



UNIVERSIDADE FEDERAL DO AMAZONAS - UFAM
INSTITUTO DE COMPUTAÇÃO- ICOMP
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA - PPGI

AnnotationUI: Padrões de Interface para Sistemas de Rotulagem de Texto

Letícia Carvalho Passos

Manaus - AM
Julho de 2023

Letícia Carvalho Passos

AnnotationUI: Padrões de Interface para Sistemas de Rotulagem de Texto

Dissertação submetida à avaliação para a obtenção do título de Mestre em Informática no Programa de Pós-Graduação em Informática pelo Instituto de Computação da Universidade Federal do Amazonas (UFAM).

Orientadora: Tayana Uchoa Conte, Dra.

Coorientador: Edson César de Oliveira, Dr.

Universidade Federal do Amazonas - UFAM

Instituto de Computação- IComp

Manaus - AM

Julho de 2023

Ficha Catalográfica

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

P289a Passos, Letícia Carvalho
AnnotationUI: padrões de interface para sistemas de rotulagem
de texto / Letícia Carvalho Passos . 2023
161 f.: il. color; 31 cm.

Orientadora: Tayana Uchoa Conte
Coorientador: Edson César de Oliveira
Dissertação (Mestrado em Informática) - Universidade Federal do
Amazonas.

1. padrões de design de interface. 2. interactive machine learning.
3. processamento de linguagem natural. 4. sistemas de rotulagem
de texto. I. Conte, Tayana Uchoa. II. Universidade Federal do
Amazonas III. Título



Ministério da Educação
Universidade Federal do Amazonas
Coordenação do Programa de Pós-Graduação em Informática

FOLHA DE APROVAÇÃO

"ANNOTATIONUI: PADRÕES DE INTERFACE PARA SISTEMAS DE ROTULAGEM DE TEXTO"

LETÍCIA CARVALHO PASSOS

Dissertação de Mestrado defendida e aprovada pela banca examinadora constituída pelos Professores:

Profa. Dra. Tayana Uchôa Conte - PRESIDENTE

Profa. Dra. Milene Selbach Silveira - MEMBRO EXTERNO

Profa. Dra. Marcela Sávia Picanço Pessoa - MEMBRO EXTERNO

Manaus, 27 de julho de 2023



Documento assinado eletronicamente por **Milene Selbach Silveira, Usuário Externo**, em 28/07/2023, às 14:33, conforme horário oficial de Manaus, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Tayana Uchoa Conte, Professor do Magistério Superior**, em 28/07/2023, às 14:35, conforme horário oficial de Manaus, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Marcela Sávia Picanço Pessoa Bastos, Usuário Externo**, em 07/08/2023, às 13:41, conforme horário oficial de Manaus, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufam.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1592221** e o código CRC **D7131CE3**.

Avenida General Rodrigo Octávio, 6200 - Bairro Coroado I Campus Universitário
Senador Arthur Virgílio Filho, Setor Norte - Telefone: (92) 3305-1181 / Ramal 1193
CEP 69080-900, Manaus/AM, coordenadorppgi@icomp.ufam.edu.br

Referência: Processo nº 23105.030930/2023-10

SEI nº 1592221

AGRADECIMENTOS

Agradeço primeiramente a Deus pela vida, inteligência e oportunidades por Ele concedidas durante esta nova fase acadêmica.

À minha família, em especial minha mãe, Nara Deybbie Carvalho Nazaré, por estar sempre ao meu lado incentivando e influenciando a conquistar o sucesso acadêmico e profissional. Assim como também por me oferecer a estrutura necessária para que nada me impedisse de continuar os meus estudos.

À minha orientadora, professora Tayana Conte, que me aceitou e me ajudou a seguir em uma direção correta no mestrado. Com certeza, foi por causa dela que estou produzindo este trabalho. Pretendo ter a mesma dedicação e força de vontade que a professora Tayana. Eu apenas tenho que dizer que sou muito grata por tudo que ela fez por mim! Também agradeço ao professor Marco Cristo por ter me direcionado à professora Tayana para ser minha orientadora.

Ao meu coorientador, Edson César, que ajudou muito na definição do problema e na validação da proposta neste trabalho. Obrigada por compartilhar tudo o que sabe para mim e me fazer refletir em tudo que faço.

Agradeço também pelas professoras Milene Silveira e Simone Barbosa pelas valiosas contribuições na minha banca de qualificação, como também os professores suplentes Bruno Gadelha e Adriana Lopes. Além disso, agradeço novamente a participação da professora Milene na minha banca de defesa de dissertação juntamente com a querida professora Marcela Pessoa, como também os professores suplentes Bruno Gadelha e Walter Nakamura.

Ao grupo de pesquisa USES que tem os melhores pesquisadores que eu conheci

na minha vida. Fico muito feliz de que cada um contribuiu e compartilhou seus conhecimentos e experiências para que este trabalho melhorasse. O USES foi minha fonte de motivação para eu continuar e fazer tudo bem feito.

Em especial, agradeço pelo Lucas Viana, que é meu braço direito neste trabalho e faz com que tudo seja mais leve. Obrigada por ser um exemplo de perseverança e por me cobrar para eu fazer tudo bem feito.

À empresa *Balsamiq* por conceder uma licença que permitiu padronizar a ferramenta de prototipação para a execução de um estudo para a minha pesquisa na turma de Interação Humano-Computador. Continuarei a usar a ferramenta em outros projetos que necessitem de prototipação.

À *Amazon Prime Video* por dar oportunidade de aplicar conhecimentos adquiridos durante o mestrado em um produto de uso global. Com essa experiência pude compreender cada vez mais a importância da IHC na perspectiva de quem desenvolve.

À Universidade Federal do Amazonas, seus servidores e prestadores de serviço que forneceram condições para continuarmos nossa pesquisa mesmo de forma remota. Também ao PPGI/ICOMP por sempre responderem minhas dúvidas e nos direcionarem nesses tempos de pandemia.

Agradecemos também à FAPEAM pelo apoio financeiro por meio do POSGRAD e dos processos de número 062.00638/2019 e 062.00150/2020, CNPq por meio do processo de número 314174/2020-6, e da CAPES - Código Financeiro 001. Esta pesquisa, de acordo com o artigo 48 do Decreto nº 6.008/2006, foi parcialmente financiada pela Samsung Electronics da Amazônia Ltda, nos termos da Lei Federal nº 8.387/1991, por meio do convênio nº 003/2019, firmado com o ICOMP/UFAM e mediante o contrato 001/2020, firmado com a UFAM e a FAEPI.

Resumo

Devido ao grande volume de dados produzido por diversas aplicações, o *Machine Learning* (ML) tem sido explorado para o uso em sistemas de diferentes domínios do conhecimento humano. Como, por exemplo, sistemas que utilizam grande volume de dados textuais, como *ChatGPT*, podem necessitar de um passo anterior de treinamento conhecido como rotulagem de dados. Comumente, a rotulagem é realizada por usuários especialistas no domínio dos dados e visa gerar uma base de treinamento para um modelo ML supervisionado. Contudo, a tarefa de rotular é muito cansativa para os usuários. Por isso, as interfaces dos sistemas devem ser projetadas considerando aspectos de User Experience (UX). Algumas pesquisas focam em apoiar o desenvolvimento dos sistemas de rotulagem de dados, porém ainda existe uma lacuna sobre como projetar tais interfaces, considerando o tipo de dado a ser rotulado. Neste sentido, esta pesquisa foca em sistemas de rotulagem de texto, atendendo os problemas de *Natural Language Processing* (NLP), a qual é área de ML que atua com dados textuais. A partir do contexto mencionado, a seguinte questão de pesquisa (QP) guia este trabalho: **Como apoiar o projeto de interface de sistemas de rotulagem de texto?** Para responder esta questão, esta pesquisa foi conduzida tendo por base *Action Research*. A partir de uma demanda da indústria, foram conduzidos estudos e análises do estado-da-prática que resultou na produção de um artefato final: os padrões *AnnotationUI*. Os resultados obtidos, após avaliação com especialistas, mostram que os padrões *AnnotationUI* apoiam o projeto de interface dos sistemas de rotulagem de dados do tipo textual.

Palavras-chave: padrões de design de interface, *interactive machine learning*, processamento de linguagem natural, sistemas de rotulagem de texto.

Abstract

Due to the large volume of data coming from various applications, *Machine Learning* (ML) has been explored for use in systems from different human knowledge domains. For example, systems that use large volume of textual data, such as *ChatGPT*, may require a previous training step known as data labeling. Commonly, users who are experts in the data domain annotate data to generate a training basis for a supervised ML model. However, the task of labeling is exhausting for users. Therefore, software designers must design system interfaces considering User Experience (UX) aspects. Some research focuses on supporting the development of data labeling systems, however there is still a gap on how to design such interfaces considering the type of data to be labeled. In this sense, this research focuses on text labeling systems, addressing the problems of *Natural Language Processing* (NLP), which is an ML area that addresses textual data. From the context we mentioned above, the following research question (QP) guides this work: **How to support the interface design of text labeling systems?** To answer this question, this research was conducted based on *Action Research*. From an industry demand, we conducted empirical studies and a state-of-the-practice analysis, which resulted in the production of the final artifact: the *AnnotationUI* patterns. After evaluation with experts, the results show that the *AnnotationUI* patterns support the interface design of textual type data labeling systems.

Keywords: interface design patterns, interactive machine learning, natural language processing, annotation systems.

LISTA DE FIGURAS

Figura 1.1 – Representação da metodologia deste trabalho.	24
Figura 2.1 – Exemplos do padrão <i>captcha</i> do site <i>UI-Patterns</i>	32
Figura 2.2 – <i>Pipeline</i> de desenvolvimento de modelos de ML proposto por Nascimento et al. (2019).	34
Figura 2.3 – Exemplo de sistema de rotulagem de dados (CHOI et al., 2019). . . .	35
Figura 2.4 – Exemplo de um sistema de rotulagem com problema de anotação de entidades (NAKAYAMA et al., 2018). A interface mostra uma instância de texto e exige que o usuário identifique as entidades relacionadas às categorias <i>Person, Location, Organization, Event, Date</i> e <i>Other</i>	38
Figura 3.1 – Exemplo de um produto classificado incorretamente.	41
Figura 3.2 – Elementos de interface de sistemas de rotulagem encontrados na literatura.	44
Figura 3.3 – Soluções utilizadas no segundo protótipo.	46
Figura 3.4 – Protótipo inicial.	46
Figura 3.5 – Gráfico de frequência de termos representativos.	48
Figura 3.6 – Soluções utilizadas no segundo protótipo.	48
Figura 3.7 – Segundo protótipo proposto.	49
Figura 3.8 – Soluções utilizadas no terceiro protótipo.	51
Figura 3.9 – Terceiro protótipo proposto.	52
Figura 4.1 – Guideline 1 - Identificação da tarefa de rotulagem. Exemplos de alternativas de <i>design</i> para as tarefas de rotulagem.	58

Figura 4.2 – Soluções de contextualização do dado.	59
Figura 4.3 – Exemplos de rotulagem por instância e por classe.	60
Figura 4.4 – Sugestões da <i>guideline</i> G4 de Exibir o <i>status</i> de andamento da rotulagem.	61
Figura 4.5 – Exemplos de como disponibilizar ajudas e atalhos.	61
Figura 4.6 – Exemplos para permitir o descanso durante a rotulagem.	62
Figura 4.7 – Protótipo construído pelo participante P1.	70
Figura 4.8 – Protótipo construído pela participante P2.	71
Figura 4.9 – Protótipo construído pela participante P3.	71
Figura 4.10–Análise da conformidade das <i>Guidelines</i> às heurísticas de Nielsen. . .	79
Figura 4.11–Protótipo da equipe 1.	84
Figura 4.12–Protótipo da equipe 2.	84
Figura 4.13–Protótipo da equipe 3.	85
Figura 4.14–Protótipo da equipe 5.	86
Figura 4.15–Protótipo da equipe 4.	87
Figura 4.16–Protótipo da equipe 9.	87
Figura 4.17–Protótipo da equipe 7.	88
Figura 4.18–Protótipo da equipe 8.	89
Figura 4.19–Protótipo da equipe 6.	89
Figura 4.20–Percepções de facilidade e utilidade consideradas pelos participantes sobre a <i>guideline</i> 1.	90
Figura 4.21–Percepções de facilidade e utilidade consideradas pelos participantes sobre a <i>guideline</i> 2.	91
Figura 4.22–Percepções de facilidade e utilidade consideradas pelos participantes sobre a <i>guideline</i> 3.	92
Figura 4.23–Percepções de facilidade e utilidade consideradas pelos participantes sobre a <i>guideline</i> 4.	93
Figura 4.24–Percepções de facilidade e utilidade consideradas pelos participantes sobre a <i>guideline</i> 5.	94

Figura 4.25–Percepções de facilidade e utilidade consideradas pelos participantes sobre a <i>guideline</i> 6.	95
Figura 4.26–Distribuição de <i>Guidelines</i> que foram consideradas fáceis de entender pelos participantes.	96
Figura 4.27–Distribuição de <i>Guidelines</i> que foram consideradas úteis pelos participantes.	97
Figura 4.28–Distribuição de <i>Guidelines</i> que poderiam ser adotadas em cenários futuros pelos participantes.	97
Figura 5.1 – Processo de seleção dos sistemas de rotulagem.	107
Figura 5.2 – Tendência de desenvolvimento de sistemas de rotulagem.	108
Figura 5.3 – A figura apresenta a quantidade de estrelas e bifurcações de cada repositório.	108
Figura 5.4 – A figura apresenta a cobertura dos sistemas de rotulagem para cada problema de NLP.	109
Figura 5.5 – Brat - sistema de rotulagem. A tarefa de rotulagem do usuário é a <i>relation extraction</i> . O usuário deve relacionar partes do texto entre si. A interação ocorre com o usuário clicando em uma das partes do texto e arrastando para a outra parte.	114
Figura 5.6 – Interfaces de sistemas de rotulagem de dados.	114
Figura A.1–Exemplo de <i>Padrão Text or Document Labeling for a Category</i> . A interface mostra uma instância de texto com as palavras mais essenciais ou mais frequentes destacadas. Há uma lista de botões para solicitar <i>feedback</i> sobre a categoria de texto. Além disso, há um botão "not sure" para indicar dúvidas sobre a rotulagem de dados.	144
Figura A.2–Exemplo de padrão <i>Text or Document Labeling for Multi-category</i> . A interface mostra a instância de texto e solicita que o usuário selecione as categorias relacionadas a ela. A interface exibe as classes em um componente <i>checkbox</i> onde o usuário pode escolher mais de um rótulo. Além disso, o usuário pode selecionar a opção "other" e inserir uma nova categoria.	145

Figura A.3–Exemplo de padrão <i>Entities Labeling</i> . A interface mostra a instância de texto e os rótulos por meio de botões. O usuário clica em um botão que representa um rótulo e seleciona a palavra relacionada. A palavra anotada tem uma cor de fundo que corresponde à mesma cor do rótulo correspondente.	147
Figura A.4–Exemplo do padrão <i>Relation Extraction Task Labeling</i> . A interface mostra uma instância de texto com algumas palavras rotuladas. O usuário deve associar duas palavras rotuladas clicando em ambas as palavras e selecionando a relação correta. A interface apresenta a nova relação por meio de uma seta com o rótulo.	148
Figura A.5–Exemplo do padrão <i>Batch Labeling</i> . A interface mostra um grupo de instâncias rotuladas e solicita que o usuário selecione as corretas. No lado direito, uma tabela apresenta as palavras mais frequentes dos casos.	149
Figura A.6–Exemplos de padrão de ajuda.	149
Figura A.7–Exemplo de Ajuda Externa. A interface fornece um link para a web para pesquisar informações sobre os dados	150
Figura A.8–Exemplo do padrão <i>Word Search</i> . A interface fornece uma <i>search bar</i> , destacando as palavras encontradas e dando o <i>feedback</i> de quantas palavras a pesquisa encontrou.	151
Figura A.9–Exemplo do padrão <i>Undo</i> . A interface fornece um botão de desfazer e mostra um ícone de exclusão ao lado das palavras anotadas.	152
Figura A.10–Exemplos do padrão <i>Navigation Between Instances</i>	152
Figura A.11–Exemplos do padrão <i>Add Category</i>	153
Figura A.12–Exemplo do padrão <i>Feedback of Lack of Data Domain</i> . A interface fornece um botão <i>not sure</i>	154
Figura A.13–Exemplos do padrão <i>Save</i>	155
Figura A.14–Exemplo do padrão <i>Cancel</i>	155
Figura A.15–Exemplos do padrão <i>Shortcuts</i>	156
Figura A.16–Exemplos do padrão <i>Tool Status</i>	157

Figura A.17-Exemplos do padrão <i>Labeling Task Progress Status</i>	158
Figura A.18-Exemplo do padrão <i>Labeling History Overview</i> . O componente de interface fornece o histórico de atividades do usuário que está rotulando.	159
Figura A.19-Exemplos do padrão <i>Statistics</i>	160

