Descrição	Exemplos de Aplicação
AT-1	Anotação Aberta	texto aberto digitado pelo worker	pontos de interesse, anotação de objetos de cena
AT-2	Agrupamento	categoria selecionada em uma lista fechada	classificação, sumarização
AT-3	Seleção	seleção de um item de mídia em uma lista de candidatos	votação, seleção de conteúdo
AT-4	Anotação Fechada	worker fornece marcação ou resposta pré-determinada	adição de marcadores, sincronização de vídeos, segmentação
AT-5	Envio de Conteúdo	o worker envia conteúdo externo pela ferramenta	coleta de dataset, aquisição de imagens
AT-6	Avaliação	conjunto formado por instantes inicial e final	delimitação de cenas, segmentação
AT-7	Ordenação	o worker recebe uma lista de itens e determina sua ordem	ranking, ordenação

4.5.7 Discussão sobre o *Framework*

Neste capítulo foi apresentado o *CrowdWaterfall*, um *framework* destinado a projetos *crowdsourcing* que demandam tarefas complexas de anotação de vídeos.

Este *framework* inclui o processo *crowdsourcing*, que contempla a decomposição e a remodelagem da tarefa complexa original como um *workflow* de tarefas simples, sendo este *workflow* uma parte integrante do *CrowdWaterfall* que visa formalizar a execução e gerência do processo.

O *framework* apresentado também inclui a formalização e descrição dos papéis envolvidos no processo utilizado, assim como as atividades e responsabilidades de cada um deles. Adicionalmente, foi apresentado o *CrowdNoter*, um ambiente computacional que foi construído com o objetivo de prover suporte computacional para o *CrowdWaterFall*.

No próximo capítulo serão apresentados estudos de caso nos quais o *CrowdWaterfall* foi aplicado em cenários reais. Esses estudos de caso foram desenvolvidos utilizando os recursos apresentados neste capítulo, incluindo o ambiente *CrowdNoter*.

Tabela 17 – Métodos de Agregação atualmente disponíveis no *CrowdNoter*.

ID	Critérios de Filtragem	Agrupamento	Agregação	Exemplo de Aplicação
AG-1	1. Existem palavras proibidas na contribuição? 2. O texto digitado é cópia de um trecho da legenda?	1. por tempo 2. por conteúdo	percentual de convergência em cada grupo	identificação de pontos de interesse
AG-2	1. A contribuição está em um formato adequado? 2. A fonte da contribuição é confiável e permitida?	1. por conteúdo	formar um <i>dataset</i> com os conteúdos sugeridos	aquisição de conteúdo extra
AG-3	1. O worker acessou todos os candidatos?	1. por item 2. por região	ranking com as contribuições mais votadas	seleção da melhor contribuição
AG-4	1. O worker movimentou o objeto pela cena? 2. A posição está em uma região válida?	1. por item	média geométrica dos intervalos na região predominante	definição da melhor posição
AG-5	1. O worker acessou todo o conteúdo enviado a ele?	1. por tempo	percentual de convergência em cada grupo	definição do instante de um evento
AG-6	1. O worker acessou todo o conteúdo enviado a ele? 2. O intervalo é válido?	1. pelo início do intervalo	intercessão dos intervalos recorrentes	definição de um intervalo temporal

5 Estudos de Caso

Neste capítulo são apresentados os três principais estudos de casos realizados com base no *framework* proposto nesta tese. Apesar de se tratarem de cenários diferentes, todos utilizaram o *framework* da mesma forma, assim como foram utilizados os mesmos recursos computacionais para executar cada um deles. Para cada um dos estudos de caso, foi selecionado um problema de requer tarefas complexas de processamento de vídeo, assim como requerem inteligência humana para suas execuções. Foram escolhidos exclusivamente problemas que demandam tarefas subjetivas e que os resultados finais devem refletir a inteligência coletiva capturada dos membros da *crowd*. Dessa forma, as tarefas necessárias para resolver os problemas abordados nos estudos de caso, apesar de triviais para os seres humanos, podem ser consideradas como difíceis para os métodos automáticos. Inclusive, um dos critérios para a escolha dos estudos de caso foi que não estivessem disponíveis, até o momento, soluções automáticas que pudessem resolvê-las. Assim, o *framework* utilizado não se apresenta como concorrente aos métodos automáticos, mas como opção de solução de problemas para os quais estes métodos não estão disponíveis ou não são adequados. Os estudos de caso utilizaram números diferentes de tarefas *crowdsourcing*, assim como produziu resultados em formatos diferentes, conforme pode ser observado na tabela 16.

Tabela 18 – Estudos de Caso.

Problema	Tarefas	Produto
1. Enriquecimento de Vídeos	4 - workflow da Figura 9	Vídeo enriquecido com recursos multimídia.
2. Anotação de Efeitos Multissensoriais em Vídeos	2 - workflow da Figura 13	Arquivo MPEG-V com metadados sobre efeitos sensoriais de vento e vibração.
3. Segmentação de Sinais LIBRAS em Vídeos	2 - workflow da Figura 10	Apresentação multimídia com navegação sobre os gestos segmentados.

Para avaliar de forma objetiva a diferença entre o resultado obtido em cada um dos estudos de caso foi utilizada a Raiz do Erro Quadrático Médio (do inglês *Root Mean Square Error* ou *RMSE*), que é adequado para comparar os valores obtidos com os valores de referência, e fornece um valor de erro médio na mesma escala dos valores analisados. O *RMSE* é usualmente aplicado no *crowdsourcing*, como pode ser observado nos trabalhos (LOWRY; FIENEN, 2013; ZHU et al., 2014; OUYANG et al., 2015)

5.1 Aspectos Operacionais dos Estudos de Caso

Uma parte importante a ser considerada na realização de experimentos envolvendo *crowdsourcing* é a definição dos aspectos operacionais. Esses aspectos dizem respeito à escolha da *crowd* e da plataforma comercial de *crowdsourcing*, assim como à alocação dos recursos necessários para arcar com os custos dos experimentos. A plataforma comercial de *crowdsourcing* escolhida para recrutar e recompensar os *workers* foi a *Microworkers*¹, pois ela se mostrou uma plataforma menos burocrática em relação às exigências para o cadastro de *owners* residentes no Brasil, e os preços cobrados eram compatíveis com o orçamento disponível para os experimentos.

Uma vez que o *CrowdWaterfall* recomenda que sejam utilizadas apenas tarefas simples, durante as atividades *crowdsourcing*, não houve necessidade de exigir habilidades ou conhecimentos específicos dos *workers*. Dessa forma foi possível recrutar uma *crowd* heterogênea para cada um dos estudos de caso. A heterogeneidade da *crowd* foi alcançada ao recrutar *workers* de diferentes países, sem outros filtros relacionados com grau de conhecimento, capacitação ou outros fatores que poderiam levar a formação de uma *crowd* com perfil direcionado.

O *CrowdWaterfall* também recomenda que as instruções apresentadas aos *workers* sejam muito simples e claras, de forma que também não foi exigido que os *workers* dominassem nenhum idioma específico. Uma vez que a plataforma *Microworkers* é completamente no idioma inglês, foi assumido que os *workers* cadastrados na plataforma seriam capazes de entender as instruções para as tarefas, que também foram escritas nesse idioma. Durante os estudos de caso foram registradas contribuições de *workers* residentes em treze diferentes países.

As tarefas eram apresentadas aos *workers* em um mural de anúncios existente na plataforma e, ao aceitar uma delas, o *worker* era direcionado para a ferramenta de anotação. O *CrowdNoter* enviava os itens a serem anotados para as ferramentas de anotação, que por sua vez enviava para o *CrowdNoter* as contribuições coletadas da *crowd*. Dessa forma foi possível validar a estratégia de comunicação e integração entre o *CrowdNoter* e as plataformas comerciais de *crowdsourcing*. A plataforma *Microworkers* foi responsável por registrar os *workers* que realizaram as contribuições e, uma vez que as contribuições eram verificadas, cada *worker* recebia o seu pagamento. O dinheiro destinado aos custos da plataforma e dos *workers* foram adicionados na *Microworkers* antes do início de cada estudo de caso.

¹ <<https://www.microworkers.com>>

5.2 Estudo de Caso 1: Enriquecimento de Vídeo

O primeiro estudo de caso está associado a um processo de enriquecimento de vídeo por meio da inserção de conteúdo suplementar. Para este experimento foi produzido um vídeo baseado em um roteiro no qual foram inseridos um conjunto de possíveis pontos de interesse, e a lista desses pontos foi utilizada para ajudar a avaliar o resultado obtido.

Este estudo de caso foi concebido para endereçar todos os aspectos do *framework* proposto, assim como foram utilizados todos os recursos do sistema desenvolvido, e foi utilizado um *workflow* formado por *microtasks* que utilizam diferentes ferramentas de anotação e métodos de agregação, a partir do qual o produto final é construído gradativamente ao longo do processo.

O produto gerado pode ser utilizado em diversas aplicações aumentando a carga semântica contida nos vídeos originais, tanto para fins de entretenimento quanto para aplicações de educação e treinamento. Esta estratégia se mostra muito interessante para enriquecer vídeos pois dispensa a necessidade de especialistas que conheçam o conteúdo transmitido ou ferramentas de edição de vídeo. Outro ponto importante nesta estratégia é que o conteúdo suplementar inserido no vídeo reflete a inteligência coletiva dos *workers*, minimizando informações direcionadas ou tendenciosas.

Este estudo de caso foi apresentado nos artigos (AMORIM; NETO; SANTOS, 2018) e (AMORIM et al., 2017), que pode ser consultado para obtenção de detalhes sobre a aplicação do *CrowdWaterfall* e execução do *workflow*.

5.2.1 Preparação do Estudo de Caso 1

A *crowd* utilizada neste estudo de caso foi recrutada e recompensada por intermédio da plataforma *Microworkers*. Ao criar as campanhas de coleta de contribuições, um funcionário da plataforma sugeriu que fossem recrutados *workers* de um grupo específico, chamado *Data Services*. Este grupo possui tamanho médio para os padrões da plataforma, com cerca de 1300 *workers*, que apesar de não possuírem habilidades específicas costumam executar as tarefas imediatamente após aceitá-las.

O orçamento alocado para este estudo de caso foi de 21 dólares, o que permitiu cobrir os custos da plataforma, e coletar 351 contribuições ao custo de 0.05 dólar por tarefa realizada.

5.2.2 Objetivo do Estudo de Caso 1

O objetivo do processo foi transformar um vídeo simples em uma apresentação multimídia interativa na qual, conforme ocorrem os pontos de interesse, são exibidos itens sobre o vídeo nos quais o usuário pode clicar para visualizar informação adicional. Além disso, a visualização do produto final inclui o índice com todos os pontos de interesse de forma que o usuário possa navegar entre eles sobre vídeo enriquecido.

O resultado desejado é uma apresentação multimídia na qual o vídeo original é exibido de forma sincronizada com o conteúdo suplementar. Este conteúdo suplementar, para este experimento, podem ser páginas da *Wikipedia*, vídeos do *Youtube*, ou um texto explicativo. Além disso a apresentação gerada deve conter um índice, com todos os pontos de interesse suplementados, a partir do qual seja possível aos usuários navegar pelo vídeo.

5.2.3 Workflow do Estudo de Caso 1

O estágio de planejamento inclui determinar as características do produto final, as *microtasks* e gerar o *workflow* (figura 9). Para alcançar o objetivo desejado foram adicionadas quatro *task-sets* ao *workflow*:

- **Task-Set 0 - Identificar Pontos de Interesse:** o *worker* recebia um segmento de vídeo, que variava entre 4.5s e 7.2s, e precisava anotar caso identificasse algum ponto de interesse. Os pontos de interesse poderia ser *Pessoas*, *Eventos* ou *Tecnologias*. A ferramenta desenvolvida para esta tarefa pode ser observada na figura 5.

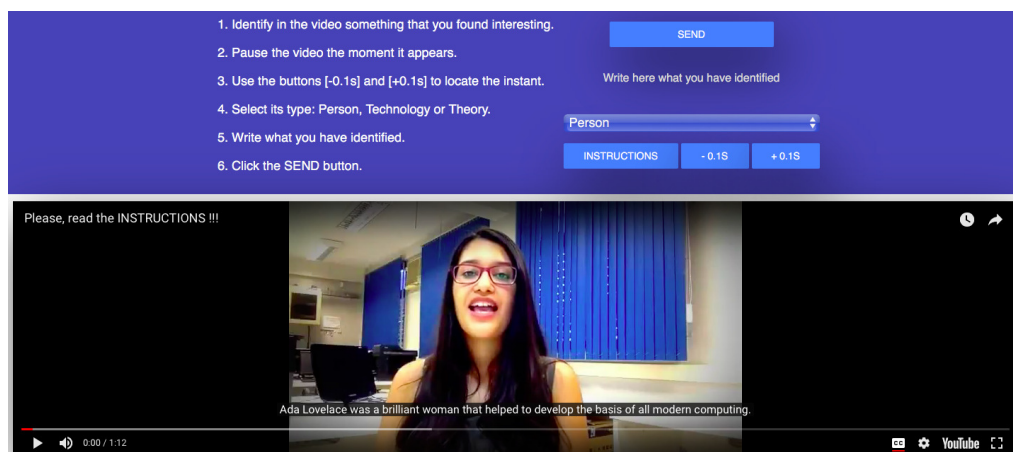


Figura 5 – Ferramenta para a execução da primeira tarefa do estudo de caso 1.