LISTA DE TABELAS

Tabela 3.1 – Requisitos principais iniciais do sistema de rotulagem.	43
Tabela 3.2 – Requisitos Refinados do Sistema de Rotulagem a partir dos requisitos iniciais e dos pontos fortes dos protótipos.	53
Tabela 4.1 – <i>Guidelines LabelUX</i>	57
Tabela 4.2 – Tipos de tarefas de ML por tipo de dado (TKACHENKO et al., 2020-2021), (AWS, 2021).	58
Tabela 4.3 – Problema de ML definidos pelas equipes.	83
Tabela 4.4 – Total de telas produzidas e <i>guidelines</i> utilizadas por cada equipe de cada perfil.	83
Tabela 5.1 – Sub-questões de pesquisa.	100
Tabela 5.2 – Consultas de pesquisa usando apenas o operador AND.	103
Tabela 5.3 – Critérios de seleção	105
Tabela 5.4 – Componentes de interfaces utilizadas para solicitar o <i>feedback</i> de rotulagem.	113
Tabela 5.5 – Os componentes de interface usados para apresentar o progresso da rotulagem ao usuário	116
Tabela 6.1 – Categorias dos padrões <i>AnnotationUI</i> e as suas rastreabilidades. . . .	120
Tabela 6.2 – Um exemplo de padrão de tarefa de rotulagem: classificação de texto para uma multi-categoria.	121
Tabela 6.3 – Um exemplo de padrão de interface UX - User needs help: Tool Help.	123
Tabela 6.4 – Um exemplo de padrão de interface UX - User needs freedom of use: Shortcuts.	124

Tabela 6.5 – Um exemplo de padrão de interface UX - User needs information:

Labeling Task Progress 125

Tabela 6.6 – Perfil dos participantes 128

SUMÁRIO

1	INTRODUÇÃO	20
1.1	Contexto	20
1.2	Problema	21
1.3	Questão de pesquisa	23
1.4	Objetivos	23
1.5	Metodologia	24
1.6	Organização do documento	27
2	REFERENCIAL TEÓRICO	29
2.1	Padrões de design de interface	29
2.2	<i>Machine Learning</i>	32
2.2.1	Sistemas de Rotulagem de Dados	34
2.2.2	Processamento de Linguagem Natural	36
2.3	Trabalhos Relacionados	38
3	MOTIVAÇÃO: CASO SEFAZ	40
3.1	Sistema de rotulagem de NF-e's SEFAZ	40
3.2	Análise de elementos de interface na literatura	43
3.3	Soluções propostas	45
3.3.1	Protótipo Inicial	45
3.3.1.1	Avaliação e lições aprendidas	47
3.3.2	Segundo Protótipo	48
3.3.2.1	Avaliação e lições aprendidas	50
3.3.3	Terceiro Protótipo	51
3.3.3.1	Avaliação e lições aprendidas	52

3.4	Discussão	53
4	ARTEFATO INICIAL: GUIDELINES LABELUX	56
4.1	Processo de criação das <i>guidelines</i>	56
4.2	Descrições das <i>guidelines</i>	57
4.2.1	G1 - Identificação da tarefa de rotulagem	57
4.2.2	G2 - Contextualização do dado	59
4.2.3	G3 - Instância vs Classe	60
4.2.4	G4 - Exibir o <i>status</i> de andamento da rotulagem	61
4.2.5	G5 - Disponibilizar ajuda e atalhos	61
4.2.6	G6 - Descanso durante a rotulagem	62
4.3	Rastreabilidade das <i>guidelines</i>	62
4.4	Estudo de viabilidade	63
4.4.1	Resultados e discussão	65
4.4.1.1	Entendimento do problema de ML dos participantes	65
4.4.1.2	Introdução das <i>guidelines LabelUX</i>	65
4.4.2	Coleta de impressões sobre cada <i>guideline</i>	66
4.4.2.1	Elaboração de protótipo	70
4.4.2.2	Perguntas finais	72
4.5	Estudo de avaliação de apoio à Usabilidade	74
4.5.1	G1 - Identificação da tarefa de rotulagem	74
4.5.2	G2 - Contextualização do dado	75
4.5.3	G3 - Rotulagem por instância ou por classe	76
4.5.4	G4 - Exibir o <i>status</i> do progresso da rotulagem	77
4.5.5	G5 - Disponibilizar ajuda e atalhos	77
4.5.6	G6 - Descanso durante a rotulagem	78
4.5.7	Considerações finais	78
4.6	Estudo <i>In Vitro</i>	81
4.6.1	Construção dos protótipos	82
4.6.2	Protótipos de perfil A	83
4.6.3	Protótipos de perfil B	86

4.6.4	Protótipos de perfil C	88
4.6.5	Percepção de uso das <i>guidelines</i>	90
4.6.6	Considerações finais	98
5	ANÁLISE DOS SISTEMAS DE ROTULAGEM DE TEXTO DO GITHUB	99
5.1	Definindo a análise	99
5.2	Estratégia de busca	101
5.2.1	Busca no <i>GitHub</i>	101
5.2.1.1	Problemas na busca por <i>string</i>	102
5.2.2	Processo de Seleção dos sistemas de rotulagem	104
5.2.3	Estratégia de extração de dados	104
5.2.4	Resultados	106
5.2.4.1	Repositórios Seleccionados	107
5.2.4.2	Respondendo à questão de pesquisa	110
6	ARTEFATO FINAL: PADRÕES ANNOTATIONUI	118
6.1	<i>AnnotationUI</i>	118
6.2	Avaliação dos padrões <i>AnnotationUI</i>	124
6.2.1	Perfil dos participantes	127
6.2.2	Resultados e Discussão	128
6.3	Ameaças à Validade	131
7	CONSIDERAÇÕES FINAIS	133
7.1	Contribuições	134
7.2	Trabalhos Futuros	134
	Referências	136
	APÊNDICE A PADRÕES ANNOTATIONUI	143
A.1	Padrões <i>AnnotationUI</i>	143
A.1.1	Padrão Text or Document Labeling for a Category	143

A.1.2	Padrão Text or Document Labeling for Multi-category	144
A.1.3	Padrão Entity Labeling Task	146
A.1.4	Padrão Relation Extraction Task	147
A.1.5	Padrão Batch Labeling	148
A.1.6	Padrão Tool Help	149
A.1.7	Padrão External Help	150
A.1.8	Padrão Word Search	150
A.1.9	Padrão Undo a labeling task	151
A.1.10	Padrão Navigation Between Instances	152
A.1.11	Padrão Add Category	153
A.1.12	Padrão Feedback of Lack of Data Domain	153
A.1.13	Padrão Save	154
A.1.14	Padrão Cancel	155
A.1.15	Padrão Shortcuts	155
A.1.16	Padrão Tool status	156
A.1.17	Padrão Labeling Task Progress	157
A.1.18	Padrão Labeling History Overview	158
A.1.19	Padrão Statistics	159

1

INTRODUÇÃO

Neste capítulo são introduzidos o problema e a pesquisa de mestrado. Esta pesquisa foi feita em colaboração com o doutorando Lucas Viana, contendo então material utilizado em sua qualificação de doutorado.

A Seção 1.1 destaca o contexto e a motivação do trabalho desenvolvido. As Seções 1.2, 1.3 e 1.4 apresentam o problema tratado, a questão de pesquisa e os objetivos a serem alcançados, respectivamente. A metodologia de pesquisa adotada é exposta na Seção 1.5. Por fim, na Seção 1.6 é apresentada a estrutura de organização deste texto.

1.1 Contexto

É crescente o volume de dados gerados pelos diversos recursos tecnológicos disponíveis devido o avanço da computação nas últimas décadas (CAI; ZHU, 2015). Esse grande volume de dados provêm de sistemas bancários, redes sociais e até de equipamentos de análise do genoma humano (MARX, 2013). Este fenômeno, chamado *Big Data*, permitiu que muitos dados fossem utilizados para apoiar outros fins (CAI; ZHU, 2015).

Uma dessas finalidades está no uso dos dados para treinar modelos de *Machine Learning* para aprender a realizar classificações automáticas (ZHOU et al., 2017). Assim, seria possível diminuir ou até mesmo anular a presença humana em tarefas repetitivas, ou complexas como, por exemplo, a detecção automática de fraude em assinaturas sem a verificação de um especialista (GHANIM; NABIL, 2018) ou até mesmo produção textual com *ChatGPT* (DIS et al., 2023).

Entretanto, apesar da visão futurística de que as “máquinas” ocupariam o lugar dos humanos em diversas tarefas, a realidade é que os modelos de ML necessitam de controle humano, não somente nas etapas de desenvolvimento, mas também no monitoramento desses modelos quando são lançados em produção (YUAN et al., 2020). Se o modelo de ML depender de dados previamente categorizados para aprender, então este modelo é chamado supervisionado. Se estes dados dependerem de pessoas para realizar essa categorização, então há necessidade de incluir a etapa de rotulagem desses dados (AMERSHI et al., 2014).

A etapa de rotulagem ocorre através de sistemas conhecidos como sistemas de rotulagem de dados. Cada sistema se difere conforme o tipo de dado a ser rotulado, podendo ser texto, imagem, áudio e vídeo (AWS, 2021). Assim, dada a necessidade de construir um sistema de rotulagem de dados, é necessário também projetar uma boa interface para o dado a ser rotulado pelos usuários (AMERSHI et al., 2014).

Nesta dissertação, foram considerados sistemas de rotulagem de texto devido a sua complexidade na área de ML. Neste sentido, o *Natural Language Processing* (NLP) é a área de ML que engloba os problemas de dados textuais, tais como análise de sentimento e reconhecimento de entidades nomeadas (OTTER; MEDINA; KALITA, 2020).

1.2 Problema

Na etapa de rotulagem de dados, existe a necessidade de interação do usuário com os dados. Assim, muitas pesquisas vêm sendo realizadas para delimitar o desenvolvimento de sistemas para intermediar a rotulagem de dados (DUDLEY; KRISTENSSON, 2018a; VICTORELLI; REIS, 2020; AMERSHI et al., 2019).

Alguns estudos, como os de Nadj et al. (2020) e Amershi et al. (2019), investigam os desafios de interação desses sistemas. Entre os desafios, são investigados os efeitos desses sistemas nos usuários, pois durante a tarefa de rotulagem, o usuário pode não estar interessado em ser um mero oráculo do sistema (KNAEBLE; NADJ; MAEDCHE, 2020). Além disso, a tarefa de rotulagem pode se tornar cansativa, gerando assim

sentimento de frustração e chateação no usuário (VIANA; OLIVEIRA; CONTE, 2021).

A atividade de rotulagem de dados realizada pelo usuário pode se tornar uma atividade desmotivadora, deixando o usuário com um sentimento de tédio, a ponto dele desistir do sistema (HUANG et al., 2013). Como consequência deste problema, a produtividade do usuário cai, podendo afetar a qualidade da rotulagem. A baixa qualidade da rotulagem faz com que o modelo de ML aprenda com dados rotulados incorretamente o deixando não condizente com o mundo real e assim não usável (ALPAYDIN, 2020).

Esses desafios com a experiência dos rotuladores nos leva ao problema relacionado à qualidade de projeto dos sistemas de rotulagem de dados (GONZALES; HORITA, 2020). Escolhas ruins para o *design* de interface também contribuem para este problema. Uma interface mal projetada afeta a experiência do usuário, causando desinteresse por parte de quem for utilizar o sistema. As interfaces desses sistemas devem ser projetadas visando auxiliar o usuário na análise e compreensão dos dados, como também permitir que o usuário tenha *feedback* sobre o andamento da atividade.

Para apoiar o *design* dos sistemas de rotulagem, diversas abordagens vêm sendo tema de publicações na literatura (HEBBALAGUPPE et al., 2013; DUDLEY; KRISTENSON, 2018b; NADJ et al., 2020; AMERSHI et al., 2014). Porém, estas focam em aspectos genéricos como transparência dos dados e visualização de desempenho de modelos, deixando uma lacuna com relação a como projetar as interfaces de sistemas de rotulagem de dados para apoiar a tarefa de rotulagem, visando uma experiência positiva do usuário.

Além disso, os trabalhos encontrados na literatura não consideram o tipo de dado a ser rotulado, porém, segundo Munzner (2014), muitos aspectos de *design* visual são orientados pelo tipo de dado apresentado na interface. Além disso, para cada tipo de dado temos diversos problemas de ML associados a cada um deles. Para cada tipo de problema, há muitas formas de interagir com o dado e obter o feedback dos rotuladores (NADJ et al., 2020; AMERSHI et al., 2019). Por isso, neste trabalho o foco foi dado aos problemas de NLP que são problemas de ML de dados textuais.

1.3 Questão de pesquisa

A partir da descrição do contexto e do problema a ser tratado, a seguinte Questão de Pesquisa (QP) foi levantada:

QP: Como apoiar o projeto de interface de sistema de rotulagem de texto?

Ao responder esta QP, busca-se apoiar a solução do problema levantado nesta dissertação. A solução consiste na disponibilização de padrões de interface para projetos de interface de sistemas de rotulagem de texto. Estes padrões foram desenvolvidos a partir de lições aprendidas de estudos empíricos e de observações em sistemas de rotulagem de dados da indústria e da literatura. Os estudos realizados confirmam que os padrões respondem à QP desta dissertação.

1.4 Objetivos

O objetivo desta dissertação é desenvolver um artefato para apoiar no projeto de interfaces de sistemas de rotulagem de texto. Este projeto deve levar em consideração o tipo de problema de NLP, visando a experiência do usuário durante a rotulagem.

Objetivos Específicos

- Entender as dificuldades de projeto de sistemas de rotulagem de dados em um caso na indústria;
- Estabelecer um referencial teórico relevante sobre sistemas de rotulagem;
- Analisar as interfaces de sistemas de rotulagem da indústria;
- Desenvolver e disponibilizar um material para apoiar o projeto de interface de sistemas de rotulagem de texto;
- Avaliar o material produzido quanto sua clareza e utilidade;
- Refinar o material produzido a partir de estudo com especialistas.

1.5 Metodologia

A metodologia adotada nesta pesquisa é do tipo pesquisa-ação. Esse tipo de pesquisa concentra-se no processo evolutivo de um determinado aspecto a partir de um problema da indústria, no qual o pesquisador participa ativamente das melhorias implementadas durante esse processo. Além disso, a cada pesquisa, o pesquisador conduz uma investigação, reflete sobre os resultados e propõe melhorias. (WOHLIN et al., 2012). Todos os estudos realizados foram conduzidos sob o Comitê de Ética da Universidade Federal do Amazonas de número 51490121.0.0000.5020.

A motivação inicial desta pesquisa foi atender a uma necessidade específica da indústria, que consistia no projeto de interface para um sistema de rotulagem de texto da Secretaria do Estado da Fazenda do Estado do Amazonas (SEFAZ/AM). Para abordar essa necessidade, foram realizadas atividades de pesquisa a fim de compreender o problema. A partir das reflexões obtidas, foram desenvolvidas diretrizes (*guidelines*) para orientar os desenvolvedores na criação de interfaces para sistemas de rotulagem.

Posteriormente, percebeu-se a importância de aprimorar as *guidelines* com exemplos práticos de uso em outros sistemas de rotulagem. Portanto, foi necessário investigar os sistemas de rotulagem utilizados na prática, para o qual foram examinados os sistemas de rotulagem disponíveis nos repositórios do *GitHub*. Os resultados dessa investigação foram fundamentais para aprimorar as *guidelines*, que evoluíram para se tornarem padrões de interface para sistemas de rotulagem de texto.

A metodologia deste trabalho é apresentada na Figura 1.1. Cada passo desta metodologia está descrito logo a seguir.

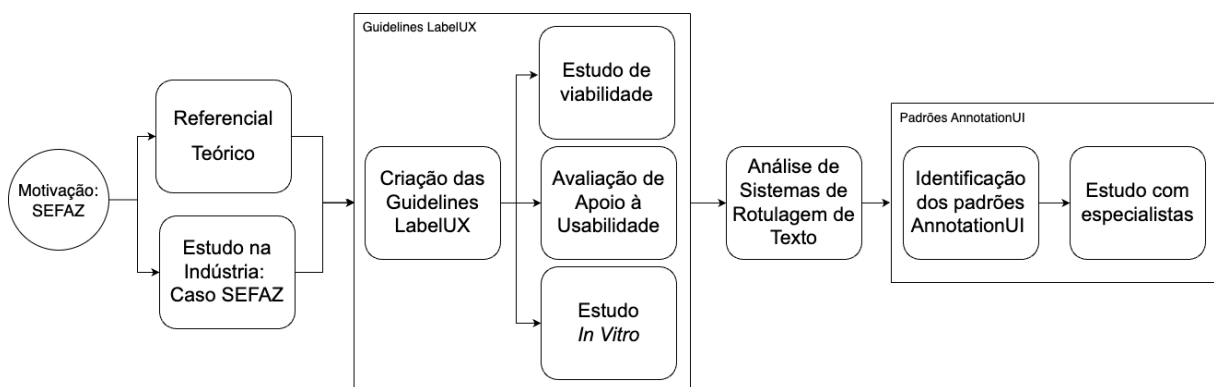


Figura 1.1 – Representação da metodologia deste trabalho.