- **Task-Set 1 - Sugerir Conteúdo Extra:** o *worker* recebia um segmento de vídeo contendo um dos pontos de interesse sobre os quais houve convergência da *crowd*, e deveria sugerir conteúdo extra para suplementá-lo. Este conteúdo extra deveria ser uma página da *Wikipedia*, um vídeo do *Youtube*, uma Imagem, ou um Texto. A ferramenta desenvolvida para esta tarefa pode ser observada na figura 6.

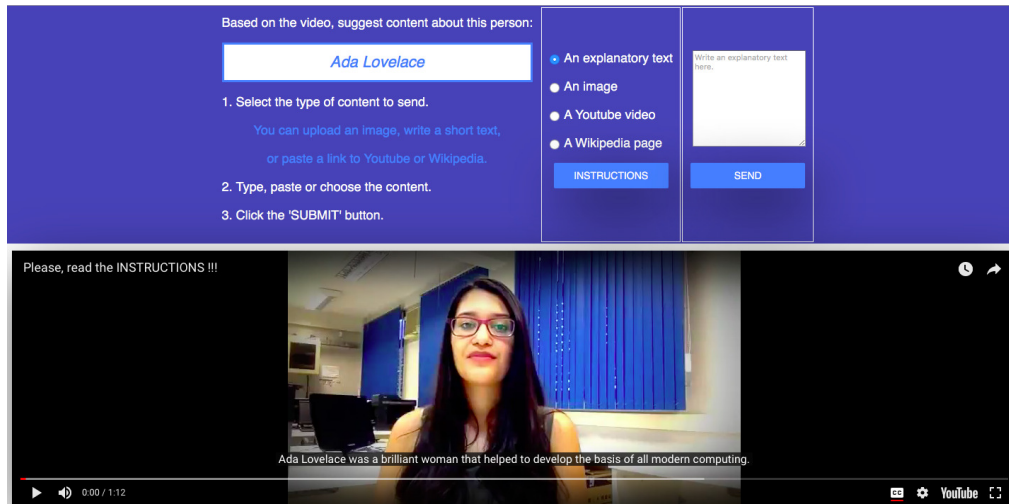


Figura 6 – Ferramenta para a execução da segunda tarefa do estudo de caso 1.

- **Task-Set 2 - Ranquear Conteúdo Extra:** o *worker* recebia um segmento de vídeo contendo um dos pontos de interesse, a lista de conteúdos extras sugeridos para suplementá-lo. O *worker* tinha como tarefa acessar as sugestões e votar no conteúdo que achou mais adequado para suplementar o ponto de interesse. A ferramenta desenvolvida para esta tarefa pode ser observada na figura 7.

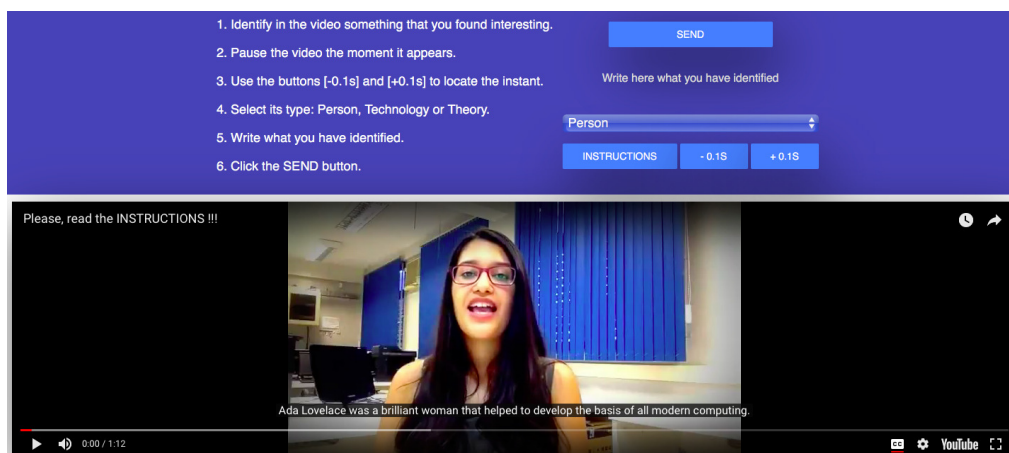


Figura 7 – Ferramenta para a execução da terceira tarefa do estudo de caso 1.

- **Task-Set 3 - Posicionar Conteúdo Extra:** o *worker* recebia uma cena do vídeo sobre o qual estava um ícone ou miniatura do conteúdo extra escolhido para suplementá-lo. A tarefa do *worker* consistia em posicionar o ícone sobre a cena de forma que causasse menos oclusão possível, e que deixasse a cena visualmente mais agradável. A ferramenta desenvolvida para esta tarefa pode ser observada na figura 8.

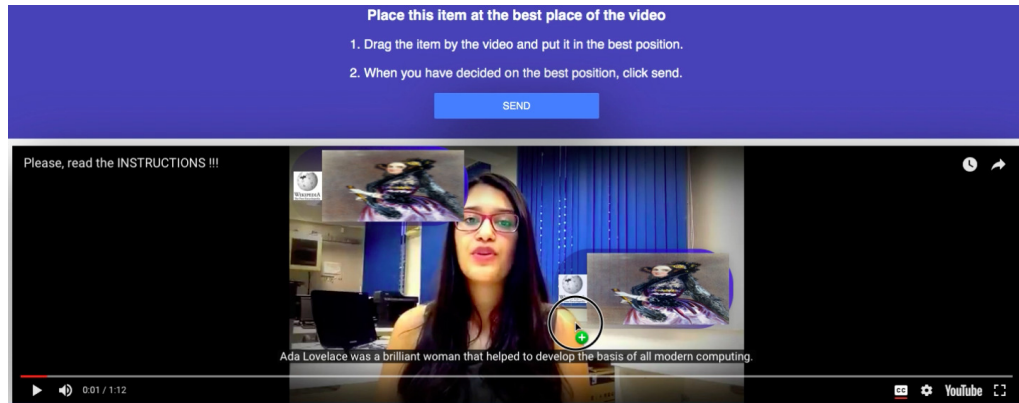


Figura 8 – Ferramenta para a execução da ultima tarefa do estudo de caso 1.

5.2.4 Produto Estudo de Caso 1

O vídeo utilizado no experimento foi produzido internamente, com base em um roteiro no qual 15 possíveis pontos de interesse foram adicionados de forma planejada. O instante inicial de cada um dos 15 pontos de interesse que foram adicionados intencionalmente, foram utilizados como valores de referência para avaliar o resultado obtido pelo processo *crowdsourcing* utilizado.

Como mostrado na tabela 19, houve convergência da *crowd* sobre 11 desses 15 pontos de interesse, com um grau de concordância de pelo menos 60%, o que equivale a 73.3% dos potenciais pontos de interesse. Além disso, os instantes no vídeo que a *crowd* determinou para os pontos de interesse foram precisos quando comparados com os instantes anotados durante a produção do vídeo.

Tabela 19 – Anotações no Roteiro X Opinião da *Crowd*

Pontos de Interesse	Instante no Roteiro (s)	Instante no Resultado (s)
Ada Lovelace	0.83	0.64
Mathematics	6.98	7.01
Babbage	13.53	13.66
Augustus DeMorgan	24.79	24.65
Logical Relations and Properties	30.32	30.01
Formalisms	33.99	
Babbage Analytics Engine	40.95	40.57
Logical Structures	47.92	48.23
George Boole	53.22	53.27
Binary Algebra System	54.53	
Turing	57.97	58.28
Universal Computer	59.31	
Claude Shannon	61.42	61.49
Boolean Tasks	65.01	
The First Computers	71.09	70.57

O valor do *RMSE* calculado entre o resultado e os valores de referência foi de 0.47s, o erro mínimo foi de 0.03s e o máximo foi de 0.52s. A distância média entre os valores de referência para os instantes iniciais é de 7.03s, de forma que o *RMSE* revela uma margem de erro de 6.69%. O resultado do processo *crowdsourcing* foi utilizado para gerar a apresentação multimídia interativa pode ser acessada no endereço <<http://crowdnoter.com/enrichment01>>.

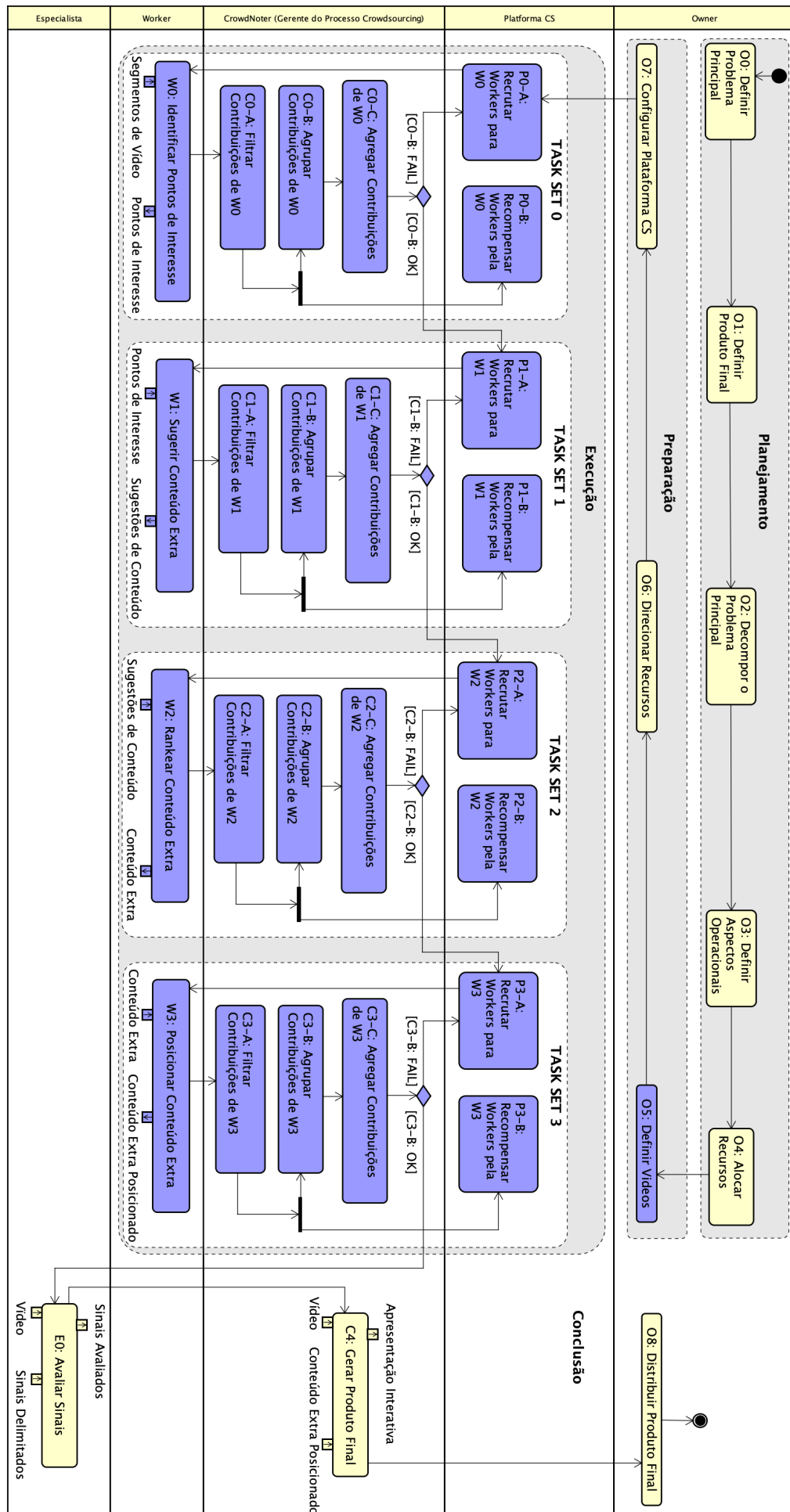


Figura 9 – Workflow para o processo crowdsourcing de enriquecimento de vídeo.

5.3 Estudo de Caso 2: Segmentação de Sinais LIBRAS de Vídeos

O segundo estudo de caso está relacionado com as tecnologias assistivas. Neste experimento, o *framework* proposto nesta tese foi utilizado para segmentar temporalmente sinais apresentados em quatro vídeos nos quais um intérprete usa sinais em LIBRAS para a comunicação. A ideia principal foi delimitar os instantes iniciais e finais em que cada sinal eram apresentados nos vídeos, em um cenário em que os sinais não estavam associados simplesmente com letras ou conceitos pré-determinados, mas poderiam ser, inclusive descritores, regionalismos, e sinais relacionados com termos técnicos, uma vez que no *framework* utilizado não importavam quais eram os sinais a serem segmentados.

Este experimento contou com o apoio de um grupo de especialistas que são intérpretes e professores do curso de Letras/LIBRAS da UFES, que forneceram anotações sobre a segmentação dos sinais, que foram utilizadas para avaliar o resultado do processo *crowdsourcing*, além de fornecerem anotações sobre o significado de cada um dos sinais segmentados.

O time de especialistas que colaboraram neste projeto foi composto de cinco intérpretes, sendo quatro deles do corpo técnico do departamento de LIBRAS e um professor do curso de graduação Letras/LIBRAS. Estes colaboradores atuaram em dois momentos diferentes. Na fase inicial do processo eles geraram os valores de referência que foram utilizados para determinar o erro médio dos instantes iniciais resultantes da agregação das contribuições da *crowd*. Os especialistas participaram novamente ao final do processo, avaliando a qualidade da delimitação dos segmentos obtidos.