1. **Referencial Teórico** - Foi utilizado como base o estudo realizado por (VIANA; OLIVEIRA; CONTE, 2021) para encontrar as soluções de interface utilizadas em sistemas de rotulagens de dados. Neste estudo, que foi uma extensão de uma RSL, foram analisados 27 artigos sobre sistemas de *Interactive Machine Learning*. Este estudo analisa os componentes de interfaces de sistemas de rotulagem e como os usuários interagem com eles. Além disso, os resultados desse trabalho foram utilizados como base teórica para a criação de algumas *guidelines*.
2. **Estudo na indústria** - Estudo de caso na SEFAZ/AM, em que a equipe de desenvolvimento tinha necessidade de projetar um sistema de rotulagem de descrições de produtos de notas fiscais. Os membros desta equipe tinham experiência na construção de sistemas WEB, porém não tinham conhecimentos profundos sobre sistemas de rotulagem de dados. No estudo, foi apurado que a equipe sentia dificuldades na prototipação de interfaces para sistemas de rotulagem, pois os desenvolvedores não tinham conhecimento sobre os requisitos de sistemas de rotulagem de dados. Como resultado, foi obtido um processo de refinamento de requisitos para sistemas de rotulagem por meio de prototipação evolutiva da interface. Além disso, as lições aprendidas desse estudo foram base para a criação das *guidelines*. Os resultados desta etapa foram submetidos no 24th *Workshop on Requirements Engineering* (WER 2021) (PASSOS et al., 2021a).
3. **Guidelines LabelUX** - A partir das conclusões obtidas dos passos anteriores foram criadas seis *Guidelines*, nomeadas de *LabelUX*. As *guidelines* visam ajudar os desenvolvedores a projetar as interfaces dos sistemas de rotulagem. Após a criação foi necessário avaliar o uso das *Guidelines LabelUX*.
 - a) **Criação das *guidelines*** - A elaboração da primeira versão das *guidelines* foi feita com base nos conhecimentos obtidos nas etapas anteriores, seguida por versões geradas a partir um ciclo de melhoria das *guidelines* com especialistas em IHC e ML. Esta versão das *guidelines* engloba os problemas de ML para o tipo de dado textual.
 - b) **Estudo de viabilidade** - Estudo para avaliar a viabilidade das *guidelines*

realizado com três engenheiros de *software*. Os participantes desse estudo faziam parte de projetos de desenvolvimento que utilizavam técnicas de ML supervisionadas e que tinham a necessidade de desenvolver um sistema de rotulagem de dados. As *guidelines* foram apresentadas aos participantes que, em seguida, criaram protótipos de interface de sistemas de rotulagem de dados. Por fim, os desenvolvedores puderam fornecer suas impressões sobre as *guidelines*. Durante a avaliação, participantes verificaram que com o uso das *guidelines* era possível projetar a interface de um sistema de rotulagem que os atendessem. Este estudo foi descrito em um artigo publicado no Simpósio Brasileiro de Qualidade de Software (SBQS) 2021 ([PASSOS et al., 2021b](#)).

- c) **Avaliação de apoio à Usabilidade** - Análise dos resultados obtidos através das etapas anteriores para avaliar se foi possível atender a aspectos relacionados à usabilidade com o uso das *guidelines*. Para isso, cada *guideline* foi analisada, visando encontrar relações sobre o uso destas e as recomendações de usabilidade, através das heurísticas de [Nielsen \(1994\)](#). Os resultados desta etapa foram publicados no XX Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais (IHC 2021) ([VIANA et al., 2021](#)).
 - d) **Estudo *In Vitro*** - Estudo realizado em uma turma de Interação Homem-Computador para avaliar a facilidade, utilidade e pretensão de uso das *guidelines*. A turma foi dividida em equipes e todas elas tiveram que construir protótipos de sistemas de rotulagem de tweets. Foram comparados se os protótipos produzidos pelas equipes que não possuíam experiência em ML tinham a mesma qualidade dos protótipos produzidos pelas equipes que tinham experiência em ML e rotulagem de dados.
4. **Análise de sistemas de rotulagem de texto do *GitHub*** - A partir das conclusões das avaliações das *guidelines* e na tentativa de aprimorar as *guidelines* foi conduzido um novo estudo focando em analisar o estado-da-prática dos sistemas de rotulagem de dados textuais. A busca desses sistemas foi restrita aos repositórios do *GitHub*, visto que é a plataforma mais popular de sistemas *open-source* ([CO-](#)

SENTINO; IZQUIERDO; CABOT, 2017), em que desenvolvedores armazenam seus projetos e deixam disponíveis ao público. Neste ponto, as *guidelines* foram utilizadas para analisar esses sistemas de rotulagem. A análise teve como objetivo descobrir as características e as limitações dessas interfaces de sistemas.

5. **Padrões AnnotationUI** - Com os resultados da análise, foi possível construir padrões de interface dado o tipo de problema textual. Assim, os padrões *AnnotationUI* se tornaram uma evolução das *guidelines LabelUX*. Em seguida, foi necessário avaliar o uso dos padrões *AnnotationUI*.

- a) **Identificação dos padrões** — A análise dos sistemas encontrados no *GitHub* permitiu obter dados suficientes para construir os padrões conforme o tipo de problema textual a ser tratado. Assim, os padrões são categorizados por tipo de problema de NLP a ser resolvido.
- b) **Avaliação dos padrões AnnotationUI com profissionais da indústria** — Para avaliar o uso dos padrões foi conduzido um estudo com quatro profissionais da indústria que trabalham com projeto de ML. Os profissionais utilizaram os padrões para criação de interfaces de sistemas de rotulagem do tipo textual. Os resultados indicaram que o uso dos padrões *AnnotationUI* ajuda o desenvolvedor a projetar as interfaces mais rapidamente e a entender as necessidades do usuário.

1.6 Organização do documento

Esta dissertação contém outros seis capítulos, além deste capítulo inicial que visa contextualizar o problema tratado neste trabalho. A seguir, é apresentada a estrutura de organização deste trabalho:

- **Capítulo 2: REFERENCIAL TEÓRICO** - Apresenta o referencial teórico sobre os conceitos Padrões de *design* de interface, aprendizado de máquina, aprendizado de máquina interativo, sistemas de rotulagem de dados, processamento de linguagem natural e os trabalhos relacionados ao contexto desta pesquisa.

- **Capítulo 3: MOTIVAÇÃO: CASO SEFAZ** - Descreve o estudo de caso realizado na SEFAZ/AM para entendimento do problema sobre projeto de interface de sistemas de rotulagens de dados e o processo de propostas de soluções.
- **Capítulo 4: ARTEFATO INICIAL: GUIDELINES LABELUX** - Apresenta a proposta de *guidelines* para apoiar o projeto de interface de sistemas de rotulagem de dados. Neste capítulo contém o processo de construção das *guidelines* e estudos realizados para avaliação das mesmas.
- **Capítulo 5: ANÁLISE DOS SISTEMAS DE ROTULAGEM DE TEXTO DO GITHUB** - Descreve como foi realizada a análise dos sistemas de rotulagem de texto do *GitHub*, tendo como base as *guidelines LabelUX*.
- **Capítulo 6: ARTEFATO FINAL: PADRÕES ANNOTATIONUI** - Apresenta os padrões criados a partir da análise dos sistemas de rotulagem textuais do *GitHub*, como também o estudo realizado com especialistas para avaliação dos padrões.
- **Capítulo 7: CONSIDERAÇÕES FINAIS** - Apresenta uma conclusão dos resultados obtidos nos estudos, discorre as contribuições e limitações desta pesquisa.

2

REFERENCIAL TEÓRICO

Neste capítulo são descritos os principais conceitos que serviram de base para o desenvolvimento deste trabalho. Primeiramente, na Seção 2.1 é abordado o referencial sobre os padrões de design de interface do usuário. Na Seção 2.2 são apresentados conceitos fundamentais de *Machine Learning* e seu processo de desenvolvimento, ML Interativo, NLP e a função dos sistemas de rotulagens de dados neste processo. Por fim, na Seção 2.3, são apresentados os trabalhos relacionados que abordam o tema de projeto de sistemas de apoio ao ML e sistemas de rotulagem de dados.

2.1 Padrões de design de interface

Padrões de design de interface são uma das opções de diretrizes de *design* (COWLEY; WESSON, 2005). As diretrizes de *design* são geralmente conhecidas como princípios de *design* ou boas práticas que são fáceis de usar e entender (REED et al., 1999). No entanto, essas diretrizes podem, às vezes, ser pouco claras e imprecisas (COWLEY; WESSON, 2005). Assim, muitos pesquisadores afirmaram vantagens consideráveis de usar padrões de *design* de interface em vez de diretrizes de *design* (DEARDEN; FINLAY, 2006).

As vantagens incluem o uso de exemplos para sistemas implementados específicos, declarações de problemas, abstração de baixo nível, o contexto da aplicação e o suporte racional (DEARDEN; FINLAY, 2006). Os padrões seguem uma estrutura que se comunica rapidamente com os desenvolvedores (BORCHERS, 2008). Além disso,

podem auxiliar os desenvolvedores a melhorar a experiência do usuário durante o *design* da interface (MALIK; MUNIR; ALI, 2020).

Conforme a ISO 9241-210:2019, experiência do usuário, do inglês *User Experience (UX)*, é a “percepção e respostas do usuário resultantes do uso ou antecipação do uso de um produto, sistema ou serviço”. A UX incorpora todos os aspectos da interação com um produto, incluindo percepções e respostas, que podem ser fatores pragmáticos relacionados a objetivos eficazes e eficientes, e fatores hedônicos relacionados aos sentimentos e bem-estar dos usuários (HASSENZAHL, 2018). Um dos aspectos da interação é a usabilidade que pode avaliar alguns elementos de UX (ISO, 2019).

Segundo Nielsen e Loranger (2006), a usabilidade está relacionada ao aprendizado, memorização, prevenção de erros, eficiência e satisfação do usuário. Uma interface que não oferece boa usabilidade é considerada uma interface problemática, podendo causar prejuízos ao usuário e até mesmo a rejeição do produto ou serviço (MONZÓN; ANGELERI; DÁVILA, 2020).

O desenvolvedor sempre deve considerar os aspectos de UX e usabilidade na construção de interfaces. Algumas pesquisas são aplicadas para ajudar o desenvolvedor nessa atividade (LIU et al., 2018; LEIVA; HOTA; OULASVIRTA, 2020; CHEN et al., 2020). Além disso, existem materiais fora da literatura que também são buscados pelos desenvolvedores para apoiar a criação de interface. Um desses materiais é o *UI-Patterns*¹.

A *UI-Patterns* é referente a uma biblioteca de padrões de interface para aplicações WEB. Os padrões possuem a seguinte estrutura: problema, exemplo, uso, solução e justificativa. Os padrões são agrupados em 11 categorias. Os padrões são categorizados conforme o tipo do problema que solucionam.

Para fins de ilustração, um dos padrões do *UI-Patterns* é o *Captcha*.² Esse padrão ajuda os desenvolvedores a projetar uma interface quando é necessário proteger o usuário e o sistema de possíveis *spam* ou tentativas de *hackers* de ataques no sistema.

A descrição e as imagens apresentadas no padrão abaixo foram retiradas diretamente do site dos *UI-Patterns*³. As descrições foram traduzidas de inglês para português brasileiro.

¹ <<https://ui-patterns.com/>>

² <<https://ui-patterns.com/patterns/Captcha>>

³ <<https://ui-patterns.com/>>

Captcha

Problema. O sistema precisa verificar se os dados enviados são originários de um humano real e não de um robô.

Uso. Sites com a capacidade de os visitantes comentarem, se registrarem como usuários ou postarem conteúdo ativamente no site são frequentemente expostos a ataques de robôs de spam. O conteúdo postado por esses robôs de spam pode ser comparado ao spam enviado por e-mail. O principal objetivo dos robôs de *spam* na WEB é criar e espalhar *links* para um site específico, a fim de aumentar a classificação de pesquisa dos sites em mecanismos de pesquisa como o google.com. Essas postagens de *spam* raramente têm algo a ver com o assunto do site, daí a categorização de *spam*. Para evitar esse tipo de *spam*, são introduzidos *captcha*. A ideia por trás do *captcha* é criar uma maneira de distinguir seres humanos reais de robôs automatizados. Use quando seu aplicativo da web sofrer ataques de robôs mal-intencionados da web tentando postar conteúdo de spam em seu site. Use para proteger seu site de robôs automatizados. Use quando a capacidade de postar conteúdo em seu site não estiver bloqueada pela necessidade de estar conectado. Os processos de registro estão incluídos.

Solução. A forma mais popular de *captchas* são imagens que representam letras e números dentro. O usuário é solicitado a escrever em um campo de formulário separado o que a imagem lê em um campo de formulário separado. Para impedir que os *spammers* usem o *software OCR* para ler a imagem, a imagem é manipulada de maneiras diferentes, dificultando os computadores, mantendo a legibilidade para os seres humanos. Se o usuário conseguir digitar o que a imagem diz, seu conteúdo será publicado no site. Caso contrário, a ação será recusada. É comum permitir que várias tentativas entrem no texto do *captcha*, pois algumas imagens do *captcha* são até ilegíveis para os seres humanos devido à forte manipulação da imagem que foi exposta.

Justificativa. Os *captchas* são a abreviação de “Teste de Turing Público Completamente Automatizado para diferenciar computadores e humanos”, do inglês *Completely Automated Public Turing test to tell Computers and Humans Apart*. A ideia por trás do *captcha* é distinguir humanos de computadores, permitindo que o usuário execute uma ação que um computador não pode. Um *captcha* é um teste de *Turing* simples.

Há uma linha tênue entre tornar um *captcha* irreconhecível para *scanners* OCR e ainda legível para seres humanos. A legibilidade para o ser humano deve vir em primeiro lugar. Outros problemas com a implementação de *captchas* para proteger seu site incluem um bloqueio de usuários com deficiência visual, pois eles não podem usar o software de voz para falar o que o *captcha* lê.

Outras formas de proteção contra *spammers* maliciosos são fazer perguntas como “o que é 2 + 3” ou “o que é dois mais três” ou usar *captchas* de voz.

Exemplos

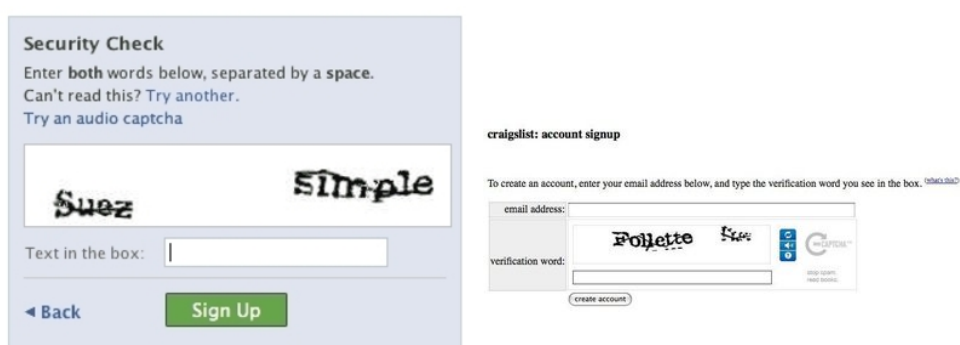


Figura 2.1 – Exemplos do padrão *captcha* do site *UI-Patterns*.

2.2 Machine Learning

Machine Learning (ML) ou Aprendizado de Máquina, é uma subárea da Inteligência Artificial que compõe algoritmos capazes de melhorar seu desempenho com a experiência. Um algoritmo neste domínio é capaz de aprender a partir de dados, capturando padrões e efetuando inferências (MITCHELL et al., 1990).

Para que ocorra o aprendizado, os algoritmos de ML passam por um processo de aquisição da experiência, chamado de treinamento. Durante essa fase, o algoritmo busca uma função, que mapeie as instâncias do espaço de entrada para o de saída. Estes modelos podem ser agrupados em diferentes categorias quanto ao tipo de aprendizado e a saída desejada para o algoritmo (FLACH, 2012).

Quanto ao tipo de aprendizado, as tarefas de ML podem ser agrupadas em três tipos diferentes, a depender da presença e do tipo de resposta dada ao algoritmo quanto ao desempenho de suas saídas. Estes são (RUSSELL; NORVIG, 2002):

1. Aprendizagem supervisionada - O algoritmo deve aprender a inferir valores a partir de atributos preditores e do respectivo atributo alvo fornecido como exemplo, ou seja, de cenários em que os valores de saída já foram previamente rotulados por especialista ou coletados na vida real.
2. Aprendizagem não-supervisionada - O algoritmo deve inferir padrões e estruturas a partir de dados não rotulados, buscando alguma estrutura interna que os caracterize.
3. Aprendizagem por reforço - O algoritmo não recebe dados nem rótulos, e deve aprender a partir das recompensas positivas ou negativas dadas por ações de simulação.

Quanto ao tipo de saída desejada, os problemas que podem ser endereçados segundo ML são de classificação, regressão, transcrição, tradução automática, detecção de anomalia, síntese e amostragem. No caso do aprendizado supervisionado, em particular, as principais tarefas realizadas são de classificação e regressão (FLACH, 2012).

A tarefa de classificação deve especificar para cada entrada x como pertencente a uma dentre k categorias pré-determinadas. O modelo que classifica deve aprender uma função f que mapeia sequências de números reais x de dimensão n para um valor inteiro y dentre k possibilidades (STONE; VELOSO, 2000).

Até o lançamento do modelo de ML, o mesmo passa por um *pipeline* de desenvolvimento, como pode ser observado na Figura 2.2. Este *pipeline* inclui as etapas de entendimento do problema, manipulação dos dados, construção do modelo e monitoramento do modelo (NASCIMENTO et al., 2019).