5.3.1 Objetivo do Estudo de Caso 2

O produto final esperado neste estudo de caso consiste em uma apresentação multimídia na qual é possível navegar entre os sinais de LIBRAS exibidos em um vídeo, em conjunto com o significado dos sinais sendo exibido de forma textual, em sincronia com o vídeo. Esta apresentação multimídia, que pode inclusive ser utilizada como material didático para o ensino de LIBRAS, é gerada a partir da segmentação obtida do processo *crowdsourcing*, juntamente com os significados fornecidos pelos especialistas para cada sinal.

Todavia é possível obter outros resultados a partir do mesmo processo, como uma base anotada de sinais de LIBRAS que pode ser utilizada para treinar modelos de inteligência artificial destinados a realizar a segmentação automática dos sinais.

A vantagem de segmentar os sinais antes de traduzi-los consiste em poder solicitar que um especialista traduza apenas alguns sinais, em vez de solicitar que o mesmo assista um vídeo inteiro e o interprete. Os detalhes sobre isso podem ser vistos no artigo, que ainda não foi publicado, mas pode ser visualizado em <https://github.com/cellonovaes/Crowdsourcing-Segmentation-of-LIBRAS-Signs>.

5.3.2 Preparação do Estudo de Caso 2

O orçamento disponível para recompensar os workers no estudo de caso 2 foi de 20 dólares americanos, sendo que 50% foi direcionado para cada uma das *microtasks* definidas. Os custos adicionais com a plataforma, além do pagamento dos *workers*, foi de 2.10 dólares.

O valor pago a cada *worker* pela execução de uma tarefa foi calculado com base na quantidade de itens a serem anotados em cada tarefa. Na primeira tarefa haviam quatro vídeos a serem anotados, de forma que foram direcionados 2.5 dólares para anotar cada um dos vídeos. O valor praticado na plataforma *Microworkers* para tarefas similares a esta varia entre 0.02 e 0.08 dólar, o que levou a um valor de 0.05 dólar por tarefa realizada. Esta estratégia possibilitou coletar 50 contribuições para cada vídeo, ou 200 contribuições para a primeira tarefa.

Também foi decidido que o orçamento da segunda tarefa também seria distribuído uniformemente entre os itens a serem anotados, e o valor seria calculado de forma a conseguir, tentando maximizar a quantidade de contribuições por item de acordo com os valores praticados na plataforma.

Não foram definidos filtros para selecionar o perfil dos *workers*, o que possibilitou coletar contribuições vindas de diferentes países, a saber, Filipinas, Venezuela, Índia, Alemanha, Rússia, Macedônia, Espanha, Bulgária, Coreia, México, Mauritânia, Malásia, Emirados Árabes Unidos, Tailândia, Noruega, Sérvia, Nova Zelândia, Estados Unidos, Bangladesh, Sri Lanka, Vietnã, Argentina, Nigéria e Indonésia.

5.3.3 Workflow do Estudo de Caso 2

A fase de planejamento do experimento ocorreu exatamente como nos outros estudos de caso, o que reforça a afirmativa de que o *framework* proposto pode ser aplicado em diferentes cenário de anotação de vídeo. A fase de execução conta com duas tarefas *crowdsourcing* em seu *workflow* (Figura 10):

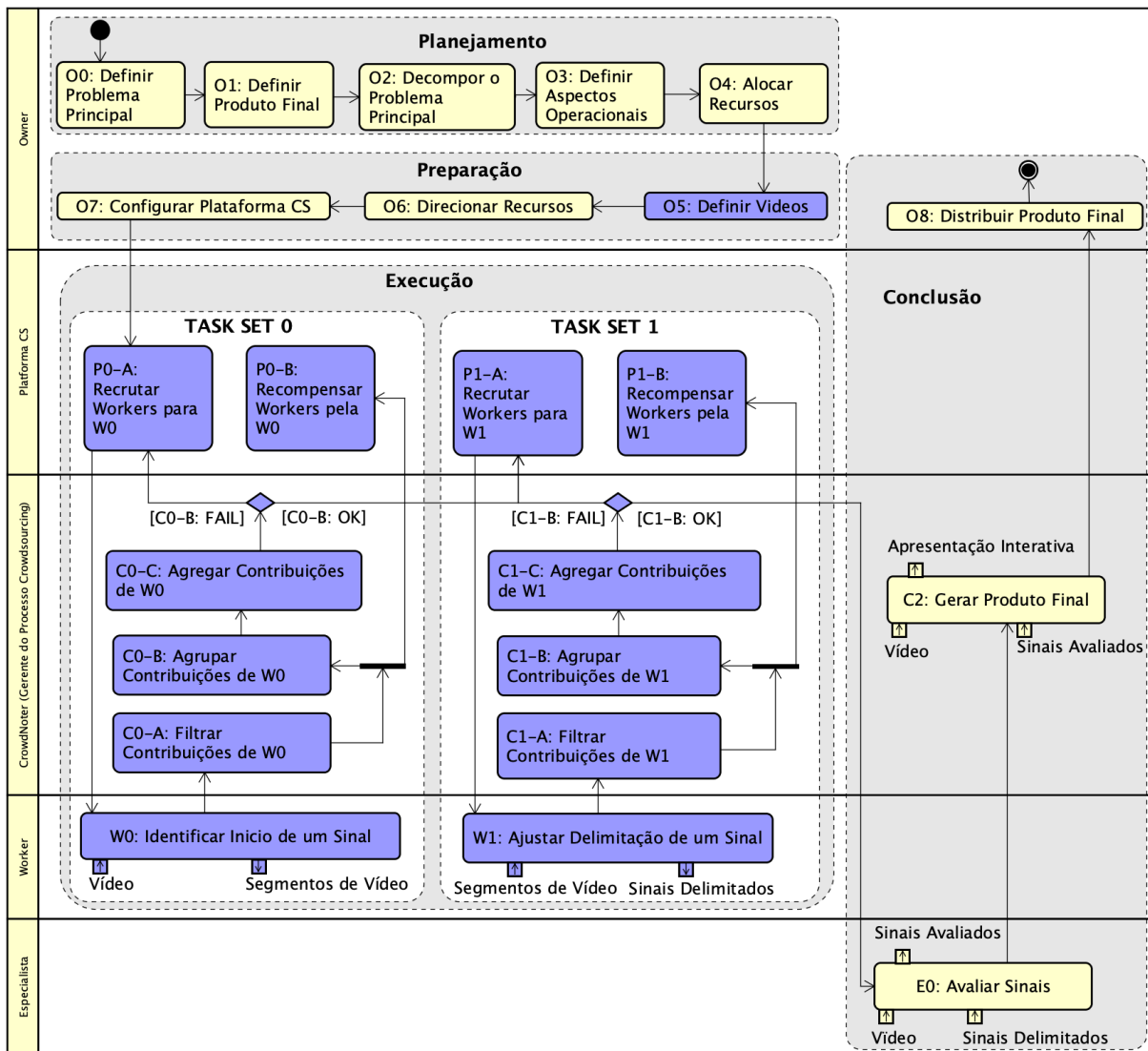


Figura 10 – Workflow do processo crowdsourcing para segmentação de sinais de LIBRAS.

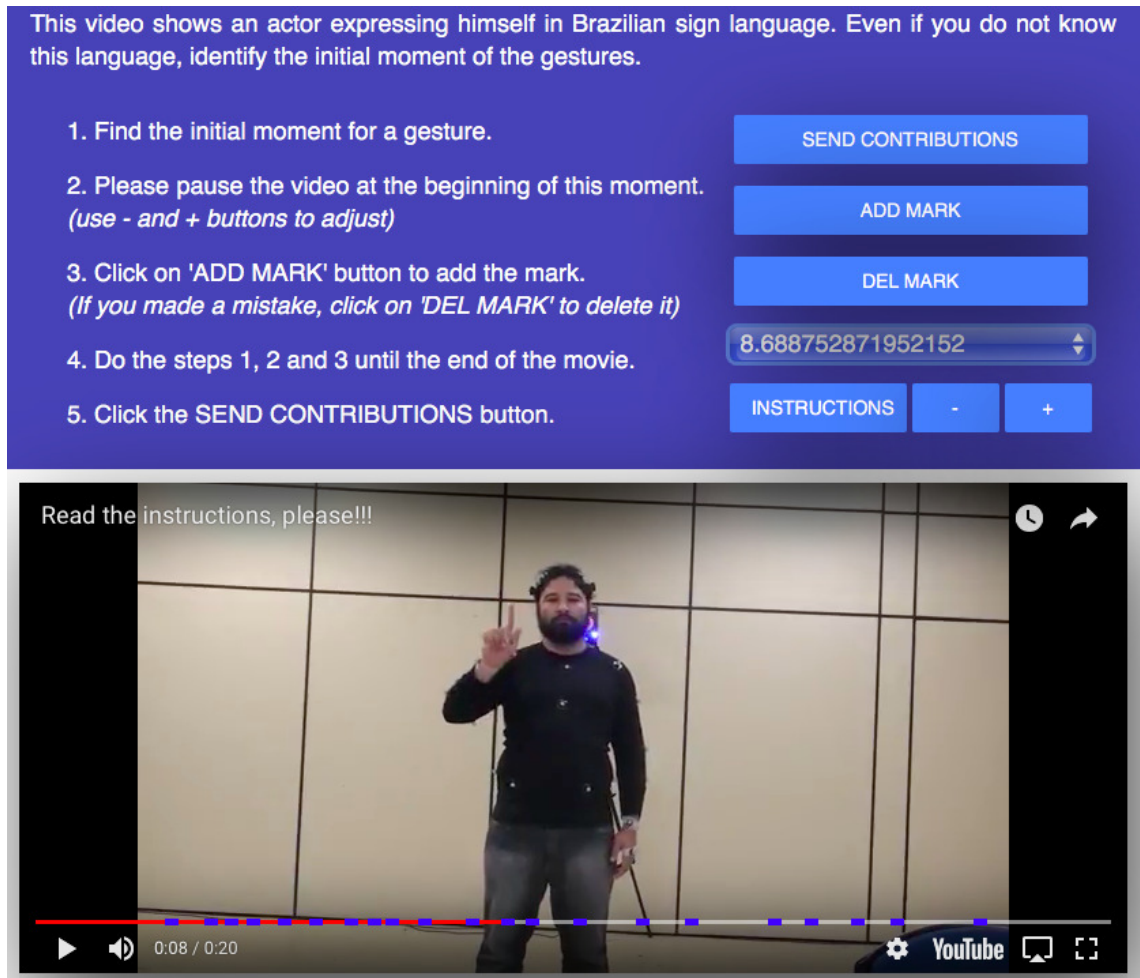


Figura 11 – Ferramenta para a execução da primeira tarefa do estudo de caso 2.

- **Task-Set 0 - Identificar Início de um Sinal:** o *worker* recebia um dos vídeos e precisava simplesmente identificar o instante inicial dos sinais, pausando o vídeo no início de cada sinal e adicionando um marcador ao clicar em um botão. Após determinar a convergência da *crowd* sobre os instantes iniciais, foram determinados segmentações iniciais para os sinais tomando o intervalo entre dois inícios consecutivos. A ferramenta desenvolvida para esta tarefa pode ser observada na figura 11.
- **Task-Set 1 - Ajustar Delimitação de um Sinal:** o *worker* recebia um segmento de vídeo contendo um sinal de LIBRAS, e deveria ajustar o início e o final do segmento de modo a definir uma delimitação mais precisa de cada sinal. A ferramenta desenvolvida para esta tarefa pode ser observada na figura 12.

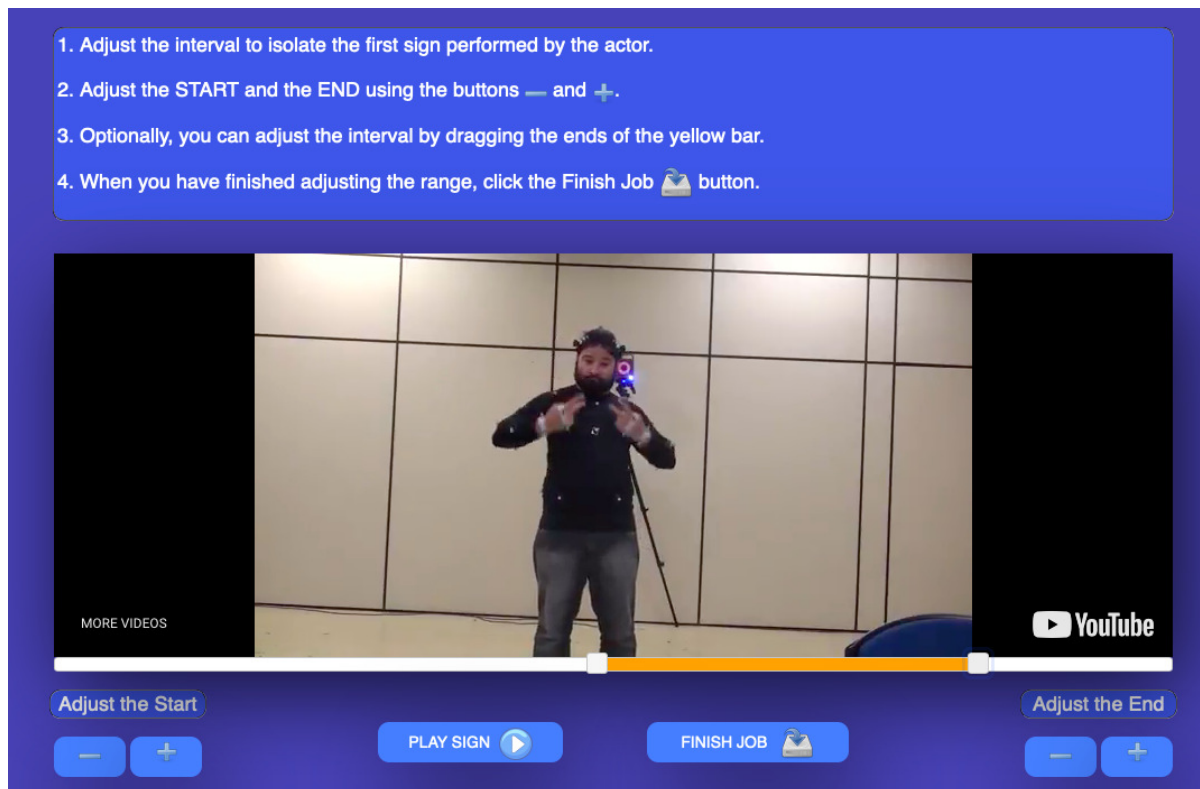


Figura 12 – Ferramenta para a execução da segunda tarefa do estudo de caso 2.

5.3.4 Produto do Estudo de Caso 2

Na tabela 20 é possível observar os segmentos resultantes do processo *crowdsourcing* em comparação aos segmentos fornecidos pelos especialistas que colaboraram com o experimento.

Em comparação com as anotações dos especialistas, o *framework* crowdsourcing proposto foi capaz de identificar 90.47% dos sinais de LIBRAS nos quatro vídeos do nosso estudo de caso, além de ser capaz de delimitar 93.75% dos sinais identificados com sucesso nos mesmos vídeos.

Uma observação importante nesses resultados é que não houve nenhum caso de falso-negativo, assim como não houve fusão de sinais no resultado. Isso significa que, embora o resultado obtido não corresponda a 100% dos sinais apresentados nos vídeos, todos os sinais delimitados podem ser utilizados para fins de exibição ou ensino, mesmo que 6,25% deles possam apresentar características que dificultem a sua compreensão.

Tabela 20 – instantes iniciais (em s) dos sinais nos vídeos V_1 a V_4

Vídeo V_1		Vídeo V_2		Vídeo V_3		Vídeo V_4	
Esp.	Crowd	Esp.	Crowd	Esp.	Crowd	Esp.	Crowd
1.94	1.97	2.01	2.03	1.73	2.04	2.35	2.29
3.25	3.44	2.71	2.72	2.93	3.38	2.76	2.96
4.36	4.68	4.01	3.91	4.16	4.86	3.47	3.48
5.65	5.75	5.74	5.76	5.33	5.52	4.07	3.99
6.34	6.58	6.82	6.97	6.14	6.03	4.67	4.67
7.47	7.58	7.57	7.61	6.76	6.73	5.28	U/D
8.47	8.64	8.62	8.61	8.29	8.24	5.89	5.80
9.26	9.40	9.74	9.59	8.65	8.87	6.28	6.18
10.3	10.46	10.6	10.84	9.69	9.74	6.79	6.64
				10.53	10.73	7.38	7.54
				11.85	12.12	8.58	8.91
				13.26	13.49	9.38	9.34
				14.05	14.20	9.88	9.88
				14.84	14.92	11.11	11.11
				15.43	15.78	12.09	12.15
				16.15	16.27	13.29	13.22
				17.37	17.44	13.89	13.82
				17.96	18.08	14.53	14.50
				18.51	18.76	15.11	U/D
				20.27	20.52	15.7	15.69
				20.83	21.12	16.22	16.50
				21.63	21.82		
				22.87	23.30		
				24.33	24.59		
				25.75	25.88		
				26.37	26.63		
				27.03	27.08		

Para o vídeo V_1 o valor de RMSE calculado para o resultado foi de 0.18s, o erro mínimo foi de 0.03s e o máximo foi de 0.32s. A distância média entre os valores de referência para os instantes iniciais é de 1.05s, de forma que o RMSE revela uma margem de erro de 17.14%.

Para o vídeo V_2 o valor de RMSE calculado para o resultado foi de 0.11s, o erro mínimo foi de 0.01s e o máximo foi de 0.24s. A distância média entre os valores de referência para os instantes iniciais é de 1.07s, de forma que o RMSE revela uma margem de erro de 10.28%.

Para o vídeo V3 o valor de RMSE calculado para o resultado foi de 0.26s, o erro mínimo foi de 0.03s e o máximo foi de 0.45s. A distância média entre os valores de referência para os instantes iniciais é de 1.12s, de forma que o RMSE revela uma margem de erro de 23.21%.

Para o vídeo V4 o valor de RMSE calculado para o resultado foi de 0.13s, o erro mínimo foi de 0.00s e o máximo foi de 0.33s. A distância média entre os valores de referência para os instantes iniciais é de 1.12s, de forma que o RMSE revela uma margem de erro de 11.61%.

Esta análise revelou um ponto bastante interessante, que é o fato de ser possível estimar a velocidade do fluxo de sinais ao longo dos vídeos, o que pode ser utilizado futuramente em uma fase de pré-processamento. Utilizando a velocidade estimada do fluxo de sinais seria possível criar, automaticamente, uma lista de sinais a serem ajustados em uma *microtask* realizada pela *crowd*. Neste experimento a velocidade média do fluxo de sinais foi de 1 sinal a cada 1.09s.

O produto final gerado na fase de conclusão, para cada um dos quatro vídeos, pode ser observado nos endereços:

Vídeo	URL
V1	http://crowdnoter.com/signs01
V2	http://crowdnoter.com/signs02
V3	http://crowdnoter.com/signs03
V4	http://crowdnoter.com/signs04

5.4 Estudo de Caso 3: Anotação de Efeitos Multissensoriais em Vídeos

O contexto do segundo estudo de caso foi o das aplicações *mulsemedia*. A ideia foi verificar de que forma a *crowd* chegaria a um consenso sobre como os efeitos multissensoriais deveriam ser inseridos em vídeos, mais especificamente os efeitos de vento e vibração. Foram selecionados três vídeos de um *dataset* anotado, de forma que as anotações originais puderam ser utilizadas na interpretação dos resultados para entender de que forma as contribuições da *crowd* poderiam ser utilizadas, de forma alternativa ou complementar.

Os principais resultados deste estudo de caso foram apresentados no artigo (AMORIM et al., 2019), no qual é possível encontrar detalhes sobre os vídeos utilizados, a construção e execução dos *workflow*, assim como discussões mais profundas sobre os resultados.

5.4.1 Preparação do Estudo de Caso 3

Assim como nos estudos de caso anteriores, o recrutamento e pagamento dos *workers* foi feito por meio da plataforma *Microworkers*. Essa plataforma agrupa seus *workers* em diferentes categorias e, no ato da criação da campanha de coleta de contribuições, sugere ao *owner* o grupo que se mostra mais adequado para o tipo de tarefa.

A plataforma sugeriu o grupo *Data Services*, com cerca de 1300 *workers* cadastrados. O orçamento alocado para este estudo de caso foi de 13 dólares, o que permitiu coletar 211 contribuições pagando 0.05 dólar por tarefa realizada, e cobrir as taxas da plataforma. Como a primeira tarefa consistia em identificar instantes que seriam ajustados na tarefa seguinte, foi definido que 150 das contribuições seriam coletadas para a segunda tarefa, e as outras 61 seriam coletadas para a primeira tarefa.

5.4.2 Objetivo do Estudo de Caso 3

O resultado esperado para este processo crowdsourcing são arquivos no padrão *MPEG-V* que pudessem ser visualizado em *softwares* como o *SEVino*² e o *PlaySEM* (SALME; SANTOS, 2015), assim como arquivos no padrão aceito pelo sistema *ELAN*³, que também permite a visualização de anotações de forma sincronizada com os vídeos.

² O *SEVino* está disponível para *download* em <<http://selab.itec.aau.at/2012/07/05/sevino-2-0>>

³ O *ELAN* está disponível para *download* em <<https://tla.mpi.nl/tools/tla-tools/elan>>.

Estes arquivos deveriam conter anotações referentes aos intervalos nos quais a *crowd* acredita ser adequado inserir efeitos de vento e/ou vibração, desconsiderando porém a intensidade pois não foi encontrada uma forma de extrair esta informação da *crowd* com os recursos disponíveis.

5.4.3 Workflow do Estudo de Caso 3

Conforme previsto no *framework*, e se pode verificar no artigo referenciado, a fase de planejamento ocorreu de forma idêntica a do estudo de caso anterior, mostrando que o *framework* pode ser aplicado em diferentes cenários e com diferentes números de tarefas. A fase de execução do *workflow* deste projeto (figura 13) apresenta dois *Task-Sets*:

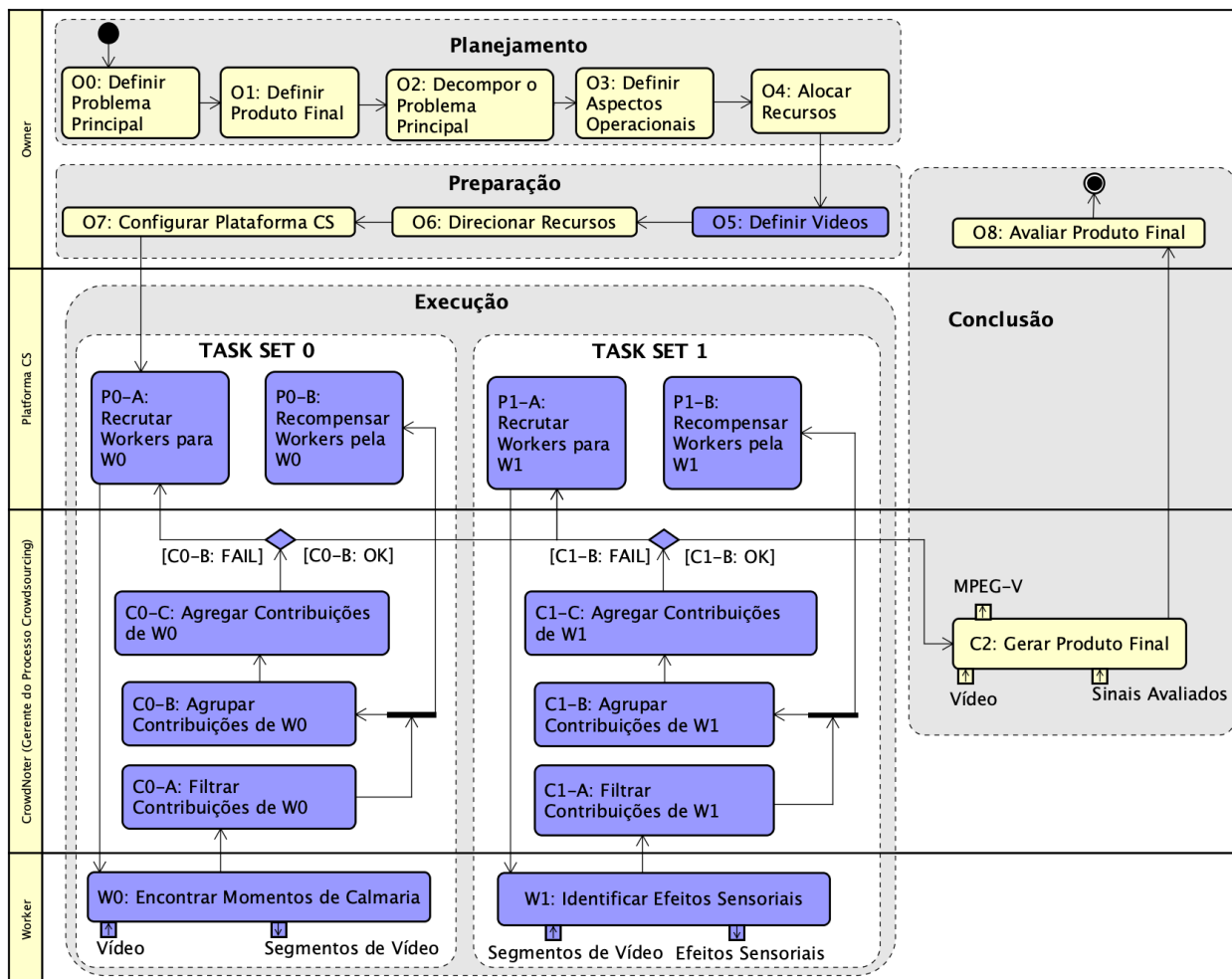


Figura 13 – *Workflow* para anotação de efeitos multissensoriais.