Um dos principais fatores para um modelo de ML ter bons resultados é possuir uma base de treinamento em grande quantidade e com qualidade nos dados (etapa de manipulação dos dados). Caso contrário, o modelo não aprenderá a inferir corretamente, mesmo sendo desenvolvido com os melhores algoritmos. No caso dos modelos supervisionados, é imprescindível que cada dado esteja corretamente categorizado para que o modelo aprenda corretamente a partir desses dados (FLACH, 2012). Em alguns casos, é necessário a presença de especialistas para categorizarem ou validarem

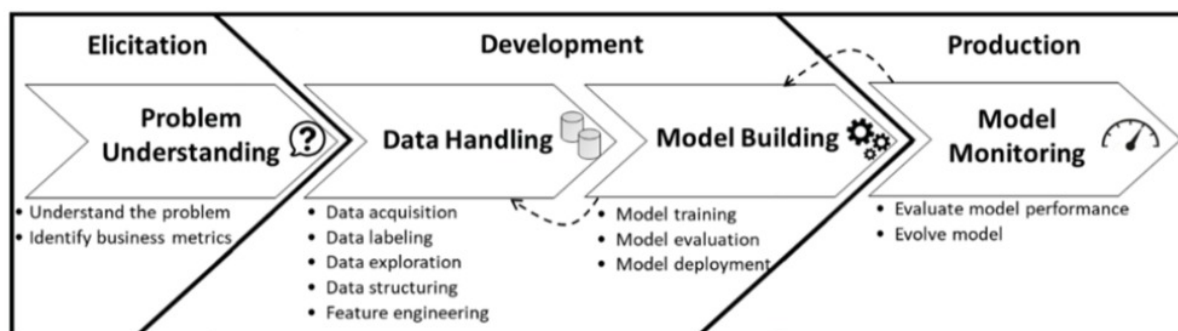


Figura 2.2 – Pipeline de desenvolvimento de modelos de ML proposto por Nascimento et al. (2019).

as classes desses dados. Este processo é chamado rotulagem de dados e encontra-se na etapa de manipulação dos dados no *pipeline* de desenvolvimento dos modelos de ML.

Esta interação humana direta com os dados e com modelos de ML é estudada por uma área denominada *Interactive Machine Learning*. Sistemas de ML interativo visam a validação da categorização ou revisão dos dados (KNAEBLE; NADJ; MAEDCHE, 2020). Esses sistemas tem a interação humana direta com os dados com o objetivo de garantir qualidade dos modelos de ML, por isso esta interação precisa ser a mais eficiente possível, tanto para quem está interagindo quanto para o modelo de ML (AMERSHI et al., 2014). Dentre os sistemas de ML interativo estão presentes os sistemas de rotulagem de dados.

2.2.1 Sistemas de Rotulagem de Dados

Os algoritmos de ML supervisionado necessitam de muitos dados rotulados para aprender. Isso quer dizer que os dados precisam estar previamente classificados. Se esta classificação depender de pessoas, é necessário apresentar estes dados da melhor forma possível. Reduzir o esforço de rotulagem é um dos principais objetivos dos sistemas de rotulagem (YUAN et al., 2020).

Sistemas de rotulagem apresentam muitos dados para serem rotulados para os usuários. A tarefa da rotulagem pode envolver problemas de classificação (CHOI et al., 2019), detecção de entidades (GROSMAN et al., 2020), segmentação de objetos (HEBBALAGUPPE et al., 2013) e muitos outros (VIANA; OLIVEIRA; CONTE, 2021).

Dependendo do domínio dos dados, estes usuários podem ser especialistas ou não, como é o caso dos sistemas de rotulagem *crowdsourcing*. Estes sistemas contam com milhares de usuários para realizar a tarefa de rotular os dados. Então, em seguida, a validação destas rotulagens pode ser feita por especialistas ou usuários experientes (YUAN et al., 2020).

Um exemplo de sistema de rotulagem é o trabalho de Choi et al. (2019). Os autores propuseram um sistema de rotulagem de documentos com um módulo inteligente que destacava as palavras consideradas importantes para classificação em cada documento como pode ser visto na Figura 2.3. Para a avaliação deste sistema, foram utilizados *reviews* de estabelecimentos para que os usuários os classificassem em positivo ou negativo. Como resultado, os autores mostraram que a eficiência de rotulagem dos participantes aumentou significativamente com a presença do módulo inteligente de destaque das palavras relevantes.

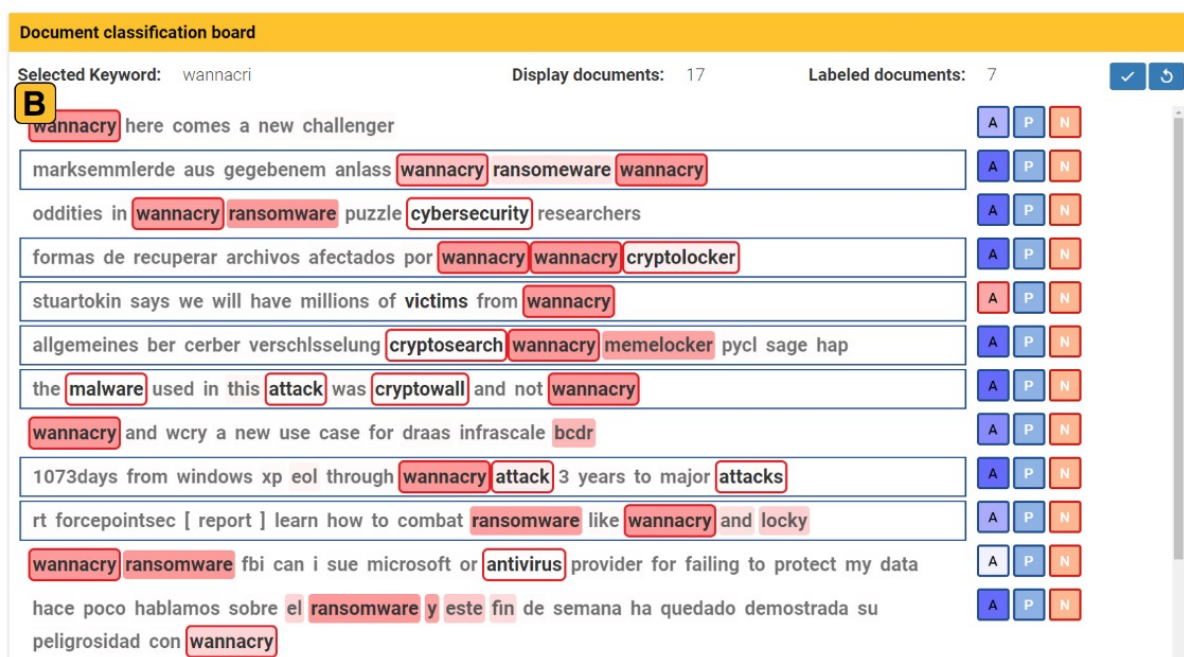


Figura 2.3 – Exemplo de sistema de rotulagem de dados (CHOI et al., 2019).

Cada vez mais é aproveitada a abundância de dados capturados e criados pelos sistemas de informação da atualidade (WARD; GRINSTEIN; KEIM, 2010). Estes dados são investigados para se adquirir conhecimentos em um determinado domínio, porém a investigação destes em grandes volumes é limitada para as pessoas pela dificuldade de interpretá-los (CUI, 2019).

Muitas iniciativas vêm sendo realizadas em torno de ajudar as pessoas na interpretação de dados em grande quantidade (PORTER; THEILER; HUSH, 2013; ACUNA et al., 2018). Essas iniciativas envolvem a criação ou uso de um sistema com técnicas que ajudam as pessoas na visualização e interpretação dos dados (THOMAS; COOK, 2005).

A definição de tipo de dado adotada neste trabalho refere-se à natureza do dado associado à tarefa de rotulagem, como dado do tipo texto, imagem, áudio e vídeo (AWS, 2021; NADJ et al., 2020). Por exemplo, quando o tipo do dado é imagem, os recursos visuais que possibilitam o usuário interagir são muito intuitivos devido à própria natureza do dado (KIM et al., 2015b). Recursos como editar a imagem, recortar, inserir elementos, entre outros são de uso comum na área de manipulação e edição de imagens e servem como base para esses sistemas.

Quando o tipo de dado é textual, os recursos desses sistemas devem auxiliar os usuários na interpretação dos textos (STENETORP et al., 2012). No entanto, a elaboração de recursos com textos não é simples (GROSMAN et al., 2020) visto que a maioria dos textos não está disponível de forma estruturada, ou seja, de fácil compreensão (GROSMAN et al., 2020). Além disso, os sistemas precisam de diferentes interações dependendo do problema de rotulagem de texto. Por exemplo, para classificar um texto como emoção positiva ou negativa, o usuário precisa ler, entender o texto e rotular o texto selecionando a classe correta. No entanto, se a tarefa é identificar nomes de pessoas, o usuário precisa interagir com muitas palavras em um único texto. Todos esses problemas de texto são chamados problemas de processamento de linguagem natural, do inglês *Natural Language Processing* (NLP).

2.2.2 Processamento de Linguagem Natural

Quando se trata de rotular texto, podemos associá-los ao NLP que é uma subárea de ML que trata de problemas de texto (CHOPRA; PRASHAR; SAIN, 2013). Os principais problemas de NLP requerem rotulagem de texto, e cada um deles tem diferentes interações de rotulagem. Esses problemas podem ser (OTTER; MEDINA; KALITA,

2020; DUDHABAWARE; MADANKAR, 2014):

- **Classificação de texto (*Text classification*)** visa rotular um texto inteiro ou linhas de texto em uma única categoria, ou multi categoria (OTTER; MEDINA; KALITA, 2020).
- **Análise de sentimentos (*Sentiment Analysis*)** visa conhecer a opinião sobre um determinado texto classificando-o como positivo ou negativo (DUDHABAWARE; MADANKAR, 2014). Neste problema, a interação de rotulagem é a mesma que a classificação do texto.
- **Extração de relação (*Relation extraction*)** é o processo de vincular entidades para entender melhor a estrutura do texto e os relacionamentos entre entidades (OTTER; MEDINA; KALITA, 2020).
- **Reconhecimento de Entidade Nomeada (*Named Entity Recognition (NER)*)** também é conhecido como extrair, separar e identificar entidades (OTTER; MEDINA; KALITA, 2020). As categorias comuns usadas para anotação NER incluem nomes de organizações, lugares, pessoas e outros (DUDHABAWARE; MADANKAR, 2014).
- **Marcação de parte do discurso (*Part-of-Speech (POS) Tagging*)** é a rotulagem de elementos de fala funcionais em dados de texto, como adjetivos e advérbios (DUDHABAWARE; MADANKAR, 2014).
- **Marcação de frase-chave (*Keyphrase Tagging*)** é o procedimento para localizar frases-chave ou palavras-chave no texto (DUDHABAWARE; MADANKAR, 2014). Também conhecido como extração de palavras-chave, ajuda a melhorar as funções relacionadas à pesquisa para bancos de dados (PAPAGIANNPOULOU; TSOUMAKAS, 2020).

Desses problemas de NLP, foram considerados os três últimos como Anotação de Entidades porque esses problemas visam localizar e rotular menções de palavras ou frases no texto. Todos os problemas de anotação de entidades têm a mesma interação

em um sistema de rotulagem. A Figura 2.4 mostra um exemplo de sistema de rotulagem com problema de anotação de entidades (NAKAYAMA et al., 2018).

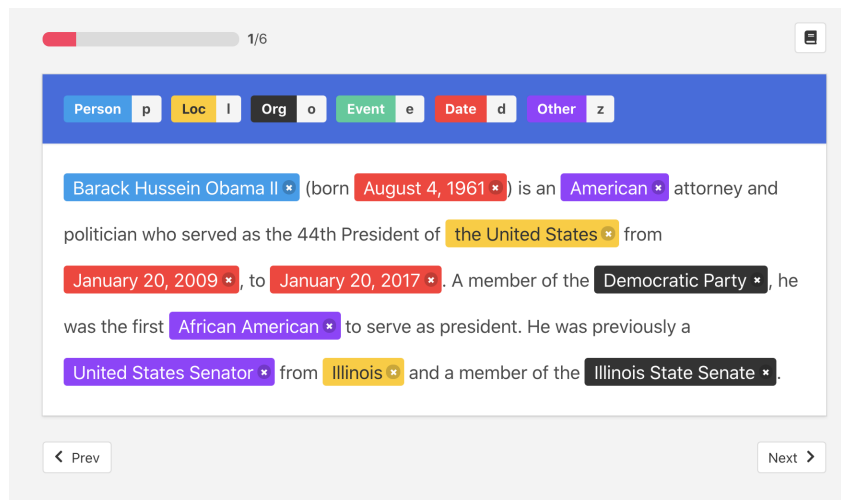


Figura 2.4 – Exemplo de um sistema de rotulagem com problema de anotação de entidades (NAKAYAMA et al., 2018). A interface mostra uma instância de texto e exige que o usuário identifique as entidades relacionadas às categorias *Person*, *Location*, *Organization*, *Event*, *Date* e *Other*.

2.3 Trabalhos Relacionados

O projeto de sistemas de rotulagem é um tópico que tem sido abordado pela academia nos últimos anos (DUDLEY; KRISTENSSON, 2018b; NADJ et al., 2020; AMERSHI et al., 2014). Segundo Amershi et al. (2014), cada vez mais tem sido destacada a importância de investigar a interação do usuário com o processo de construção de modelos ML. Com base nisso, os pesquisadores realizaram um estudo experimental para investigar os desafios dos usuários nesses sistemas e destacar o papel e a importância do usuário dentro do *Interactive Machine Learning*. Os autores encontraram desafios e oportunidades pertinentes para o avanço do estado-da-arte nesta área.

Dentre os desafios e oportunidades, Amershi et al. (2014) destacam a necessidade de definir princípios e diretrizes sobre como projetar a interação humana com ML. É levantado a seguinte desafio: como projetar a interação com o ML interativo, assim como são projetadas as interfaces de sistemas tradicionais. Assim, há uma oportunidade de explorar como os princípios atuais de *design* se aplicam aos sistemas de ML interativo.

Além disso, os autores incentivam a colaboração entre áreas de IHC e ML para tratar dos problemas de interações em sistemas de ML interativo e aumentar o desempenho dos modelos ML.

Segundo [Dudley e Kristensson \(2018b\)](#), existe uma lacuna sobre os princípios de projeto de sistemas dentro do *Interactive Machine Learning*. Para cobrir essa lacuna, os pesquisadores realizaram um levantamento da literatura, onde abstraíram quatro canais de comunicação presentes nas interfaces de sistemas de ML interativo que inclui sistemas de rotulagem. Esses quatro canais são: revisão de amostra, atribuição de *feedback*, inspeção de modelo e visão geral da tarefa. O objetivo desses canais é fornecer uma visão geral sobre o que apresentar ao usuário neste tipo de sistema. Além disso, o artigo apresenta um fluxo de atividades que o usuário pode desempenhar para o desenvolvimento de um modelo ML.

Assim como na pesquisa de [Dudley e Kristensson \(2018b\)](#), [Nadj et al. \(2020\)](#) apresentam uma lacuna focada no *design* de sistemas de rotulagem. Os pesquisadores afirmam que faltam princípios consolidados sobre como projetar esse tipo de sistema. Para isso, os autores realizaram uma Revisão Sistemática da Literatura (RSL), onde analisaram 44 artigos sobre sistemas de rotulagem de dados.

Como resultado, no trabalho de [Nadj et al. \(2020\)](#), foram definidos cinco princípios de *design* para sistemas de rotulagem, focando em soluções que reduzam a frustração e chateação dos usuários durante a tarefa de rotulagem. Estes cinco princípios são: inclua a intenção dos usuários, apoie o envolvimento dos usuários, aumente a relevância percebida dos usuários, avalie o grau de interrupção e ajuste a transparência dos usuários.

A partir da análise desses trabalhos relacionados, foi possível destacar a importância de propor orientações para auxiliar no projeto de interface de sistemas de domínios específicos. A diferença das deste trabalho dos demais trabalhos relacionados sobre sistemas de *Interactive Machine Learning* está em apresentar exemplos visuais sobre o que e como as interfaces devem apresentar aos usuários durante a atividade de rotulagem de dados. Além disso, esta pesquisa considera apenas o dado textual, pois cada tipo de dado apresenta uma complexidade de interação e visualização.

3

MOTIVAÇÃO: CASO SEFAZ

Neste capítulo é descrito o estudo de caso do projeto de sistema de rotulagem de descrições de produtos de notas fiscais da SEFAZ-Amazonas. Este estudo foi a motivação inicial para esta pesquisa de mestrado e mostra a relevância deste trabalho na perspectiva da indústria.

Na Seção 3.1 é introduzido o problema das Secretarias da Fazenda e a necessidade de sistemas de rotulagem. Em seguida, foi realizada uma análise de sistemas de rotulagem na literatura (Seção 3.2) e processo de evolução dos protótipos criados com as soluções para sistemas de rotulagem (Seção 3.3). Por fim, na Seção 3.4 é apresentada a discussão dos resultados obtidos.

Os resultados deste capítulo foram registrados em um artigo com o título *Rotule-me! Uma experiência de engenharia de requisitos para um sistema de rotulagem*. Este artigo foi submetido e aceito no *Workshop* de Engenharia de Requisitos (WER 2021) (PASSOS et al., 2021a).

3.1 Sistema de rotulagem de NF-e's SEFAZ

Esta seção descreve o problema enfrentado pelas Secretarias da Fazenda dos Estados de todo o Brasil, que levou ao desenvolvimento do sistema de rotulagem de dados. Além disso, é detalhada a elicitação de requisitos deste sistema, que motivou esta pesquisa.

As Secretarias da Fazenda já processaram, desde outubro de 2005, sob o projeto Nota Fiscal Eletrônica (NF-e), mais de 26.614 bilhões de NF-e's (Secretaria da Fazenda,

2021). As NF-e's registram as operações comerciais realizadas nos estabelecimentos e indústrias, podendo, cada nota, conter até 200 descrições dos produtos comercializados.

A partir dos dados constantes nas NF-e's, os governos estaduais realizam o cálculo do Imposto sobre Circulação de Mercadorias e Serviços (ICMS), que, por fim, financiará todas as políticas governamentais, desde a educação até a saúde pública. Caso a identificação do produto descrito na NF-e seja incorreta, haverá injustiça fiscal, incorrendo na cobrança a maior do ICMS, prejudicando o contribuinte, ou a cobrança a menor, prejudicando o governo estadual (Governo do Estado de São Paulo, 2021).

Uma das variáveis do cálculo do ICMS é o código NCM ou Nomenclatura Comum do Mercosul (Receita Federal - Ministério da Economia, 2021). Cada produto descrito nas NF-e's é classificado em um código NCM pelos contribuintes. Porém, é comum que existam muitos erros nestas classificações, seja pela dificuldade em classificar um produto em uma das 10.200 possibilidades de NCM, ou seja, pela intenção de sonegação fiscal (Receita Federal - Ministério da Economia, 2021). Um exemplo de um código NCM com produto classificado incorretamente pode ser visto na Figura 3.1. Neste exemplo, uma garrafa térmica foi classificada com um código NCM de inseticida.

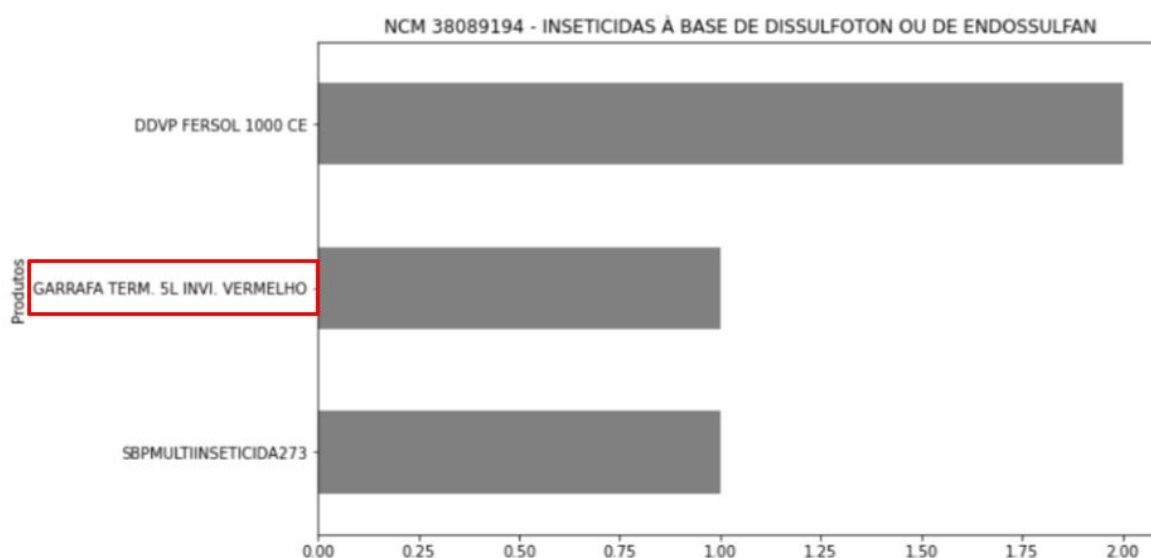


Figura 3.1 – Exemplo de um produto classificado incorretamente.

Tendo em vista as milhões de descrições de produtos em NF-e's recebidas pelo Fisco Estadual, é necessário verificar se a cada uma dessas descrições foi atribuído

o código NCM correto pelos contribuintes. Entretanto, o grande volume de NF-e's recebidas pelas secretarias diariamente, torna humanamente impossível a tarefa de validar item a item os milhares de produtos comercializados.

A solução escolhida para resolver este problema foi desenvolver um modelo de ML para classificar automaticamente cada produto. Porém, é necessário, anteriormente, ter uma boa base de dados para treinamento (ALPAYDIN, 2020). Isto significa dizer que esta base precisa possuir uma grande quantidade de dados distintos que representem cada classe corretamente (ROH; HEO; WHANG, 2019).

Para obter esta base para treinamento é necessário construir um sistema de rotulagem de dados onde especialistas avaliam a classificação de milhões de produtos de acordo com o código NCM correto. Entretanto, esta tarefa de rotulagem dos dados demanda muito tempo dos especialistas, o que pode causar desconforto e desmotivação ao usar o sistema (FELIX; DASGUPTA; BERTINI, 2018).

O sistema de rotulagem deste problema em específico considera que seus usuários são especialistas em classificar títulos de produtos para um determinado código NCM. Devido ao grande volume de dados a serem apresentados e a especialidade dos usuários, é necessário, portanto, propor soluções de interface para entender como melhor apresentar estes dados e como classificar cada título corretamente com menos esforço.

Para propor estas soluções, foram realizadas interações entre o gerente de produto e a equipe de pesquisa deste trabalho para definir quais eram os requisitos iniciais deste sistema. Estas interações tiveram como base um processo de elicitação baseado em prototipagem (MANNIO; NIKULA, 2001). Na Tabela 3.1, são apresentados os requisitos principais iniciais para este sistema. Os requisitos foram elaborados considerando o objetivo do sistema de rotulagem.

ID	Requisitos Iniciais
RI1	O sistema deve possuir elementos na interface para acelerar o processo de rotulagem (por exemplo, palavras destacadas para facilitar a leitura dos dados ou visualização de múltiplas instâncias para rotulagem em grupo.)
RI2	O sistema deve apresentar os produtos e as suas respectivas categorias para a rotulagem
RI3	O sistema deve fornecer informações complementares para auxiliar o usuário em tomadas de decisão quanto a tarefa de rotulagem

Tabela 3.1 – Requisitos principais iniciais do sistema de rotulagem.

Estes requisitos principais iniciais foram utilizados como base para a construção dos protótipos de interface. Com as apresentações dos protótipos, os requisitos foram refinados, tornando-se mais concretos durante as iterações do processo.

3.2 Análise de elementos de interface na literatura

Para resolver os problemas apresentados na Seção 3.1, foi primeiramente realizada uma pesquisa exploratória na literatura científica com o objetivo de analisar estudos sobre sistemas de rotulagem e seus elementos de interface. Foram analisados: o trabalho de [Dudley e Kristensson \(2018b\)](#) e as RSL de [Amershi et al. \(2014\)](#) e [Nadj et al. \(2020\)](#).

O resultado da pesquisa exploratória permitiu identificar elementos de interface aplicados nos sistemas de rotulagem que serviram como base para a criação dos protótipos do sistema de rotulagem proposto neste estudo. Conforme a Figura 3.2, estes elementos foram identificados como: Listas Cruzadas (E1) ([KIM et al., 2015a](#)), *Feedback* de Atributos Positivos e Negativos (E2) ([AMERSHI et al., 2014](#)), Palavras Relevantes em Cores (E3) ([WALLACE et al., 2012](#)) e Escolha Binária Sobre a Relevância do Dado (E4) ([PLUMMER et al., 2019](#)).

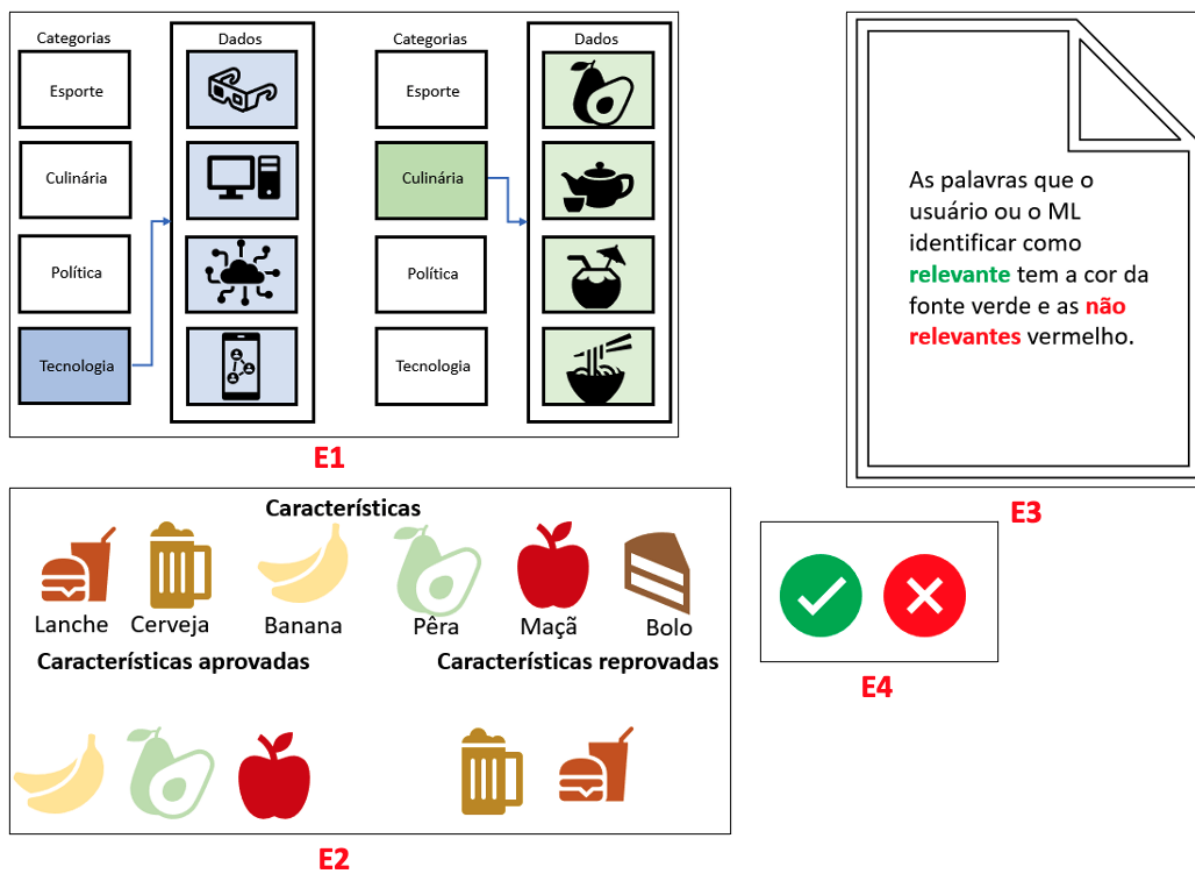


Figura 3.2 – Elementos de interface de sistemas de rotulagem encontrados na literatura.

1. **E1.** As listas cruzadas se dividem em duas listas, a do lado esquerdo apresenta as categorias de dados e a do lado direito são apresentados os dados de cada categoria conforme o usuário clique em uma delas. O E1 auxilia na visualização de dados de cada categoria numa disposição paralela entre as listas. Assim, é possível visualizar rapidamente várias instâncias pertencentes a um determinado grupo.

Na Figura 3.2.E1 são apresentados dois exemplos de listas cruzadas. No primeiro exemplo, a categoria selecionada é tecnologia (cor azul), e como consequência são apresentados os dados desta categoria. No segundo exemplo, a categoria selecionada é culinária (cor verde) e assim como no exemplo anterior, são apresentados os dados desta categoria.

2. **E2.** O usuário fornece *feedback* identificando atributos positivos e negativos. Este elemento agrupa essas características em um grupo de inclusão e outro de exclusão. Assim, o usuário pode verificar os dados e excluir as características que

não representem sua respectiva classe. Por exemplo, se o objetivo do usuário é classificar textos que tenham apenas as palavras banana, pêra e maçã, mas também não queira textos com as palavras cerveja e lanche, o resultado desse elemento será apresentar apenas os textos que atendam as características aprovadas e que não tenham as características reprovadas, conforme a Figura 3.2.E2.

3. E3. Palavras relevantes em cores destacam palavras importantes em um texto. Este elemento auxilia na visualização rápida de termos em um texto através de cores que cativem a atenção do usuário. Assim, a interpretação do texto na totalidade é agilizada e traz a atenção do usuário para o assunto central do dado analisado. Na Figura 3.2.E3, as palavras identificadas como relevantes possuem a cor da fonte verde e as não relevantes da cor vermelha.
4. E4. Escolha sobre a relevância do dado é uma interação do usuário ao indicar se um dado é relevante ou não. Este elemento auxilia na coleta do *feedback* do usuário sobre um determinado dado com o toque de um único botão. Além disso, este elemento é indicado para problemas de classificação binária, ou seja, escolher se um dado pertence em uma dentre duas classes. Na Figura 3.2.E4, são utilizados o ícone “check” para indicar o texto relevante e “x” para indicar o texto como não relevante.

3.3 Soluções propostas

Nesta Seção, são apresentados os protótipos construídos com os elementos encontrados na pesquisa exploratória (Seção 3.2) para solucionar o problema descrito na Seção 3.1.

3.3.1 Protótipo Inicial

Conforme a Figura 3.4, inicialmente pensou-se na criação de um protótipo modelado para a rotulagem dos códigos NCM associados aos produtos, ou seja, a tarefa do usuário seria de selecionar o código NCM correto dentre os códigos sugeridos. Neste protótipo,

foi utilizado uma adaptação do 3.2.E4 como solução para obter *feedback* do usuário sobre o rótulo do produto com apenas um toque (Figura 3.3).

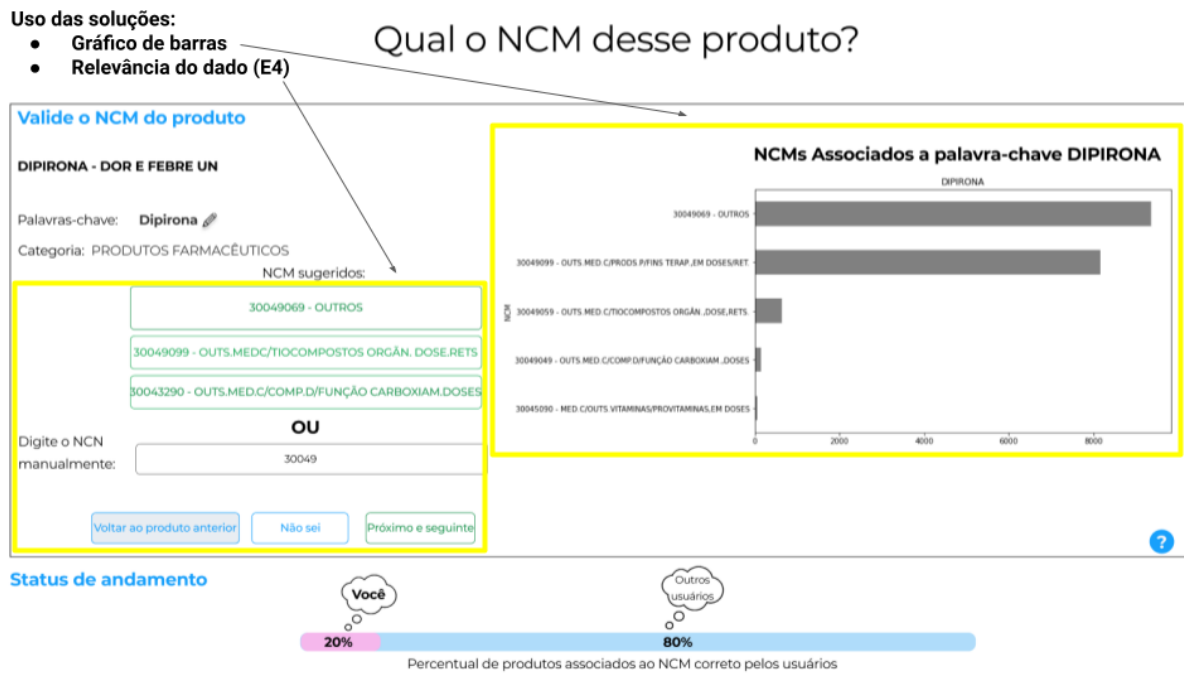


Figura 3.3 – Soluções utilizadas no segundo protótipo.

Outros elementos de interface também foram utilizados. Um deles foi um gráfico de barras mostrando os códigos NCM associados aos produtos que tinham uma determinada palavra na sua descrição, como também o status do andamento da tarefa.