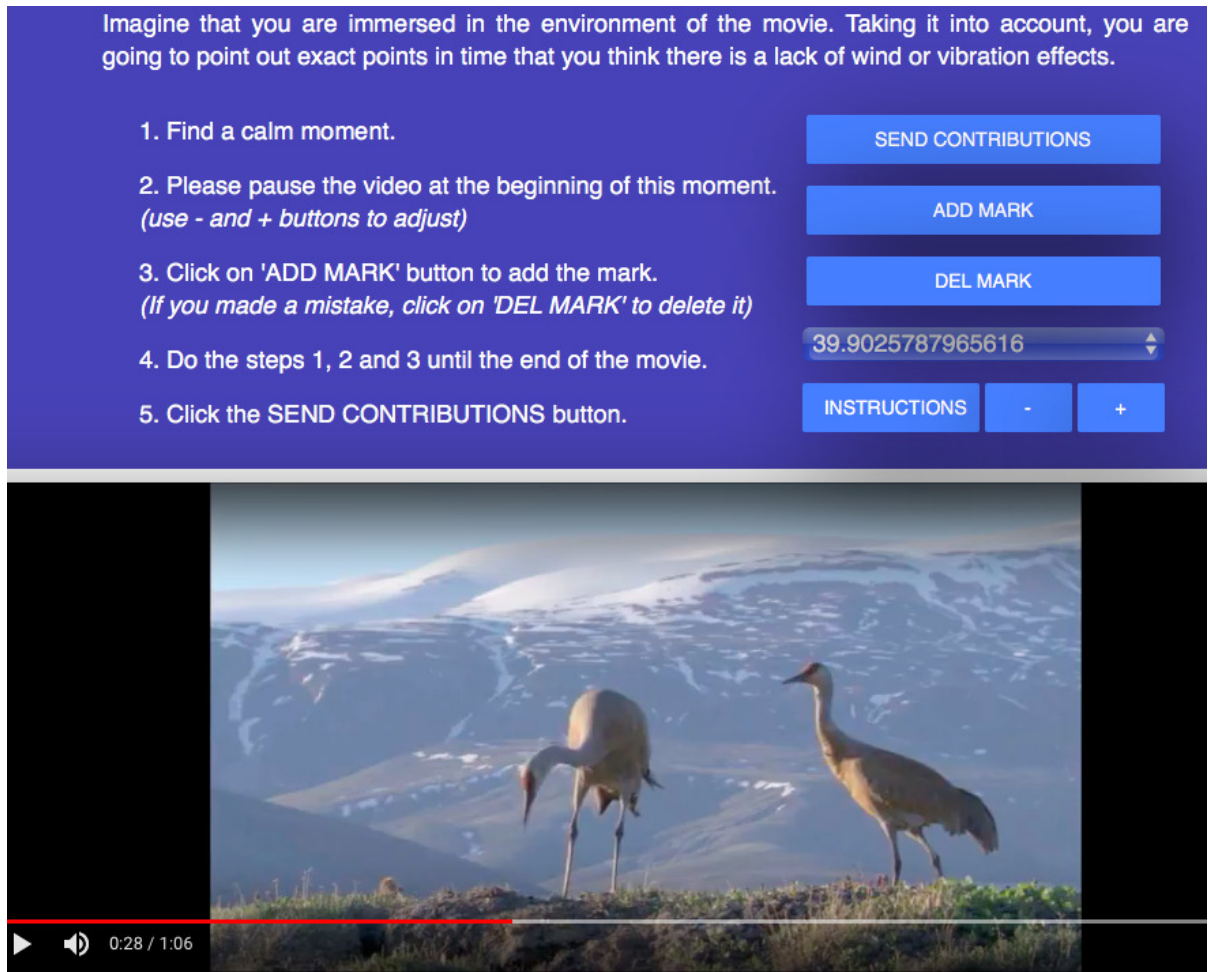


Figura 14 – Ferramenta para a execução da primeira tarefa do estudo de caso 3.

- **Task-Set 0 - Encontrar momentos de calma:** o *worker* recebia o vídeo inteiro e deveria marcar com um clique de botão o início de cada momento de calma em que não devem ser inseridos efeitos sensoriais. Estes momentos resultaram em pontos de segmentação dos intervalos que deveriam receber os efeitos sensoriais. A ferramenta desenvolvida para esta tarefa pode ser observada na figura 14.
- **Task-Set 1 - Identificar efeitos sensoriais:** o *worker* recebia um dos segmentos de vídeo resultantes da tarefa anterior, e deveria identificar qual o efeito a ser inserido nele, assim como ajustar o início e final da inserção. A ferramenta desenvolvida para esta tarefa pode ser observada na figura 15.

Na fase de conclusão, as anotações de efeitos sensoriais obtidas do processo *crowdsourcing* foram utilizadas para gerar os arquivos de metadados no formato do *ELAN* e *MPEG-V*.

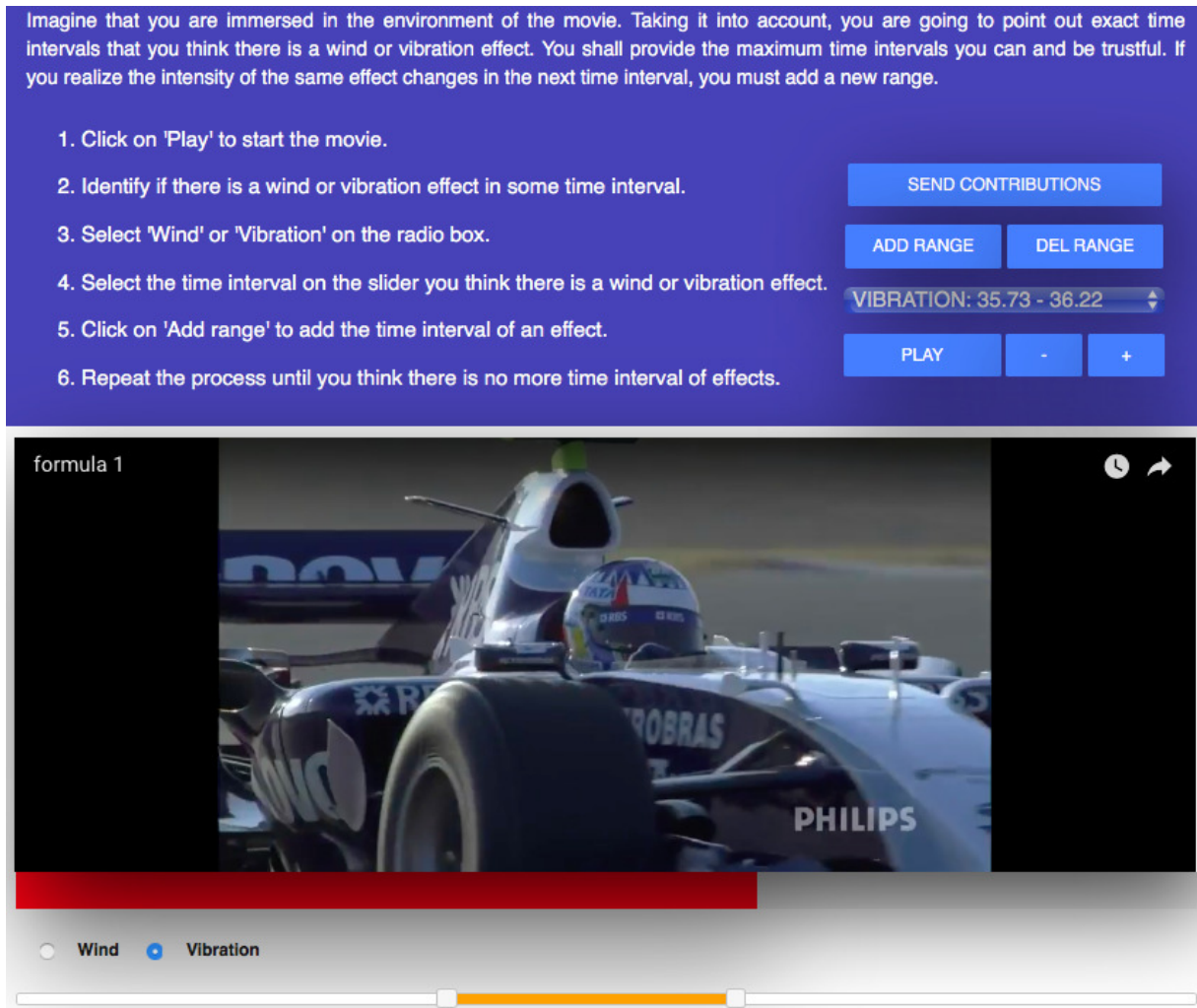


Figura 15 – Ferramenta para a execução da segunda tarefa do estudo de caso 3.

5.4.4 Produto do Estudo de Caso 3

Conforme é possível observar na tabela 21, as contribuições da *crowd* na primeira tarefa resultaram em 38 segmentos para os 3 vídeos utilizados no experimento. Na segunda tarefa, as convergência das contribuições da *crowd* geraram 60 indicações de inserção de efeitos de vento e vibração nos vídeos.

Tabela 21 – Contribuições coletadas e agregadas em cada tarefa.

Task	Itens	Contribuições	Anotações	Filtradas	Agregadas
1	3 vídeos	61	434	306	38 segmentos
2	38 segmentos	150	312	278	60 efeitos
Total	41 itens	211	746	584	98 itens

Este estudo de caso forneceu um resultado interessante, no qual foi possível observar que as inserções de efeitos apontadas pela *crowd* não correspondiam às anotações do *dataset* original. A *crowd* fez associações semânticas próprias entre os efeitos sensoriais e os eventos dos vídeos, como ao associar o efeito de vibração aos momentos em que os carros aceleravam no vídeo que apresentava corridas de Fórmula 1.

A conclusão deste estudo de caso é que as anotações de efeitos multissensoriais realizadas pela *crowd*, ao menos da forma como foi conduzido o experimento, não se mostraram uma alternativa às anotações dos especialistas, mas podem ser utilizadas para identificar pontos de inserção esperados pelos usuários e que foram esquecidos pelos autores. Sendo assim, não foi possível calcular o RMSE para este experimento, uma vez que os vetores com os intervalos obtidos não puderam ser alinhados com os vetores de referência.

Adicionalmente, foi possível concluir que seria positivo adicionar uma tarefa adicional, executada por especialistas, na qual eles se baseariam nas anotações geradas pelo processo para complementar as anotações originais ou as suas próprias anotações.

5.5 Discussão sobre os Resultados

Os estudos de caso realizados demonstraram que o *framework* pode ser aplicado com sucesso em diferentes tipos de aplicação que requerem anotação de vídeo, independente da quantidade de tarefas que o processo demanda.

Os resultados alcançados mostraram que para alguns casos é possível utilizar o *CrowdWaterfall* para obter anotações similares às que seriam produzidas por especialistas, e para outros casos é possível obter anotações que podem ser utilizadas para complementar e melhorar as anotações feitas por eles, como visto no estudo de caso 5.4. Vale destacar que realmente foi possível gerar artefatos complexos com base nas anotações geradas por meio da utilização do *framework*, como é o caso das apresentações multimídia interativas geradas para os estudos de caso 5.2 e 5.3.

Nos estudos de caso 5.2 e 5.3, as tarefas simples foram modeladas de forma que os *workers* não precisariam de nenhum recurso sofisticado para compreender exatamente o que estavam anotando e produzindo. Sendo assim, o resultado obtido da *crowd* foi muito próximo das anotações geradas pelos especialistas.

Por outro lado, no estudo de caso 5.4 foram necessárias tarefas nas quais os *workers* precisavam imaginar a forma como perceberiam os efeitos de vento e vibração, pois não haviam dispositivos disponíveis para isso. A consequência foi que o resultado obtido não foi equivalente ao dos especialistas que contavam com estes equipamentos. Todavia, esse estudo mostrou que anotações geradas sob condições diferentes podem gerar resultados complementares, e foi isso que houve, as anotações geradas pela *crowd* complementava as feitas pelos especialistas no *dataset* original.

Ainda sobre o estudo de caso apresentado na seção 5.4, foi possível observar um ponto interessante. Os *workers* fizeram associações semânticas claramente diferentes das pretendidas pelo autor, e o exemplo mais evidente disso foi a associação do evento de vibração a cada vez que os carros aceleravam no vídeo sobre Fórmula 1. Isso revela que, em alguns casos, as anotações da *crowd* podem ser utilizadas para complementar as feitas pelos especialistas, em vez de oferecer uma alternativa a elas.

6 Conclusão e Trabalhos Futuros

Nesta tese foi apresentado um *framework* chamado *CrowdWaterfall*, que visa facilitar o uso de *crowdsourcing* e computação humana na execução de tarefas complexas de anotação em vídeos. Os cenários atacados foram aqueles nos quais as tarefas requerem características humanas como subjetividade, emoção e empatia para serem executadas.

O foco da pesquisa não foi apresentar uma estratégia concorrente ao trabalho dos especialistas, mas fornecer uma solução que possa ser utilizada quando o trabalho dos especialistas é um fator limitador seja pelo custo associado ou pela dificuldade na prospecção desses profissionais. De forma semelhante, a ideia não foi apresentar uma alternativa aos métodos automáticos, mas um método capaz de cobrir os cenários em que eles encontram dificuldades para serem aplicados, seja por não existirem *datasets* anotados para serem utilizados no treinamento ou seja por não existirem regras claras sobre como estas anotações poderia ser produzidas automaticamente.

O *framework* desenvolvido é baseado em um *workflow* modular no qual é possível inserir tarefas *crowdsourcing* assim como poderiam ser adicionadas tarefas executadas por especialistas ou mesmo por métodos automáticos, não existe nenhum fator que impeça esse tipo de composição de tarefas.

O *CrowdWaterfall* utiliza a estratégia de decompor uma tarefa complexa de anotação de vídeo em um *workflow* formado por tarefas simples de anotação, que são adequadas para serem executadas em campanhas *crowdsourcing*. Foi apresentada uma estratégia para realizar a decomposição, assim como foram apresentados os passos para a definição dos *workflows*.

6.1 Revisitando a Questão de Pesquisa

No início deste trabalho foi levantada uma questão de pesquisa, que será discutida a seguir com base nos resultados alcançados:

QP: *"Como as anotações de vídeo obtidas de um processo crowdsourcing, baseado em um workflow de tarefas pequenas e simples, executadas por pessoas que não precisam ser especialistas, se comparam aos resultados gerados por tarefas grandes ou complexas executadas por especialistas?"*

A questão de pesquisa não trata somente da qualidade das anotações obtidas pelo processo *crowdsourcing*, mas também das suas características. Os estudos de caso demonstraram que é possível utilizar o *CrowdWaterfall* para se obter anotações que apenas seria possíveis utilizando especialistas executando tarefas complexas de anotação de vídeo. Os resultados mostraram que as anotações obtidas pelo *framework* são equivalentes às geradas por especialistas em alguns casos, porém os experimentos também revelaram que tudo depende da forma como as tarefas são concebidas, como o *workflow* é construído, e dos recursos utilizados.