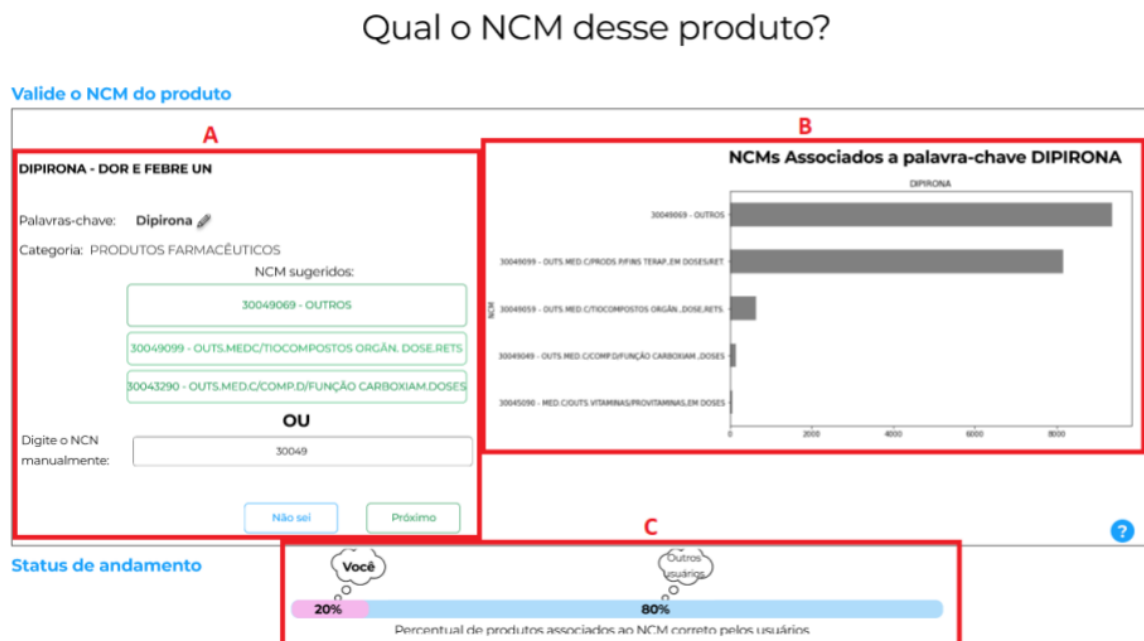


Figura 3.4 – Protótipo inicial.

Em conjunto com o gerente de produto do projeto foram avaliados os Pontos Fortes (PF) e fracos do protótipo apresentado na Figura 3.4. Os pontos fortes são:

- **PF1: gráfico de barra de uma palavra representativa de um produto (Figura 3.4.B)**
O gráfico auxilia o usuário na escolha do código NCM correto apresentando os códigos NCM mais associados a um determinado produto.
- **PF2: status de andamento (Figura 3.4.C)** Através do status de andamento, o usuário é informado quanto à quantidade de produtos categorizados por ele e por outros usuários.

Porém, um ponto fraco encontrado é com relação à estratégia de rotulagem utilizada. Um dos pontos levantados pelo gerente de produto é o fato do usuário precisar avaliar o código NCM de cada produto, porém as secretarias recebem milhões de notas por dia o que torna inviável ao usuário avaliar todos os produtos. Outro ponto é que a avaliação do código NCM de cada produto pode provocar fadiga no usuário, o que pode afetar a qualidade da avaliação devido ao cansaço gerado durante a atividade de rotulagem.

3.3.1.1 Avaliação e lições aprendidas

Tendo em vista a inviabilidade da solução empregada no protótipo da Figura 3.4, foram discutidas outras estratégias para contornar este problema. Uma das estratégias que resolveria este problema é solicitar do usuário a avaliação de grupos de produtos associados a um código NCM, visto que estes códigos são um conjunto de grandeza menor (10.200) comparado com o número de produtos descritos nas NF-e's recebidas pelas secretarias por dia (≈ 1 MI). Além disso, foi observado que existem termos que são mais representativos em descrições de produtos de um determinado código NCM, como poder ser visto na Figura 3.5.

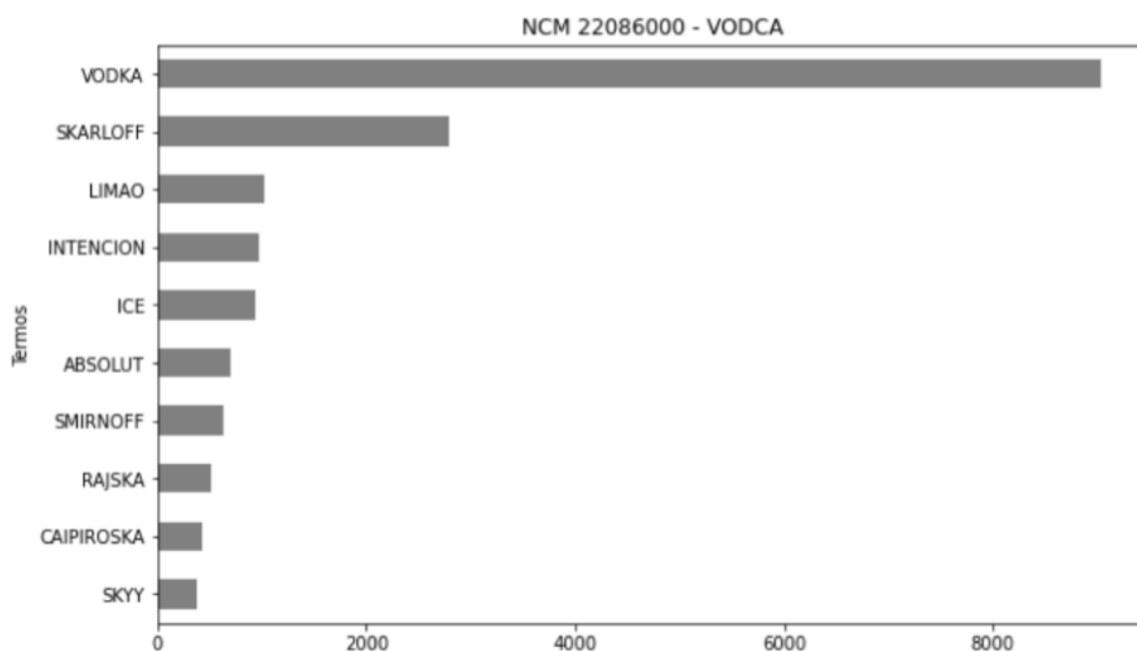


Figura 3.5 – Gráfico de frequência de termos representativos.

3.3.2 Segundo Protótipo

Para solucionar o problema encontrado pelo gerente de produto na proposta anterior, foi construído um novo protótipo analisando quais elementos podiam ser mais compatíveis. Assim, foram utilizadas todas as soluções E1, E2, E3 e E4 para este protótipo (Figura 3.6).

Uso das soluções:

- Listas Cruzadas (E1)
- Grupos de inclusão e exclusão (E2)

Valide os NCM aos produtos

NCM selecionado: 22086000 - VODCA

Uso das soluções:

- Palavras destacadas (E3)
- Relevância do dado (E4)

NCMs	Palavras Relevantes	Palavras de Inclusão	Palavras de Exclusão												
22086000 - VODCA	<table border="1"> <thead> <tr> <th>Palavras</th> <th>Distrib</th> </tr> </thead> <tbody> <tr> <td>Alcool</td> <td>550</td> </tr> <tr> <td>Vodka</td> <td>350</td> </tr> <tr> <td>Vodca</td> <td>200</td> </tr> <tr> <td>Vinagre</td> <td>150</td> </tr> <tr> <td>Beats</td> <td>50</td> </tr> </tbody> </table>	Palavras	Distrib	Alcool	550	Vodka	350	Vodca	200	Vinagre	150	Beats	50	VODCA, VODKA	Cerveja, Alcool
Palavras	Distrib														
Alcool	550														
Vodka	350														
Vodca	200														
Vinagre	150														
Beats	50														

Produtos (Geral) | Produtos não rotulados | Produtos rotulados

VODKA MARCA| LIMAO 1000ML
VODKA MARCA| GREEN APPLE 1000ML
CAIPIRINHA VODKA
VODKA MARCA| 275ML LIM
MARCA| VODKA TRIDESTILADA 965 ML

Total Geral 24

Aprovar todos os produtos desta página

Figura 3.6 – Soluções utilizadas no segundo protótipo.

O protótipo apresentado na Figura 3.7 foi criado com foco na análise dos produtos por NCM. O usuário precisa primeiro selecionar um código NCM, conforme Figura 3.7.A. Em seguida, os componentes são carregados com as informações de todos os produtos associados ao código NCM selecionado, conforme o uso da solução E1.

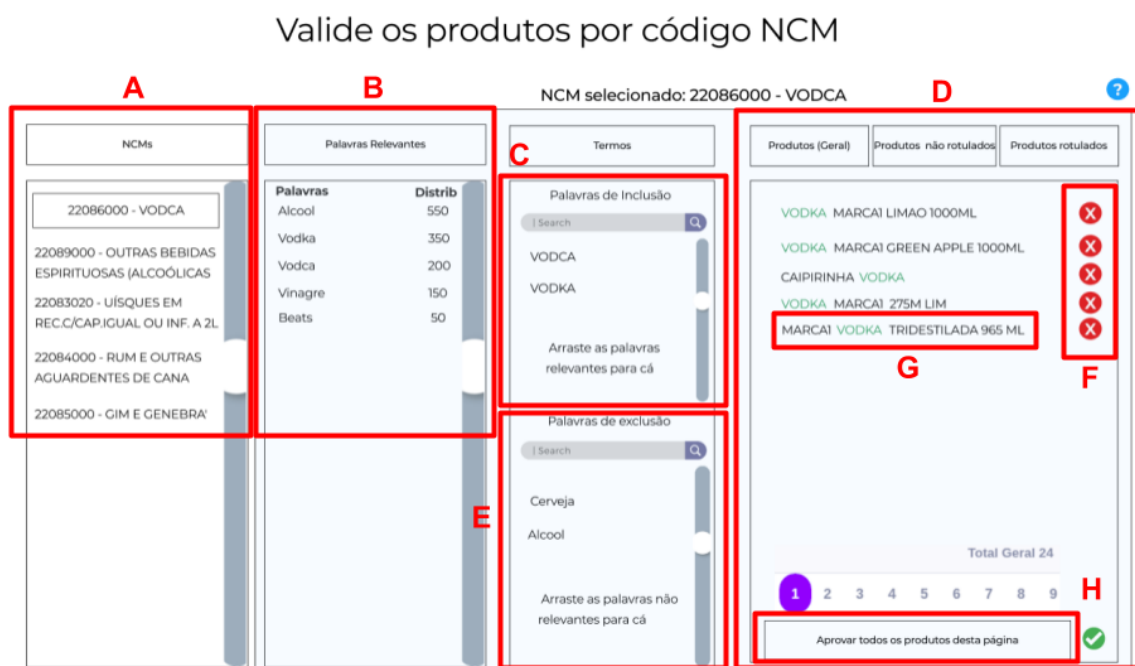


Figura 3.7 – Segundo protótipo proposto.

Após a finalização do protótipo, novamente foi realizada uma validação do protótipo com o gerente de produto. Nesta validação, foram identificados os seguintes pontos fortes:

- **PF3: Agrupamento e filtros de termos representativos (Figura 3.7.B)** O filtro de termos apresenta todos os termos únicos dos produtos associados ao código NCM selecionado. Além disso, é apresentado a frequência que esses termos ocorrem, possibilitando ao usuário verificar quais os termos mais utilizados nos produtos. O uso da solução E2 permitiu que os termos fossem “arrastados” para os componentes de palavras de inclusão e exclusão conforme a Figura 3.7.C, assim estes termos refletem nos dados apresentados no componente D da Figura 3.7.
- **PF4: Destaque de termos aprovados e reprovados por cor (Figura 3.7.G)** A inserção de termos no componente C e E da Figura 3.7 altera a cor da fonte na descrição do produto (Figura 3.7.G), sendo verde para indicar que a palavra é um termo

de relevância para a categorização do produto e vermelho para indicar o termo como não relevante, conforme o uso da solução E3. Além disso, com o uso da solução E4, foi possível aprovar grupos de produtos ou reprovar um produto em específico.

Os pontos fracos identificados pelo gerente de produto no segundo protótipo foram:

- **Sobrecarga de informações no protótipo** Apesar de todas as informações serem necessárias, existe uma sobrecarga de informações que pode confundir o usuário e dificultar a análise ao expor as informações de diversos códigos NCM junto com muitos títulos de produtos. Uma sugestão do gerente de produto foi a possibilidade de analisar os produtos pertencentes a um código NCM pré-selecionado anteriormente.
- **Impossibilidade de agrupar termos** Há termos que combinados com outros devem ser classificados em conjunto como critérios de inclusão ou exclusão. Um exemplo seria de um código NCM que descreve os produtos de combustível do tipo álcool que, neste caso, possui o termo “álcool” como o mais representativo, porém este em conjunto com o termo “gel” não deve ser considerado. Mesmo com a utilização da solução E4 para remover itens específicos do grupo, ainda se tornaria inviável remover item a item, caso existissem uma quantidade significativa desses produtos incorretos.

3.3.2.1 Avaliação e lições aprendidas

O problema destacado pelo gerente de produto no protótipo inicial foi resolvido através da estratégia de apresentar os produtos agrupados por código NCM e assim possibilitando a avaliação de produto por grupo e não mais individualmente.

Além disso, neste protótipo foi utilizado o elemento E2 como solução para a inclusão e exclusão de termos representativos, dado que alguns termos estão presentes

em uma grande quantidade de produtos. Entretanto, foi levantada a necessidade de fazer a separação por grupos de termos e um design mais simples e limpo.

3.3.3 Terceiro Protótipo

Dado os problemas relatados pelo gerente de produto no protótipo apresentado na Subseção 3.3.2, foi realizada uma nova iteração para incluir as soluções de elementos que atendam estes problemas. Novamente, foram utilizadas todas as soluções E1, E2, E3 e E4, como também o gráfico de barras (Figura 3.8).

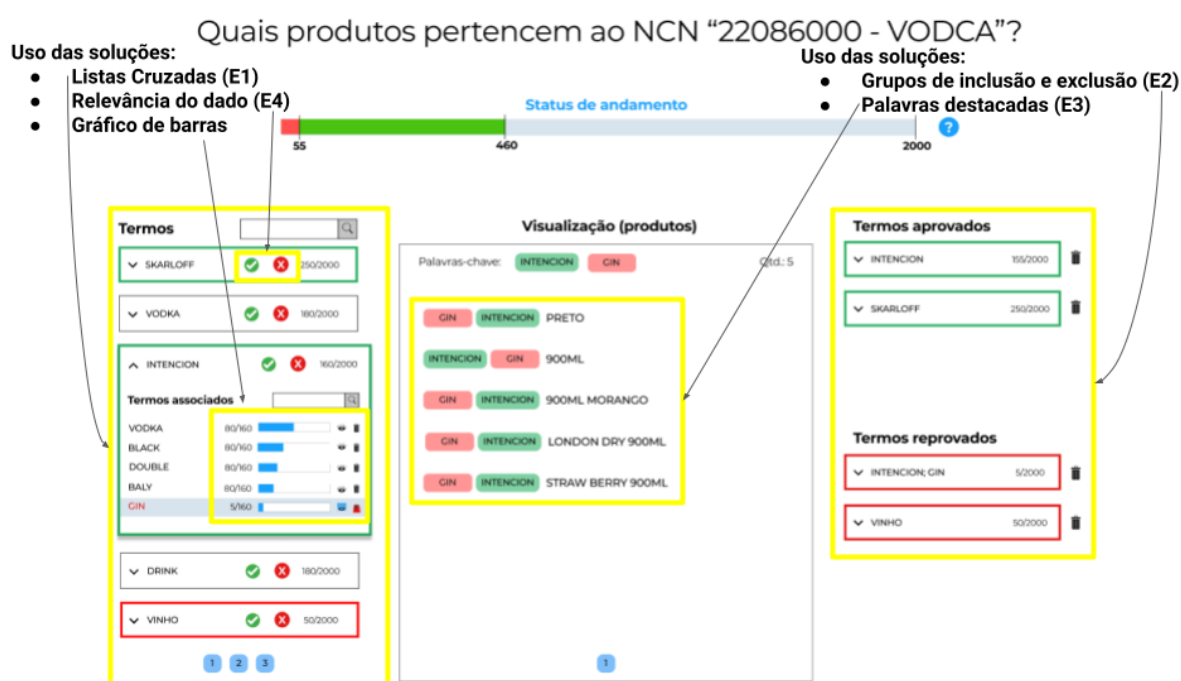


Figura 3.8 – Soluções utilizadas no terceiro protótipo.

O novo protótipo foca no design mais simples através da avaliação de produtos de apenas um código NCM por vez, como pode ser visto na Figura 3.9. Além disso, neste protótipo é possível classificar grupos de termos, como também visualizar a frequência de palavras associadas a cada termo.

Após a finalização do protótipo, foi realizada uma nova validação de protótipo com o gerente de produto do projeto. Os seguintes pontos fortes foram identificados:

- **PF5: Design clean (Figura 3.9)** O novo protótipo aparenta conter menos informações por conta da exibição de produtos de apenas um NCM por vez. A

Quais produtos pertencem ao NCN "22086000 - VODCA"?



Figura 3.9 – Terceiro protótipo proposto.

visualização de produtos no componente C está associada a interação do usuário nos componentes B e E, conforme a utilização da solução E1.

- **PF6: Inclusão de grupos de termos para exclusão ou inclusão (Figuras 3.9.D e 3.9.F)** No componente F, é possível ver que o termo principal "MARCA2" em conjunto com "gin" foram reprovados pelo usuário. Assim, conforme utilização da solução E4, um termo pode ser aprovado, mas também pode ser reprovado em conjunto com outro termo, como pode ser visto nos componentes D e F que representam o uso da solução E2. Quando um termo é aprovado, a borda do *card* fica da cor verde, e quando reprovado, a borda fica da cor vermelha.