Sendo assim, a resposta para a *QP* é a de que as anotações geradas com base no *CrowdWaterfall* podem ser equivalentes às geradas pelos especialistas ou complementares a elas, de acordo com o planejamento e condições do processo aplicado.

Isso ocorre porque os processos *crowdsourcing* tendem a gerar resultados que refletem a inteligência coletiva, que está relacionada com a convergência da opinião da *crowd*, enquanto os resultados gerados por um especialista podem estar subordinados ao seu ponto de vista e suas opiniões. Nos casos em que isso ocorre, como no estudo de caso 5.4, as anotações produzidas a partir do *CrowdWaterfall*, apesar de não substituírem as anotações dos especialistas, podem ser utilizada para complementa-las com pontos e aspectos que podem ter passados despercebidos pelos especialistas.

6.2 Revisitando os Objetivos

Para avaliar se os objetivos do trabalho foram alcançados, a seguir será feita uma análise de cada um deles considerando os resultados obtidos.

6.2.1 Objetivo Geral

Esta tese visa apresentar um *framework* que inclui o processo *crowdsourcing* que contempla desde a decomposição e a remodelagem da tarefa complexa original como um *workflow* de tarefas simples, a formalização e descrição dos papéis, fases e atividades do processo, os métodos de agregação específicos para pontos e intervalos temporais, e os recursos de software para suportar projetos *crowdsourcing* que demandam tarefas complexas de anotação de vídeo.

Este objetivo foi alcançado pelo *framework CrowdWaterfall*, que nos estudos de caso se mostrou adequado tanto para guiar a decomposição do problema principal de anotação complexa em um conjunto de tarefas simples de anotação de vídeo, quanto para gerar e executar um *workflow* para o processo. Sendo assim, é possível dizer que o objetivo geral da tese foi alcançado.

6.2.2 Objetivo Específico 1

OB1: *Definir as atividades, as fases e os papéis envolvidos no processo que operacionaliza o framework proposto, assim como apresentar um workflow formal para ele.*

Durante este trabalho foi definido processo que conta com um *workflow* bem estruturado, e que é capaz de cobrir todos os requerimentos do *framework* proposto. Além disso foram definidos todos os papéis envolvidos no processo, assim como todas as atividades que cada um deles deve executar. Por fim, os estudos de caso mostraram que o processo é adequado para processos *crowdsourcing* de anotação de vídeo baseados em múltiplas tarefas simples de anotação.

6.2.3 Objetivo Específico 2

OB2: *Desenvolver um ambiente de software que forneça o suporte computacional para a execução do workflow associado ao processo crowdsourcing.*

Com o objetivo de fornecer a parte operacional do *framework CrowdWaterfall*, foi desenvolvido o ambiente computacional *CrowdNoter*, que provê todas as funcionalidades necessárias para suportar a execução do *workflow* definido para o processo, incluindo uma coleção de métodos de agregação e ferramentas de anotação, métodos para a gerência e monitoramento do fluxo de execução, armazenamento das contribuições e resultados agregados, entre outros.

6.3 Contribuições

A contribuição principal desta tese é o *framework crowdsourcing* chamado *CrowdWaterfall*, que se baseia na decomposição de uma tarefa complexa de anotação de vídeo em um conjunto de tarefas *crowdsourcing* simples, que são organizadas em um *workflow* que pode ser executado para produzir anotações de vídeo que tradicionalmente requerem o trabalho de especialistas. As tarefas simples utilizadas para compor o *workflow* seguem o conceito de *microtarefa*, podendo ser executadas a partir de campanhas *crowdsourcing* por *workers* sem qualificação ou habilidades específicas, e sem a necessidade de dedicar tempo ou esforço significativo. Este *framework* facilita a utilização de *crowdsourcing* em um conjunto de aplicações que demandam atividades complexas de anotação de vídeo que necessitam do trabalho de especialistas para serem realizadas.

Outra contribuição importante é a definição do *workflow* bem estruturado que é utilizado no processo apresentado. Este *workflow* mostrou-se adequado para guiar processos de anotação de vídeo, mas nada impede que seja utilizado em outras aplicações *crowdsourcing* que requerem a utilização de múltiplas tarefas simples.

As técnicas e métodos desenvolvidos no decorrer do projeto também são contribuições importantes uma vez que oferecem recursos que podem ser úteis mesmo para projetos que não utilizem o *CrowdWaterfall*.

Entre estas funcionalidades vale ressaltar a possibilidade de criar uma separação entre a plataforma comercial *crowdsourcing* e o sistema responsável por gerenciar o fluxo de execução do processo *crowdsourcing* e a distribuição das tarefas. Isso permite que se utilize as plataformas comerciais de *crowdsourcing* apenas para recrutar e recompensar os *workers*. A divisão entre a plataforma comercial e o ambiente de execução do processo permite que as contribuições coletadas e resultados agregados sejam armazenados exclusivamente no local em que o *owner* ache adequado. Isso permite que empresas que lidam com dados sigilosos passem a considerar o *crowdsourcing* como uma possível abordagem para seus projetos que utilizam dados sigilosos.

Mais uma contribuição produzida neste trabalho é uma solução para o controle de latência nas contribuições *crowdsourcing*. Essa funcionalidade mostra-se especialmente útil quando é necessário acelerar o processo de coleta de contribuições sem abrir mão do monitoramento da convergência dos resultados em tempo real.

A generalização dos métodos de agregação temporal voltados para tarefas *crowdsourcing* de anotação de instantes ou intervalos também é uma contribuição importante, uma vez que eles podem ser reaproveitados em um grande número de tarefas muito comuns nos processos *crowdsourcing*.

A própria estratégia para monitoramento e gerência do *workflow crowdsourcing* é uma contribuição relevante desta tese, uma vez que fornece meios para o gerente do processo *crowdsourcing* acompanhar, em tempo real, o estado do resultado obtido e controlar o *workflow*, seja interrompendo ou retomando o processo, seja direcionando arbitrariamente o fluxo de execução para uma das tarefas.

Por fim, foi criado um ambiente de software, chamado *CrowdNoter*, que implementa todas as funcionalidades citadas nesta seção. Este ambiente é livre e aberto, de forma que pode ser modificado e utilizado por grupos de pesquisa ou qualquer pessoa que deseje implementar processos *crowdsourcing* baseados em um *workflow* de tarefas simples. Este ambiente pode colaborar na popularização deste tipo de processo *crowdsourcing*.

Entres as contribuições acadêmicas é possível listar as publicações alcançadas, que serão apresentadas a seguir.

6.3.1 Publicações

A seguir serão apresentadas as publicações produzidas ao longo desta pesquisa.

1. AMORIM, M. N. de; SALEME, E. B.; NETO, F. R. d. A.; SANTOS, C. A. S.; GHINEA, G. **Crowdsourcing authoring of sensory effects on videos**. Multimedia Tools and Applications, Feb 2019. ISSN 1573-7721. DOI:10.1007/s11042-019-7312-2. URL: <<https://doi.org/10.1007/s11042-019-7312-2>>
2. AMORIM, M. N.; NETO, F. R. A.; SANTOS, C. A. S. **Achieving complex media annotation through collective wisdom and effort from the crowd**. In: 2018 25th International Conference on Systems, Signals and Image Processing (IWSSIP). 2018. p. 1–5. ISSN 2157-8702. DOI: 10.1109/IWSSIP.2018.8439402. URL: <<https://ieeexplore.ieee.org/document/8439402>>
3. AMORIM, M.; MENDES, R.; SANTOS, C. A. S.; TAVARES, O. **Crowdnote: Crowdsourcing environment for complex video annotations**. In: XVI WFA 2017. ISBN: 978-85-7669-380-2. URL: <<https://sol.sbc.org.br/index.php/webmediaestendido/article/view/4862/4768>>

4. MENDES, R.; AMORIM, M. de; SANTOS, C. **Crowdsync: User generated videos synchronization using crowdsourcing**. In: IEEE. Systems, Signals and Image Processing (IWSSIP), 2017 International Conference on 2017. p. 1–5. DOI: 10.1109/IWSSIP.2017.7965599. URL: <<https://ieeexplore.ieee.org/document/7965599>>
5. AMORIM, M.; SEGUNDO, R.; SANTOS, C. S. **Livesync: a method for real time videostreaming synchronization from independent sources**. 12 2017. URL: <<https://periodicos.ufmg.br/index.php/jidm/article/view/377/266>>
6. AMORIM, M. N. de; SEGUNDO, R. M.; SANTOS, C. A.; TAVARES, O. d. L. **Video annotation by cascading microtasks: A crowdsourcing approach**. In: Proceedings of the 23rd Brazillian Symposium on Multimedia and the Web. New York, NY, USA: ACM, 2017. (WebMedia '17), p. 49–56. ISBN 978-1-4503-5096-9. DOI: 10.1145/3126858.3126897. URL: <<http://doi.acm.org/10.1145/3126858.3126897>>
7. SEGUNDO, R.; AMORIM, M.; SANTOS, C. S. **Crowdsourcing multimedia: Enhancing multimedia activities with the power of crowds**. 2016. p. 11–12. DOI: 10.1145/2976796.2988223. URL: <<https://sol.sbc.org.br/livros/index.php/sbc/catalog/view/14/47/97-1>>
8. AMORIM, M. N.; SANTOS, C. A. S.; TAVARES, O. **Excam - uma metodologia acrowdsourcing para a autoria de conteúdo extra para videos**. In: WebMedia 2016 WTD. Teresina - PI, Brazil: [s.n.], 2016. ISBN: 978-85-7669-332-1. URL: <https://sol.sbc.org.br/index.php/webmedia_estendido/article/download/4881/4787>
9. AMORIM, M. N. de; SEGUNDO, R. M.; SANTOS, C. A. **Livesync: a tool for real time video streaming synchronization from independent sources**. In: WebMedia 2016 WFA. Teresina - PI, Brazil, 2016. ISBN: 978-85-7669-332-1. URL: <https://sol.sbc.org.br/index.php/webmedia_estendido/article/view/4894/4800>

6.4 Limitações

O trabalho apresentado alcançou os objetivos pretendidos, porém ele ainda apresenta um série de limitações que precisam ser consideradas. Algumas delas são relacionadas com os recursos produzidos, outras residem nas características do *crowdsourcing*.

A primeira limitação do *framework*, e dos processos *crowdsourcing* como em geral, é que não se pode garantir uma expectativa de qualidade para as contribuições da *crowd*. Existem diversos fatores que influenciam na qualidade das contribuições e existem técnicas de filtragem e validação que podem melhorar a qualidade, porém o fato de o *framework* utilizar o trabalho de *crowd workers* adiciona uma variável imprevisível ao sistema, que é o comportamento da *crowd*. Existem formas de estimar este comportamento e o nível de qualidade, e estes meios geralmente fornecem boas estimativas, mas não se deve excluir do processo o fator associado com a imprevisibilidade do comportamento humano.

A segunda limitação do trabalho está relacionada à automatização do *workflow*. Apesar do *framework* suportar diferentes tipos de controle de fluxo, sejam manuais ou automáticos, a transição automática entre as tarefas não foi implementada. Na atual versão é necessário que o gerente do processo *crowdsourcing* realize manualmente a transição entre as tarefas do *workflow*.

A terceira limitação diz respeito à praticidade na utilização do *framework*. Não foi produzida uma ferramenta visual e intuitiva para a decomposição da tarefa original e criação do *workflow*. Adicionalmente, o ambiente computacional requer alguma programação para ser adaptado a cada cenário de aplicação. Estes fatores podem inibir pessoas com menos habilidades relacionadas com desenvolvimento a utilizar o *CrowdWaterfall* em seus projetos.

6.5 Trabalhos Futuros

O *framework* apresentado abre diversas possibilidades de trabalhos futuros, alguns de caráter mais acadêmico outros mais técnicos, havendo ainda alguns estudos de caso que poderiam gerar contribuições importantes para a comunidade científica e para a sociedade em geral. A seguir serão listados alguns trabalhos interessantes que poderiam ser realizados com o uso do *CrowdWaterfall*, assim como melhorias para cobrir algumas de suas limitações.

- **Criação e Execução de *Workflows* Híbridos**

Um ponto relevante a se pesquisar é de que forma se comportaria o *CrowdWaterfall* caso fossem utilizadas, em um mesmo *workflow*, tarefas *crowdsourcing*, tarefas feitas por especialistas e tarefas realizadas por métodos automáticos. O uso conjunto dessas três modalidades é algo que merece ser investigado pois isso abriria diversas possibilidades.

- **Geração de *Datasets* Anotados**

Uma aplicação direta do *CrowdWaterfall* seria na implementação de sistemas com o objetivo de anotar *datasets* que possam ser utilizados no treinamento de modelo de *machine learning*. As anotações produzidas seriam aquelas que não poderiam ser obtidas por *workers* não capacitados em um processo *crowdsourcing* tradicional, justificando o uso do *framework* para produzir anotações que atualmente requeiram o trabalho de especialistas.

- **Construção de um Assistente para o *CrowdWaterfall***

Uma contribuição técnica importante para a difusão do *framework* desenvolvido seria a criação de um módulo visual capaz de permitir que o *owner* cadastre um novo projeto e o auxilie na decomposição da tarefa principal e na construção do *workflow*.

- **Criação de Sistemas de Anotação por Demanda**

Outra aplicação interessante seria utilizar a possibilidade que o *CrowdNoter* oferece de ter uma lista circular de itens a serem anotados sendo executada continuamente, para se gerar um sistema que fique em execução contínua, permitindo que a qualquer momento sejam adicionados novos itens para serem anotados.

- **Objetos de Aprendizagem e Tecnologias Assistivas**

Conforme observado nos estudos de caso, o *CrowdWaterfall* é adequado para aplicações relacionada com tecnologias assistivas e com a geração de apresentações multimídia interativas. Sendo assim, seria interessante que fosse conduzido um trabalho que integrasse os estudos de caso 5.2 e 5.3 em um sistema capaz de produzir objetos de aprendizagem voltados para o estudo de LIBRAS.

Referências

- ABBAS, Q.; IBRAHIM, M. E.; JAFFAR, M. A. Video scene analysis: an overview and challenges on deep learning algorithms. *Multimedia Tools and Applications*, Springer, p. 1–39, 2017. Citado na página 2.
- AHN, L. V. *Human Computation*. Tese (Doutorado) — Carnegie Mellon University, Pittsburgh, PA, USA, 2005. AAI3205378. Citado 3 vezes nas páginas 3, 13 e 18.
- AHN, L. V. Games with a purpose. *Computer*, IEEE, v. 39, n. 6, p. 92–94, 2006. Citado na página 15.
- AHN, L. v. *Massive-scale online collaboration*. TED, 2011. Disponível em: http://www.ted.com/talks/luis_von_ahn_massive_scale_online_collaboration. Citado na página 13.
- AMBATI, V.; VOGEL, S.; CARBONELL, J. Collaborative workflow for crowdsourcing translation. In: *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*. New York, NY, USA: ACM, 2012. (CSCW '12), p. 1191–1194. ISBN 978-1-4503-1086-4. Disponível em: <http://doi.acm.org/10.1145/2145204.2145382>. Citado na página 35.
- AMORIM, M. N.; NETO, F. R. A.; SANTOS, C. A. S. Achieving complex media annotation through collective wisdom and effort from the crowd. In: *2018 25th International Conference on Systems, Signals and Image Processing (IWSSIP)*. [S.l.: s.n.], 2018. p. 1–5. ISSN 2157-8702. Citado na página 71.
- AMORIM, M. N. de; SALEME, E. B.; NETO, F. R. d. A.; SANTOS, C. A. S.; GHINEA, G. Crowdsourcing authoring of sensory effects on videos. *Multimedia Tools and Applications*, Feb 2019. ISSN 1573-7721. Disponível em: <https://doi.org/10.1007/s11042-019-7312-2>. Citado na página 84.
- AMORIM, M. N. de; SEGUNDO, R. M.; SANTOS, C. A. Livesync: a tool for real time video streaming synchronization from independent sources. In: *WebMedia 2016 WFA*. Teresina - PI, Brazil: [s.n.], 2016. Citado na página 29.
- AMORIM, M. N. de; SEGUNDO, R. M.; SANTOS, C. A.; TAVARES, O. d. L. Video annotation by cascading microtasks: A crowdsourcing approach. In: *Proceedings of the 23rd Brazillian Symposium on Multimedia and the Web*. New York, NY, USA: ACM, 2017. (WebMedia '17), p. 49–56. ISBN 978-1-4503-5096-9. Disponível em: <http://doi.acm.org/10.1145/3126858.3126897>. Citado 2 vezes nas páginas 3 e 71.
- ANEGEKUH, L.; SUN, L. A screening methodology for crowdsourcing video qoe evaluation. In: IEEE. *Global Communications Conference (GLOBECOM), 2014 IEEE*. [S.l.], 2014. p. 1152–1157. Citado na página 20.

- BERNSTEIN, M. S. et al. Soylent: A word processor with a crowd inside. In: *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology*. New York, NY, USA: ACM, 2010. (UIST '10), p. 313–322. ISBN 978-1-4503-0271-5. Disponível em: <<http://doi.acm.org/10.1145/1866029.1866078>>. Citado 5 vezes nas páginas 32, 35, 36, 37 e 40.
- BERTINI, M. et al. Socially-aware video recommendation using users' profiles and crowdsourced annotations. In: ACM. *Proceedings of the 2nd international workshop on Socially-aware multimedia*. [S.l.], 2013. p. 13–18. Citado na página 29.
- BIEL, J.-I.; GATICA-PEREZ, D. The youtube lens: Crowdsourced personality impressions and audiovisual analysis of vlogs. *Multimedia, IEEE Transactions on*, IEEE, v. 15, n. 1, p. 41–55, 2013. Citado 3 vezes nas páginas 13, 28 e 30.
- CHEN, C.; MENG, X.; ZHAO, S.; FJELD, M. Retool: Interactive microtask and workflow design through demonstration. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2017. (CHI '17), p. 3551–3556. ISBN 978-1-4503-4655-9. Disponível em: <<http://doi.acm.org/10.1145/3025453.3025969>>. Citado 2 vezes nas páginas 35 e 40.
- CHEN, S.; LI, M.; REN, K.; QIAO, C. Crowd map: Accurate reconstruction of indoor floor plans from crowdsourced sensor-rich videos. In: IEEE. *Distributed Computing Systems (ICDCS), 2015 IEEE 35th International Conference on*. [S.l.], 2015. p. 1–10. Citado na página 29.
- CHILTON, L. B.; LITTLE, G.; EDGE, D.; WELD, D. S.; LANDAY, J. A. Cascade. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, p. 1999, 2013. Disponível em: <<http://dl.acm.org/citation.cfm?doid=2470654.2466265>>. Citado na página 33.
- CHITTILAPPILLY, A. I.; CHEN, L.; AMER-YAHIA, S. A survey of general-purpose crowdsourcing techniques. *IEEE Transactions on Knowledge and Data Engineering*, IEEE, v. 28, n. 9, p. 2246–2266, 2016. Citado 6 vezes nas páginas 17, 30, 31, 32, 44 e 46.
- CROSS, A.; BAYYAPUNEDI, M.; RAVINDRAN, D.; CUTRELL, E.; THIES, W. Vidwiki: enabling the crowd to improve the legibility of online educational videos. In: ACM. *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. [S.l.], 2014. p. 1167–1175. Citado 2 vezes nas páginas 13 e 26.
- CROSS, A.; BAYYAPUNEDI, M.; RAVINDRAN, D.; CUTRELL, E.; THIES, W. Vidwiki: Enabling the crowd to improve the legibility of online educational videos. In: ACM. *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. [S.l.], 2014. p. 1167–1175. Citado 2 vezes nas páginas 30 e 31.
- DESELL, T.; GOEHNER, K.; ANDES, A.; ECKROAD, R.; ELLIS-FELEGE, S. On the effectiveness of crowd sourcing avian nesting video analysis at wildlife@ home. *Procedia Computer Science*, Elsevier, v. 51, p. 384–393, 2015. Citado na página 29.

- DESHPANDE, R.; TUNA, T.; SUBHLOK, J.; BARKER, L. A crowdsourcing caption editor for educational videos. In: IEEE. *Frontiers in Education Conference (FIE), 2014 IEEE*. [S.l.], 2014. p. 1–8. Citado 2 vezes nas páginas 13 e 27.
- DIFALLAH, D. E.; CATASTA, M.; DEMARTINI, G.; IPEIROTIS, P. G.; CUDRÉ-MAUROUX, P. The dynamics of micro-task crowdsourcing: The case of amazon mturk. In: *Proceedings of the 24th International Conference on World Wide Web*. New York, NY, USA: ACM, 2015. (WWW '15), p. 238–247. ISBN 978-1-4503-3469-3. Disponível em: <<http://doi.acm.org/10.1145/2736277.2741685>>. Citado na página 13.
- ELLIS, C. A.; GIBBS, S. J.; REIN, G. Groupware: some issues and experiences. *Communications of the ACM*, ACM, v. 34, n. 1, p. 39–58, 1991. Citado na página 4.
- FAYAD, M. E.; SCHMIDT, D. C.; JOHNSON, R. E. *Building application frameworks: object-oriented foundations of framework design*. [S.l.]: John Wiley & Sons, Inc., 1999. Citado 2 vezes nas páginas 6 e 43.
- FORD, C. Visual Storytelling. *Art History*, v. 15, n. 4, p. 533–537, 2016. ISSN 14678365. Disponível em: <<http://arxiv.org/abs/1604.03968>>. Citado na página 38.
- FREIBURG, B.; KAMPS, J.; SNOEK, C. G. Crowdsourcing visual detectors for video search. In: ACM. *Proceedings of the 19th ACM international conference on Multimedia*. [S.l.], 2011. p. 913–916. Citado na página 29.
- FREITAS, W. R.; JABBOUR, C. J. Utilizando estudo de caso (s) como estratégia de pesquisa qualitativa: boas práticas e sugestões. *Revista Estudo & Debate*, v. 18, n. 2, 2011. Citado na página 9.
- GADGIL, N. J.; TAHBOUB, K.; KIRSH, D.; DELP, E. J. A web-based video annotation system for crowdsourcing surveillance videos. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *IS&T/SPIE Electronic Imaging*. [S.l.], 2014. p. 90270A–90270A. Citado 4 vezes nas páginas 13, 25, 27 e 31.
- GALTON, F. Vox populi (the wisdom of crowds). *Nature*, Springer Nature, v. 75, n. 1949, p. 450–451, 1907. Disponível em: <<https://doi.org/10.1038%2F075509f0>>. Citado 3 vezes nas páginas 3, 11 e 18.
- GOTTLIEB, L.; CHOI, J.; KELM, P.; SIKORA, T.; FRIEDLAND, G. Pushing the limits of mechanical turk: qualifying the crowd for video geo-location. In: ACM. *Proceedings of the ACM multimedia 2012 workshop on Crowdsourcing for multimedia*. [S.l.], 2012. p. 23–28. Citado na página 29.
- HAAS, D.; ANSEL, J.; GU, L.; MARCUS, A. Argonaut: Macrotask crowdsourcing for complex data processing. *Proc. VLDB Endow.*, VLDB Endowment, v. 8, n. 12, p. 1642–1653, ago. 2015. ISSN 2150-8097. Disponível em: <<http://dx.doi.org/10.14778/2824032.2824062>>. Citado na página 13.
- HAN, C.-H.; LEE, J.-S. Quality assessment of on-line videos using metadata. In: IEEE.

Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on. [S.l.], 2014. p. 1385–1388. Citado na página 29.

HEILBRON, F. C.; NIEBLES, J. C. Collecting and annotating human activities in web videos. In: ACM. *Proceedings of International Conference on Multimedia Retrieval*. [S.l.], 2014. p. 377. Citado na página 36.

HENDRICKX, I. et al. A multilingual wikified data set of educational material. 05 2018. Citado 2 vezes nas páginas 13 e 24.

HOSSEINI, M.; PHALP, K.; TAYLOR, J.; ALI, R. The four pillars of crowdsourcing: A reference model. In: *2014 IEEE Eighth International Conference on Research Challenges in Information Science (RCIS)*. [S.l.: s.n.], 2014. p. 1–12. ISSN 2151-1349. Citado 2 vezes nas páginas 13 e 14.

HOWE, J. The rise of crowdsourcing. *Wired Magazine*, v. 14, n. 6, 06 2006. Disponível em: <<http://www.wired.com/wired/archive/14.06/crowds.html>>. Citado 2 vezes nas páginas 13 e 18.

HUBERMAN, B. A.; ROMERO, D. M.; WU, F. Crowdsourcing, attention and productivity. *Journal of Information science*, Sage Publications Sage UK: London, England, v. 35, n. 6, p. 758–765, 2009. Citado na página 18.

HUNG, N. Q. V.; TAM, N. T.; TRAN, L. N.; ABERER, K. An evaluation of aggregation techniques in crowdsourcing. In: SPRINGER. *International Conference on Web Information Systems Engineering*. [S.l.], 2013. p. 1–15. Citado 2 vezes nas páginas 18 e 19.

HURON SAMUEL, I. P. F. J. D. Polemictweet: Video annotation and analysis through tagged tweets. In: *Human-Computer Interaction–INTERACT 2013*. [S.l.]: Springer, 2013. p. 135–152. Citado na página 29.

KACORRI, H.; SHINKAWA, K.; SAITO, S. Introducing game elements in crowdsourced video captioning by non-experts. In: ACM. *Proceedings of the 11th Web for All Conference*. [S.l.], 2014. p. 29. Citado na página 29.

KHATTAK, F. K.; SALLEB-AOUISSI, A. Quality control of crowd labeling through expert evaluation. In: *Proceedings of the NIPS 2nd Workshop on Computational Social Science and the Wisdom of Crowds*. [S.l.: s.n.], 2011. v. 2. Citado na página 20.

KIM, G.; SIGAL, L.; XING, E. P. Joint summarization of large-scale collections of web images and videos for storyline reconstruction. In: *Proceedings of the 2014 IEEE CCVPR*. Washington, DC, USA: IEEE Computer Society, 2014. (CVPR '14), p. 4225–4232. ISBN 978-1-4799-5118-5. Disponível em: <<http://dx.doi.org/10.1109/CVPR.2014.538>>. Citado 2 vezes nas páginas 35 e 37.

KITTUR, A.; KHAMKAR, S.; ANDRÉ, P.; KRAUT, R. Crowdweaver: visually managing complex crowd work. In: ACM. *CSCW - Proceedings of the ACM 2012*

conference on Computer Supported Cooperative Work. [S.l.], 2012. p. 1033–1036. Citado na página 39.