3.3.3.1 Avaliação e lições aprendidas

Os problemas destacados pelo gerente de produto no protótipo anterior foram solucionados através de um design mais simples e com a possibilidade de agrupar termos para classificá-los como termos aprovados ou reprovados. Por fim, este protótipo foi aprovado para dar continuidade ao projeto.

3.4 Discussão

Este capítulo apresenta um processo evolutivo de engenharia de requisitos utilizando protótipos de sistemas de rotulagem de dados, visando atender o problema das Secretarias da Fazenda sobre a tarefa de rotulagem de milhões de produtos. Para isso, foram realizadas três iterações com os protótipos para o sistema proposto.

A cada iteração, os requisitos do sistema foram evoluídos (Tabela 3.2). A evolução aconteceu a partir dos requisitos iniciais apresentados na Tabela 3.1. Além disso, foram analisados pontos fortes e fracos de cada protótipo e a partir disso foi feita uma reflexão entre esses pontos e os requisitos iniciais, tendo como resultado os requisitos refinados. Estes requisitos refinados foram validados pelo gerente de produto e pela equipe de desenvolvimento. A Tabela 3.2 apresenta o ID dos requisitos iniciais da Tabela 3.1, os requisitos refinados, e por fim, os pontos fortes que possibilitaram a evolução dos requisitos.

Apesar dos pontos fracos não estarem relacionados na Tabela 3.2 com os requisitos refinados, esses foram importantes no processo evolutivo dos requisitos, pois possibilitaram a reflexão sobre o que o sistema não devia fazer. Além disso, os pontos fracos evidenciaram problemas como a avaliação individual de produtos, sobrecarga de informações no protótipo e a impossibilidade de agrupar termos.

Req. Iniciais	Requisitos Refinados	Pontos Fortes
RI2	RR1: Visualizar títulos de produtos por NCM pré-selecionado	PF5: Design <i>clean</i>
RI3	RR2: Visualizar status de progresso de títulos rotulados	PF2: Status de andamento
RI3	RR3: Visualizar distribuição e quantidade de títulos de produtos	PF1: Gráfico de barra de uma palavra representativa de um produto
RI1	RR4: Identificar visualmente os termos aprovados e reprovados	PF4: Destaque de termos aprovados e reprovados por cor
RI1	RR5: Rotular títulos por termos ou grupos de termos	PF6: Inclusão de grupos de termos para exclusão ou inclusão
RI1	RR6: Possibilitar filtros por termos	PF3: Filtros de termos

Tabela 3.2 – Requisitos Refinados do Sistema de Rotulagem a partir dos requisitos iniciais e dos pontos fortes dos protótipos.

Como exemplo, o requisito RR1 foi um refinamento do requisito RI2, pois deixa

claro como devem ser apresentados os produtos e suas categorias. Este refinamento foi necessário devido à ambiguidade do requisito inicial RI2. Além disso, durante a construção dos protótipos, foi observado que a visualização de produtos por NCM proporcionava uma interface mais simples e limpa, conforme o ponto forte PF5 apontado pelo gerente de produto.

Os requisitos RR1 ao RR4 referem-se a requisitos de visualização. Esses requisitos de visualização foram adaptados de outros sistemas de rotulagem encontrados na literatura. Nesse sentido, os elementos encontrados contribuem para a melhor identificação das informações e contextualizar o usuário do que está acontecendo no sistema.

O requisito RR5 representa uma abordagem de rotular por termos ou grupos de termos. A importância desse requisito é gerada pela necessidade do usuário em avaliar muitos dados. A partir disso, foi identificado que existem muitos produtos com termos semelhantes e como solução foram criados componentes de interface que apresentassem ao usuário os termos semelhantes mais usados nas descrições dos produtos. Assim, o especialista, ao invés de classificar milhões de produtos individualmente, identifica apenas os termos mais representativos, como também os termos menos representativos que na maioria das vezes são produtos que foram classificados com o código NCM incorreto pelos contribuintes.

O requisito RR6 diz respeito a possibilidade de identificar termos através de filtros de pesquisa. Assim, os usuários podem pesquisar um termo específico, levando em consideração que esses usuários são especialistas e podem ter uma ideia prévia de quais termos devem ser analisados primeiro.

A última evolução do protótipo atendeu a todos os requisitos refinados, o que permitiu à equipe de desenvolvimento prosseguir com a construção do sistema tomando como base esses requisitos refinados.

Por fim, este foi um estudo de caso participativo na SEFAZ que trouxe uma experiência prática sobre projeto de sistemas de rotulagem de dados. Com esse estudo, foi possível perceber que projetar um sistema de rotulagem para atender um problema de ML pode ser difícil devido à complexidade de entendimento de conceitos de ML e rotulagem de dados. Além disso, há pouco suporte para a equipe de desenvolvimento

quanto a questão de como projetar estes sistemas. Esse estudo motivou uma discussão de como orientar os profissionais neste contexto. Assim, os próximos passos desta pesquisa de mestrado foram orientados em desenvolver *guidelines* para apoiar no projeto de interface de sistemas de rotulagens de dados textuais.

4

ARTEFATO INICIAL: GUIDELINES LABELUX

Neste capítulo, são apresentadas as *guidelines LabelUX*. As *guidelines* foram o primeiro artefato construído nesta pesquisa com intuito de responder à QP desta dissertação.

Na Seção 4.1, é descrito o processo conduzido durante a criação das *guidelines*. Na Seção 4.2, são descritas cada uma das *guidelines*. Na Seção 4.3, é apresentado como cada guideline está associada com os estudos anteriores.

Por fim, as Seções 4.4, 4.5 e 4.6 descrevem estudos de avaliação das *guidelines*. O primeiro estudo foi um estudo de viabilidade, o segundo foi uma análise das *guidelines* quanto a aspectos de usabilidade e por último foi um estudo em uma turma de IHC para avaliar a facilidade, utilidade e pretensão de uso das *guidelines*.

4.1 Processo de criação das *guidelines*

Após o estudo na SEFAZ (capítulo 3), surgiu uma discussão sobre a falta de orientações para projetar interfaces de sistemas de rotulagem de dados e reduzir o tempo de entendimento do problema para quem não é tão experiente em ML. Então, foi concebida a ideia de preparar *guidelines* com orientações para auxiliar no projeto desses sistemas. Assim, uma primeira versão das *guidelines* foi desenvolvida a partir dos requisitos refinados obtidos no estudo do capítulo anterior, como também da revisão da literatura e da análise de sistemas da indústria. A partir disso, novas versões foram geradas a

partir um ciclo de melhoria das *guidelines*.

Esse ciclo de melhorias foi composto de inspeções conduzidas com uma especialista em IHC, com mais de 10 anos de experiência, para verificar questões de usabilidade empregadas nas *guidelines*. Após as inspeções, refinamentos foram conduzidos com um especialista de ML, com mais de 5 anos de experiência tanto na academia quanto na indústria. Esses refinamentos buscavam verificar se as *guidelines* eram úteis e se a linguagem utilizada era adequada aos profissionais que podem não ser especialistas em IHC ou ML, mas que precisavam utilizá-las no projeto de seus sistemas. As *guidelines* desenvolvidas são apresentadas na Tabela 4.1.

ID	Guideline
G1	Identificação da tarefa de rotulagem
G2	Contextualização do dado
G3	Rotulagem por instância ou por classe
G4	Exibir o <i>status</i> do progresso da rotulagem
G5	Disponibilizar ajuda e atalhos
G6	Descanso durante a rotulagem

Tabela 4.1 – *Guidelines LabelUX*.

4.2 Descrições das *guidelines*

As próximas subseções apresentam as *guidelines* para sistemas de rotulagem. Estas descrições também estão disponíveis em uma página WEB¹. Esta página, além de disponibilizar as *guidelines*, também foi utilizada nos estudos dos próximos capítulos.

4.2.1 G1 - Identificação da tarefa de rotulagem

Nesta *guideline*, primeiramente, são apresentadas as principais tarefas de rotulagem para cada tipo de dado (texto, imagem, áudio, vídeo), conforme a Tabela 4.2. Em seguida, após a identificação da tarefa de rotulagem que o sistema a ser projetado deve atender, são sugeridas técnicas de rotulagem para cada tipo de tarefa.

¹ <<https://sites.google.com/view/labelux>>

Tipo de dado	Tarefa de rotulagem
Texto	1. Detecção de entidades 2. Classificação de texto 3. Classificação de texto multi-rótulo
Imagem	1. Classificação de imagem 2. Classificação de imagem multi-rótulo 3. Detecção de objetos 4. Segmentação de objetos
Vídeo	1. Classificação de Vídeo 2. Classificação de Vídeo multi-rótulo 3. Detecção de objetos 4. Rastreamento de objetos
Áudio	1. Transcrição de áudio 2. Classificação de áudio

Tabela 4.2 – Tipos de tarefas de ML por tipo de dado (TKACHENKO et al., 2020-2021), (AWS, 2021).

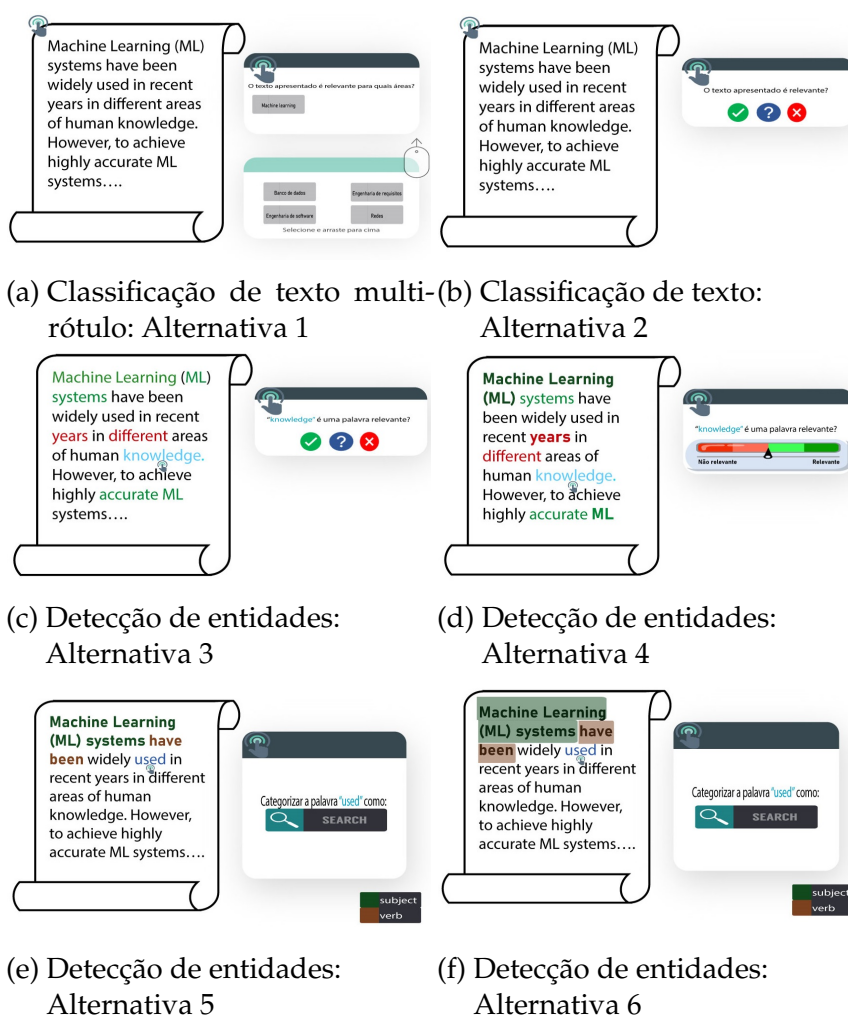


Figura 4.1 – Guideline 1 - Identificação da tarefa de rotulagem. Exemplos de alternativas de design para as tarefas de rotulagem.

Na *guideline*, são apresentadas técnicas de rotulagem para as tarefas do tipo de dado textual que são detecção de entidades (Figura 4.1- alternativas 4 e 6), classificação de texto (Figura 4.1- alternativa 2) e classificação de texto multi-rótulo (Figura 4.1 alternativa 1). As técnicas são soluções para os engenheiros de *software* apresentarem o dado ao usuário e obterem *feedback* do usuário sobre a rotulagem do dado.

4.2.2 G2 - Contextualização do dado

Na *guideline* G2, são apresentadas soluções complementares para auxiliar o usuário a compreender melhor sobre o dado a ser rotulado no momento. O uso de técnicas de visualização de dados, informações descritivas e estatísticas, exemplos e dados semelhantes contribuem para o entendimento do usuário sobre o dado. Como, por exemplo, para o tipo de dado textual, há as sugestões de visualização por tópicos (Figura 4.2a) que ajudam a compreender os assuntos mais abordados e a distribuição de termos frequentes (Figura 4.2b) que indicam em um gráfico as palavras mais frequentes em um conjunto de dados. Além disso, essa contextualização contribui para uma rotulagem de maior qualidade, pois informa o usuário sobre aspectos importantes que auxiliam na compreensão do dado.

Advogado
Caso Direito
Tribunal Escritório
Deveres Juiz Sentença

(a) Visualização por tópicos.

Termos mais frequentes

Advogado	80/160	<div style="width: 50%;"><div style="width: 50%;"></div></div>	👁️
Caso	60/160	<div style="width: 37.5%;"><div style="width: 37.5%;"></div></div>	👁️
Direito	40/160	<div style="width: 25%;"><div style="width: 25%;"></div></div>	👁️
Tribunal	30/160	<div style="width: 18.75%;"><div style="width: 18.75%;"></div></div>	👁️
Escritório	5/160	<div style="width: 3.125%;"><div style="width: 3.125%;"></div></div>	👁️

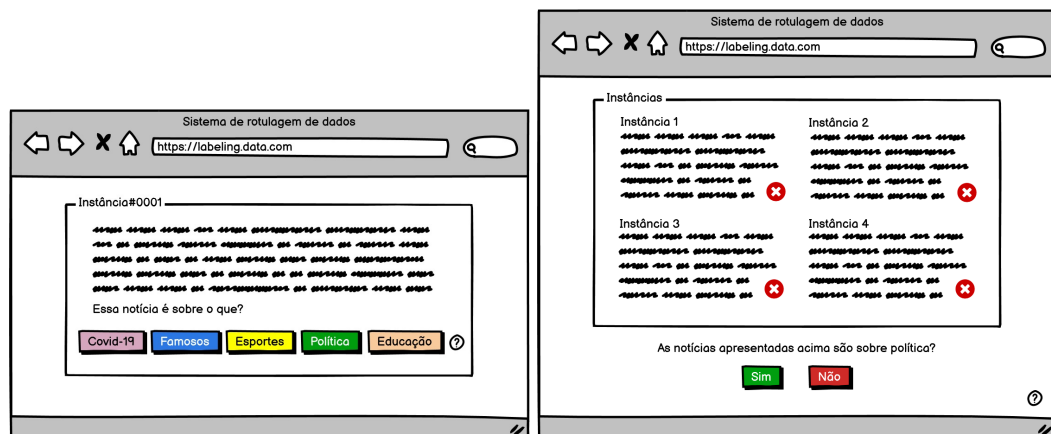
(b) Distribuição de termos.

Figura 4.2 – Soluções de contextualização do dado.

4.2.3 G3 - Instância vs Classe

Nesta *guideline*, são descritas formas para a apresentação dos dados para rotulagem, por instância ou por classe, dependendo do problema de ML. Na rotulagem por instância (Figura 4.3a), o usuário deve rotular um dado (instância) por vez. Contudo, na rotulagem por classe ou categoria (Figura 4.3b), são apresentados os dados que podem estar associados a uma classe específica. Dessa forma, o usuário pode avaliar se cada dado pertence ou não à classe observada, tornando o problema em classificação binária.

Por fim, nesta *guideline*, é destacada que a alternativa de rotulagem por classe apresenta limitações quanto à liberdade do usuário em desfazer erros na rotulagem e ao *design clean* (PASSOS et al., 2021a). Neste tipo de rotulagem, fica mais difícil do usuário desfazer uma rotulagem incorreta, e como a interface contém mais informações, o usuário pode ter a impressão de que o *design* está poluído, o que vai contra o conceito de *design clean* (NIELSEN, 1994). No entanto, o usuário tem a vantagem de estar rotulando mais dados do que se estivesse interagindo apenas com um dado por vez como é feito na rotulagem por instância que, dependendo do problema, pode ser uma atividade cansativa e repetitiva para o usuário.



(a) Exemplo de rotulagem por instância. (b) Exemplo de rotulagem por classe.

Figura 4.3 – Exemplos de rotulagem por instância e por classe.

4.2.4 G4 - Exibir o *status* de andamento da rotulagem

O objetivo desta *guideline* é apresentar ao engenheiro de *software* mecanismos importantes para situar o usuário na tarefa de rotulagem. Soluções de como fornecer informações sobre o *status* da rotulagem para o usuário, como, por exemplo, mensagens sobre o andamento da rotulagem, barras de progresso, tempo de rotulagem e estimativas, como também exibição de instâncias rotuladas e não rotuladas. Dentro desta *guideline*, também é sugerido que haja sessões de rotulagem com metas alcançáveis para não frustrar o usuário. Exemplos de como exibir o *status* da rotulagem pode ser contemplado na Figura 4.4.