KITTUR, A.; SMUS, B.; KHAMKAR, S.; KRAUT, R. E. Crowdforge: Crowdsourcing complex work. In: *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*. New York, NY, USA: ACM, 2011. (UIST '11), p. 43–52. ISBN 978-1-4503-0716-1. Disponível em: <<http://doi.acm.org/10.1145/2047196.2047202>>. Citado 2 vezes nas páginas 34 e 39.

KOZAR, O. Towards better group work: Seeing the difference between cooperation and collaboration. In: ERIC. *English Teaching Forum*. [S.l.], 2010. v. 48, n. 2, p. 16–23. Citado na página 5.

KULKARNI, A.; CAN, M.; HARTMANN, B. Collaboratively crowdsourcing workflows with turkomatic. In: *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*. New York, NY, USA: ACM, 2012. (CSCW '12), p. 1003–1012. ISBN 978-1-4503-1086-4. Disponível em: <<http://doi.acm.org/10.1145/2145204.2145354>>. Citado 2 vezes nas páginas 35 e 39.

KUNCHEVA, L. I.; WHITAKER, C. J.; SHIPP, C. A.; DUIN, R. P. Limits on the majority vote accuracy in classifier fusion. *Pattern Analysis & Applications*, Springer, v. 6, n. 1, p. 22–31, 2003. Citado na página 19.

LAW, E.; AHN, L. von. Human computation. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, v. 5, n. 3, p. 1–121, 2011. Disponível em: <<http://dx.doi.org/10.2200/S00371ED1V01Y201107AIM013>>. Citado na página 12.

LE, J.; EDMONDS, A.; HESTER, V.; BIEWALD, L. Ensuring quality in crowdsourced search relevance evaluation: The effects of training question distribution. In: *SIGIR 2010 workshop on crowdsourcing for search evaluation*. [S.l.: s.n.], 2010. p. 21–26. Citado na página 20.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, Nature Research, v. 521, n. 7553, p. 436–444, 2015. Citado na página 13.

LÉVY, P.; COSTA, C. I. da. *tecnologias da inteligência, As*. [S.l.]: Editora 34, 1993. Citado 3 vezes nas páginas 3, 12 e 18.

LI, G.; WANG, J.; ZHENG, Y.; FRANKLIN, M. Crowdsourced data management: A survey. In: IEEE. *Data Engineering (ICDE), 2017 IEEE 33rd International Conference on*. [S.l.], 2017. p. 39–40. Citado 3 vezes nas páginas 17, 30 e 65.

LITTLE, G.; CHILTON, L. B.; GOLDMAN, M.; MILLER, R. C. TurkIt: Human computation algorithms on mechanical turk. In: *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology*. New York, NY, USA: ACM, 2010. (UIST '10), p. 57–66. ISBN 978-1-4503-0271-5. Disponível em: <<http://doi.acm.org/10.1145/1866029.1866040>>. Citado 2 vezes nas páginas 33 e 39.

- LOWRY, C. S.; FIENEN, M. N. Crowdhydrology: crowdsourcing hydrologic data and engaging citizen scientists. *GroundWater*, Wiley Online Library, v. 51, n. 1, p. 151–156, 2013. Citado na página 69.
- LUZ, N.; SILVA, N.; NOVAIS, P. A survey of task-oriented crowdsourcing. *Artificial Intelligence Review*, v. 44, n. 2, p. 187–213, Aug 2015. ISSN 1573-7462. Disponível em: <<https://doi.org/10.1007/s10462-014-9423-5>>. Citado na página 47.
- MANZATO, M. G. *Uma arquitetura de personalização de conteúdo baseada em anotações do usuário*. Tese (Doutorado), 2011. Citado na página 4.
- MATTSSON, M. Object-oriented frameworks-a survey of methodological issues, 'licentiate thesis. *Department of Computer Science, Lund University*, 1996. Citado 2 vezes nas páginas 6 e 43.
- MENDES, R.; AMORIM, M. de; SANTOS, C. Crowdsync: User generated videos synchronization using crowdsourcing. In: IEEE. *Systems, Signals and Image Processing (IWSSIP), 2017 International Conference on*. [S.l.], 2017. p. 1–5. Citado 3 vezes nas páginas 13, 21 e 30.
- MIHALCEA, R.; CSOMAI, A. Wikify!: Linking documents to encyclopedic knowledge. In: *Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management*. New York, NY, USA: ACM, 2007. (CIKM '07), p. 233–242. ISBN 978-1-59593-803-9. Disponível em: <<http://doi.acm.org/10.1145/1321440.1321475>>. Citado na página 1.
- MISANCHUK, M.; ANDERSON, T. Building community in an online learning environment: Communication, cooperation and collaboration. ERIC, 2001. Citado na página 3.
- MO, L. et al. Optimizing plurality for human intelligence tasks. In: *Proceedings of the 22Nd ACM International Conference on Information & Knowledge Management*. New York, NY, USA: ACM, 2013. (CIKM '13), p. 1929–1938. ISBN 978-1-4503-2263-8. Disponível em: <<http://doi.acm.org/10.1145/2505515.2505755>>. Citado na página 13.
- NEGRI, M.; BENTIVOGLI, L.; MEHDAD, Y.; GIAMPICCOLO, D.; MARCHETTI, A. Divide and conquer: Crowdsourcing the creation of cross-lingual textual entailment corpora. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. [S.l.], 2011. p. 670–679. Citado na página 33.
- NETO, F. R. A. *A general-purpose conceptual model for crowdsourcing project*. 88 p. Dissertação (Mestrado) — UFES - Universidade Federal do Espírito Santo, Vitória, ES, Brasil, 2018. Citado na página 15.
- NETO, F. R. A.; SANTOS, C. A. S. Understanding crowdsourcing projects: A systematic review of tendencies, workflow, and quality management. *Information Processing &*

Management, Elsevier, v. 54, n. 4, p. 490–506, 2018. ISSN 0306-4573. Citado na página 15.

NORONHA, J.; HYSEN, E.; ZHANG, H.; GAJOS, K. Z. Platemate: Crowdsourcing nutritional analysis from food photographs. In: *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*. New York, NY, USA: ACM, 2011. (UIST '11), p. 1–12. ISBN 978-1-4503-0716-1. Disponível em: <<http://doi.acm.org/10.1145/2047196.2047198>>. Citado na página 33.

OUYANG, R. W. et al. Debiasing crowdsourced quantitative characteristics in local businesses and services. In: ACM. *Proceedings of the 14th International Conference on Information Processing in Sensor Networks*. [S.l.], 2015. p. 190–201. Citado na página 69.

PAL, G.; ACHARJEE, S.; RUDRAPAUL, D.; ASHOUR, A. S.; DEY, N. Video segmentation using minimum ratio similarity measurement. *International journal of image mining*, Inderscience Publishers (IEL), v. 1, n. 1, p. 87–110, 2015. Citado na página 2.

PARK SUNGHYUN, S. P. M. L.-P. Toward crowdsourcing micro-level behavior annotations: the challenges of interface, training, and generalization. In: ACM. *Proceedings of the 19th international conference on Intelligent User Interfaces*. [S.l.], 2014. p. 37–46. Citado na página 29.

PÉREZ, D. G. de G.; SUOKAS, M.; BEDNARIK, R. Crowdsourcing pupil annotation datasets: Boundary vs. center, what performs better? In: *Proceedings of the 7th Workshop on Pervasive Eye Tracking and Mobile Eye-Based Interaction*. New York, NY, USA: ACM, 2018. (PETMEI '18), p. 3:1–3:7. ISBN 978-1-4503-5789-0. Disponível em: <<http://doi.acm.org/10.1145/3208031.3208036>>. Citado 2 vezes nas páginas 13 e 23.

PIMENTEL, M. d. G. C.; CATTELAN, R. G.; MELO, E. L.; FREITAS, G. B.; TEIXEIRA, C. A. Watch-and-comment as an approach to collaboratively annotate points of interest in video and interactive-tv programs. In: _____. *Mobile TV: Customizing Content and Experience: Mobile Storytelling, Creation and Sharing*. London: Springer London, 2010. p. 349–368. ISBN 978-1-84882-701-1. Citado na página 4.

PINTO, J. P.; VIANA, P. Tag4vd: a game for collaborative video annotation. In: ACM. *Proceedings of the 2013 ACM international workshop on Immersive media experiences*. [S.l.], 2013. p. 25–28. Citado na página 29.

RAINER, B.; TIMMERER, C. A quality of experience model for adaptive media playout. In: IEEE. *Quality of Multimedia Experience (QoMEX), 2014 Sixth International Workshop on*. [S.l.], 2014. p. 177–182. Citado na página 20.

RIEK, L. D.; O'CONNOR, M. F.; ROBINSON, P. Guess what? a game for affective annotation of video using crowd sourcing. In: *Affective computing and intelligent interaction*. [S.l.]: Springer, 2011. p. 277–285. Citado na página 29.

- ROHWER, P. A note on human computation limits. In: *Proceedings of the ACM SIGKDD Workshop on Human Computation*. New York, NY, USA: ACM, 2010. (HCOMP '10), p. 38–40. ISBN 978-1-4503-0222-7. Disponível em: <http://doi.acm.org/10.1145/1837885.1837897>. Citado 2 vezes nas páginas 13 e 14.
- SALEME, E. a. B.; SANTOS, C. A. S. Playsem: A platform for rendering mulsemmedia compatible with mpeg-v. In: *Proceedings of the 21st Brazilian Symposium on Multimedia and the Web*. New York, NY, USA: ACM, 2015. (WebMedia '15), p. 145–148. ISBN 978-1-4503-3959-9. Disponível em: <http://doi.acm.org/10.1145/2820426.2820450>. Citado na página 84.
- SANTOS, C.; SANTOS, A.; TAVARES, T. *Uma estratégia para a construção de ambientes para a descrição semântica de vídeos*. [S.l.]: WEBMEDIA, 2007. Citado na página 3.
- SANTOS, F. S. Validação de requisitos através da prototipação de software. *CONNECTION LINE*, n. 9, 2014. Citado na página 9.
- STEPANOV, E. A. et al. Cross-language transfer of semantic annotation via targeted crowdsourcing: task design and evaluation. *Language Resources and Evaluation*, v. 52, n. 1, p. 341–364, Mar 2018. ISSN 1574-0218. Disponível em: <https://doi.org/10.1007/s10579-017-9396-5>. Citado 2 vezes nas páginas 13 e 22.
- STEPANOV, E. A.; RICCARDI, G.; BAYER, A. O. et al. The development of the multilingual luna corpus for spoken language system porting. In: *LREC*. [S.l.: s.n.], 2014. p. 2675–2678. Citado na página 22.
- SULSER, F.; GIANGRECO, I.; SCHULDT, H. Crowd-based semantic event detection and video annotation for sports videos. In: *ACM. Proceedings of the 2014 International ACM Workshop on Crowdsourcing for Multimedia*. [S.l.], 2014. p. 63–68. Citado na página 29.
- SUROWIECKI, J. *The wisdom of crowds*. [S.l.]: Anchor, 2005. Citado 2 vezes nas páginas 11 e 21.
- TEEGAVARAPU, S.; SUMMERS, J. D. Case study method for design research. *Proceedings of IDETC/DTM*, p. 3–6, 2008. Citado na página 9.
- VONDRICK CARL, P. D. R. D. Efficiently scaling up crowdsourced video annotation. *Int. J. Comput. Vision*, Kluwer Academic Publishers, Hingham, MA, USA, v. 101, n. 1, p. 184–204, jan. 2013. ISSN 0920-5691. Disponível em: <http://dx.doi.org/10.1007/s11263-012-0564-1>. Citado na página 29.
- WANG, M.; HUA, X.-S. Active learning in multimedia annotation and retrieval: A survey. *ACM Trans. Intell. Syst. Technol.*, ACM, New York, NY, USA, v. 2, n. 2, p. 10:1–10:21, fev. 2011. ISSN 2157-6904. Disponível em: <http://doi.acm.org/10.1145/1899412.1899414>. Citado na página 1.

WANG, M.; HUA, X.-S.; TANG, J.; HONG, R. Beyond distance measurement: Constructing neighborhood similarity for video annotation. *Trans. Multi.*, IEEE Press, Piscataway, NJ, USA, v. 11, n. 3, p. 465–476, abr. 2009. ISSN 1520-9210. Disponível em: <http://dx.doi.org/10.1109/TMM.2009.2012919>. Citado na página 1.

WU, B.; ZHONG, E.; TAN, B.; HORNER, A.; YANG, Q. Crowdsourced time-sync video tagging using temporal and personalized topic modeling. In: ACM. *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. [S.l.], 2014. p. 721–730. Citado na página 29.

YU, H.; SHEN, Z.; MIAO, C.; AN, B. Challenges and opportunities for trust management in crowdsourcing. In: IEEE. *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2012 IEEE/WIC/ACM International Conferences on*. [S.l.], 2012. v. 2, p. 486–493. Citado na página 21.

ZHAO, M.; HOEK, A. van der. A brief perspective on microtask crowdsourcing workflows for interface design. In: IEEE PRESS. *Proceedings of the Second International Workshop on CrowdSourcing in Software Engineering*. [S.l.], 2015. p. 45–46. Citado na página 5.

ZHU, T.; BEHAR, J.; PAPASTYLIANOU, T.; CLIFFORD, G. D. Crowlabel: A crowdsourcing platform for electrophysiology. In: IEEE. *Computing in Cardiology 2014*. [S.l.], 2014. p. 789–792. ISBN 978-1-4799-4347-0. Citado na página 69.