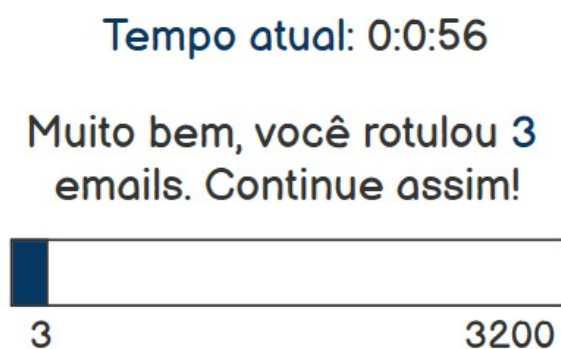


Figura 4.4 – Sugestões da *guideline* G4 de Exibir o *status* de andamento da rotulagem.

4.2.5 G5 - Disponibilizar ajuda e atalhos

Nesta *guideline*, é sugerido que existam atalhos visíveis (Figura 4.5a) e componentes de ajuda para os usuários, como *tooltips* e botões de *ajuda* (Figura 4.5b). O objetivo desta *guideline* é fazer com que o engenheiro de *software* reflita sobre formas do usuário obter ajuda ou informações adicionais sobre a tarefa. Além disso, como empregar o uso de atalhos para otimizar as ações do usuário.



Figura 4.5 – Exemplos de como disponibilizar ajudas e atalhos.

4.2.6 G6 - Descanso durante a rotulagem

Nesta *guideline*, são apresentadas soluções para permitir que o usuário pause a rotulagem a qualquer momento sem cancelar todo trabalho já realizado. Isso se deve ao fato da rotulagem ser uma atividade cansativa, além disso, é sugerido que haja um registro automático a cada interação do usuário (Figura 4.6a) ou uma opção de salvar e sair (Figura 4.6b).



(a) Salvar automaticamente.



(b) Botão de Salvar e Sair.

Figura 4.6 – Exemplos para permitir o descanso durante a rotulagem.

4.3 Rastreabilidade das *guidelines*

Cada uma das *guidelines* foram inspiradas a partir da pesquisa exploratória (Seção 3.2) em sistemas de rotulagem da literatura e indústria, e das lições aprendidas do estudo de caso da SEFAZ (Seção 3.4). A seguir, encontra-se a associação de cada uma das *guidelines* com os resultados até então obtidos:

- (G1) **Identificação da tarefa de rotulagem** - Esta *guideline* foi construída partir da abstração das técnicas abstraídas das interfaces de sistemas de rotulagem dos artigos encontrados na RSL de [Dudley e Kristensson \(2018a\)](#) e dos trabalhos de [Choi et al. \(2019\)](#), [Felix, Dasgupta e Bertini \(2018\)](#), [Poursabzi-Sangdeh et al. \(2016\)](#). As definições de problemas de ML quanto ao tipo de dado foi obtida a partir do sistema de rotulagem da indústria ([AWS, 2021](#)).
- (G2) **Contextualização do dado** - Esta *guideline* foi obtida a partir do requisito [RR3](#) de *Visualizar distribuição e quantidade de títulos de produtos* que utiliza gráficos de barras para visualizar a frequência dos termos ([MUNZNER, 2014](#)). Além disso,

foi observado no trabalho de [Poursabzi-Sangdeh et al. \(2016\)](#) que a visualização por tópicos contribuía para o entendimento de um grupo de instâncias textuais.

- (G3) **Rotulagem por instância ou por classe** - Esta *guideline* foi obtida a partir do requisito [RR5](#) de *Rotular títulos por termos ou grupos de termos*. Além disso, em um sistema de rotulagem da indústria ([PRODIGY, 2021](#)), a rotulagem por instância é justificada por focar na rotulagem de um dado por vez.
- (G4) **Exibir o status de andamento da rotulagem** - Esta *guideline* foi obtida a partir do requisito [RR2](#) de *Visualizar o status de progresso de títulos rotulados*. Além disso, a visualização do progresso da tarefa de rotulagem está relacionado com um dos canais de comunicação proposto por [Dudley e Kristensson \(2018a\)](#). Este canal é a *Task Overview* que indica a visualização do status atual da tarefa de rotulagem.
- (G5) **Disponibilizar ajuda e atalhos** - Esta *guideline* foi obtida a partir da análise de um sistema de rotulagem *open-source* ([TKACHENKO et al., 2020-2021](#)). Além disso, os elementos de ajuda eram um elemento encontrado nos sistemas de rotulagem analisados nos trabalhos encontrados na RSL de [Dudley e Kristensson \(2018a\)](#).
- (G6) **Descanso durante a rotulagem** - Esta *guideline* foi obtida a partir da análise da ausência do elemento do botão de salvar nos protótipos desenvolvidos no estudo anterior na SEFAZ. Além disso, isso também foi observado em um sistema de rotulagem da indústria ([MICROSOFT, 2021](#)). Neste sistema, não é possível parar a rotulagem após iniciar uma sessão com uma quantidade pré-definida e exacerbada de dados.

As Seções a seguir descrevem os estudos realizados para avaliar as *guidelines LabelUX*.

4.4 Estudo de viabilidade

O objetivo do estudo foi analisar o uso das *guidelines* para o projeto de sistemas de rotulagem de dados. O propósito foi avaliar se as *guidelines* auxiliam engenheiros de

software a implantarem soluções nesses sistemas que sejam fáceis e intuitivas, visando apoiar assim o usuário durante a tarefa de rotulagem.

O estudo foi conduzido com 3 participantes. Os participantes escolhidos fazem parte de projetos que envolvem a construção de modelos ML. Nos projetos de cada participante, era necessário rotular uma grande quantidade de dados, pois não tinham uma base previamente rotulada.

O estudo foi executado com as seguintes etapas (para cada participante):

1. **Entendimento do problema de ML do participante:** nesta etapa, o participante foi questionado sobre o seu contexto de uso das técnicas de ML. Além disso, os participantes descreveram a etapa de rotulagem de dados, explicando o tipo de dado a ser rotulado, a quantidade de dados, quem são os possíveis usuários especialistas que podem auxiliar no processo de rotulagem e como está sendo feita a rotulagem de dados no projeto.
2. **Introdução das *guidelines LabelUX*:** nesta etapa, foram apresentadas as *guidelines* e discutidas o que são cada uma delas e para que servem. Além disso, foi disponibilizado para o participante o *link* da página *WEB* das *guidelines* para melhor acompanhamento.
3. **Coleta de impressões sobre cada *guideline*:** nesta etapa, foram realizadas perguntas aos participantes sobre suas impressões de cada *guideline* durante a apresentação da etapa 2.
4. **Elaboração de protótipo:** os participantes foram convidados a construir um protótipo simples (não navegável) de um sistema de rotulagem de dados. Este protótipo deve atender ao problema de rotulagem dos participantes e pode ser construído em papel, em uma ferramenta de prototipagem ou ferramenta de edição de imagens. No momento da prototipagem, os participantes poderiam acessar as *guidelines* pelo site, assim como tirar dúvidas a qualquer momento.
5. **Perguntas finais:** ao final da atividade de criação do protótipo, foram realizadas perguntas aos participantes sobre a etapa de elaboração de protótipo sem navegação. As perguntas são: Quais *guidelines* foram mais úteis? Quais você considerou

por primeiro? Quais não ficaram claras e que podiam melhorar? Você teve alguma dúvida com relação à ordem das *guidelines*?

O estudo foi realizado com 3 profissionais com mais de 3 anos de experiência com projetos de ML (projeto e desenvolvimento). Além disso, os participantes estão atuando em projetos supervisionados, com necessidade de rotulagem de dados.

4.4.1 Resultados e discussão

Foram executadas as etapas do estudo com cada um dos participantes, de forma a não haver viés causado por comunicação entre os mesmos. O tempo do estudo foi de, aproximadamente, duas horas e 30 minutos com cada participante.

4.4.1.1 Entendimento do problema de ML dos participantes

Na primeira etapa do estudo, cada um dos participantes descreveu o problema de ML que buscava resolver atualmente. O problema do participante P1 consistia em classificar *reviews* de lojas de aplicativos através de análise de sentimento.

Quanto ao problema da participante P2, os usuários deveriam classificar sentenças em útil ou não útil dentro de um determinado contexto de um projeto de desenvolvimento de *software*. O problema da participante P3 consistia em classificar cenários de especificação de *softwares* em categorias de requisitos de sistemas.

4.4.1.2 Introdução das *guidelines LabelUX*

Após o entendimento do problema de ML de cada participante, foi feita uma breve introdução sobre as *guidelines LabelUX* (segunda etapa). Em seguida, antes de iniciar a apresentação de cada *guideline*, os participantes deveriam considerar um cenário em que muitos usuários especialistas os ajudariam na tarefa de rotular os dados referentes aos problemas que cada um tratava. Assim, os participantes teriam que começar a pensar em um projeto de sistema de rotulagem de dados.

4.4.2 Coleta de impressões sobre cada *guideline*

Na terceira etapa do estudo, conforme cada *guideline* era apresentada, os participantes respondiam às perguntas propostas. Por questões de brevidade, a seguir, são apresentadas as perguntas e trechos das respostas sobre cada *guideline*.

(G1) Identificação da tarefa de rotulagem

a) **Qual sua opinião sobre as alternativas apresentadas?**

P1: *"Essa opção de dúvida é interessante para que depois você possa analisar esses casos... Sem ter essa opção de deixar em aberto, talvez acabe gerando rotulagens incorretas... Poderia ter também um espaço para o usuário dar feedback, caso não tenha entendido muito bem uma classe..."*

"Nas abordagens de detecção de entidades, é interessante esse destaque dos termos não só com cores, colocando esse background..."

P2: *"Elas são de certa forma evolutivas porque tentam sanar deficiências presentes em uma outra abordagem... As escalas podem fazer o usuário tender à neutralidade ou ao ponto central..."*

"Seria bom ter um campo opcional para o usuário colocar uma justificativa da rotulagem de um certo dado..."

P3: *"As técnicas de rotulagem poderiam ter a opção de, após selecionar a palavra, clicar com o botão direito para classificar como relevante e com o botão esquerdo para classificar como irrelevante. Assim, o usuário não precisaria ficar mexendo o mouse para um lado e para o outro..."*

"...Na tarefa de classificação multi-rótulo, as classes poderiam ser mostradas diretamente na tela, caso tenham um número limitado, algo como no máximo 10 classes. Isso evitaria a busca a cada palavra rotulada..."

b) **Você combinaria várias dessas técnicas para resolver o teu problema?**

P1: *"Por conta do meu problema, teria que combinar sim..."*

P2: *"Sim, pois eu vejo pontos fortes em cada uma delas..."*

P3: *"Sim, e, provavelmente, faria adaptações..."*

O participante P1 levantou muitos pontos pertinentes ao problema dele na *guideline* G1, mas pontuou que as alternativas apresentadas interessavam às tarefas descritas. O participante, a princípio, não havia entendido que combinações entre as alternativas poderiam ser feitas. Isto aconteceu, porque esta possibilidade não estava descrita nas *guidelines*.

A participante P2 concluiu que as alternativas são evolutivas e que algumas seriam complexas para o usuário iniciante. A participante P3 sugeriu outras formas de interação para uma mesma técnica de rotulagem, considerando um usuário mais experiente. Além disso, fez uma sugestão para a tarefa de classificação multi-rótulo, considerando uma quantidade pequena de classes (até 10), seria possível apresentar todas as classes possíveis na interface sem a necessidade de um componente de busca.

Todos os participantes concordaram que além da rotulagem, seria interessante que o usuário pudesse justificar sua escolha ou incluir novas classes. Além disso, todos consideraram como mais úteis as técnicas que continham opções de "não rotular" através dos botões de dúvida (Figura 4.1 - Alternativa 2), como também o destaque no *background* das palavras relevantes (Figura 4.1 - Alternativa 6).

(G2) Contextualização do dado

a) **Nesta *guideline*, que sugestões você usaria ou que você faria de diferente se não usasse?**

P1: *"A visualização de distribuição por barras seria mais interessante... A visualização por tópicos não é considerada uma ferramenta analítica."*

P2: *"Utilizaria uma combinação da visualização por tópicos e depois complementar com a exibição da distribuição de termos frequentes... A visualização por tópicos me parece mais agradável em um primeiro contato."*

P3: *"...Eu manteria os termos mais frequentemente utilizados na parte superior da tela para serem selecionados mais facilmente... prefiro a solução com números, não gosto muito da visualização por tópicos."*

Na *guideline* G2, os participantes tiveram opiniões divergentes. Os participantes P1 e P3 não se agradaram da proposta de visualização por tópicos (Figura 4.2a), pois

consideravam não ser uma boa ferramenta para análise e visualização, porque não proporciona uma visão precisa das palavras mais importantes nos dados. Quanto ao uso da distribuição de termos frequentes (Figura 4.2b), todos concordaram que o utilizariam, porém, a participante P2 considerava que a visualização por tópicos era mais agradável para o usuário.

(G3) Rotulagem por instância ou por classe

a) **Você concorda com esta *guideline*? No seu problema de ML, qual abordagem você adotaria?**

P1: *"Nunca pensei assim em colocar por classe, porque considero que por instância seja o mais natural, mas do ponto de vista do usuário, ele vai se sentir mais produtivo."*

P2: *"A rotulagem sendo feita parcialmente, evita a ocorrência de erros e melhora a visão do usuário quanto ao que ele está fazendo... Utilizaria a rotulagem por classes."*

P3: *"Gostei das duas perspectivas de classificação. No meu problema eu adotaria a de instâncias, porque as classes são definidas durante o processo de rotulagem"*

Todos os participantes relataram que não haviam pensado anteriormente em uma estratégia de rotulagem por classe, onde cada instância é avaliada se pertence ou não a uma classe, ao invés de determinar qual é a classe específica de uma instância. Eles somente contemplaram esta possibilidade após a leitura da *guideline* G3.

(G4) Exibir o status de andamento da rotulagem

a) **Nesta *guideline*, qual sua opinião sobre cada uma dessas sugestões?**

P1: *"É interessante para o usuário ver o progresso dele... Eu não tinha pensado muito nesse sentido de controlar por sessão ou de ter metas de rotulagens... Esse histórico de rotulados e não rotulados não sei se é tão necessário."*

P2: *"Eu gostei muito das sugestões, porém eu não gostei tanto da exibição de instâncias rotuladas e não rotuladas... Seria interessante de ser mantida em uma outra tela, separada, como um histórico."*

P3: *"...Eu gostei da barra de progresso, porquê tem a informação passada de maneira visual."*

Os três participantes concordaram com o uso de todas as sugestões, com exceção da exibição de instâncias rotuladas e não rotuladas que não agradaram os participantes P1 e P2. Foi relatado que esta sugestão não ajudaria o usuário no momento da tarefa de rotulagem, mas P2 indicou a possibilidade de uso como visão de histórico em outra tela. A participante P3 mostrou maior interesse pelo uso da sugestão de barra de progresso devido à visualização da informação.

(G5) Disponibilizar ajuda e atalhos

a) **Qual sua opinião sobre esta guideline?**

P1: *"É importante deixar até visível esses atalhos ao invés de tooltips... É interessante também deixar claro sobre as classes para não embaralhar os conceitos"*

P2: *"É fundamental para que o usuário não tenha muitas dúvidas... Atalhos podem facilitar muito a jornada do usuário que está fazendo a mesma tarefa a mais tempo."*

P3: *"...Importante para que o usuário tenha ajuda sem precisar sair da tela de classificação."*

Todos os participantes concordaram que a *guideline* G5 é muito importante, tanto para os usuários iniciantes, quanto para os experientes no sistema. A participante P3 considerou importante o uso desta *guideline* porque o usuário pode obter ajuda sem precisar sair da tela da rotulagem. Além disso, um ponto importante levantado pelo participante P1 é que a ajuda deve estar presente também sobre os significados das classes e sobre a tarefa da rotulagem.

(G6) Descanso durante a rotulagem

a) **Qual sua opinião sobre esta guideline?**

P1: *"Essa guideline é bem interessante... O sistema deve permitir que ele (o usuário) faça uma pausa e tenha o feedback visual de que tudo foi salvo."*

P2: *"Gosto das duas opções apresentadas e penso que as duas são para tipos diferentes de usuários..."*

P3: *"Muito importante ter opção de salvar o andamento do trabalho a qualquer momento."*

Sobre a última *guideline*, os participantes relataram não ter pensado anteriormente sobre esses aspectos relacionados à pausa e o salvar dos dados rotulados. Porém, após ler as *guidelines*, eles as consideraram indispensáveis para um projeto de sistema de rotulagem. A participante P2 relatou que utilizaria as duas sugestões apresentadas, pois contemplam tipos diferentes de usuários. Esses usuários podem ser aqueles acostumados com ferramentas que já salvam automaticamente, como também os usuários que precisam interagir com um botão de salvar.

4.4.2.1 Elaboração de protótipo

Na quarta etapa do estudo, os participantes fizeram um protótipo de baixa fidelidade de um sistema de rotulagem de dados com auxílio das *guidelines*. Foi informado aos participantes que estes protótipos deveriam atender ao problema que eles trabalhavam.

Durante a construção do protótipo, os participantes não realizaram perguntas pertinentes às *guidelines*. O tempo total para a construção dos protótipos, para ambos participantes, foi de 50 minutos aproximadamente. O participante P1 optou por uma ferramenta de edição, como pode ser visto na Figura 4.7. Os protótipos das participantes P2 e P3 foram desenvolvidos em uma ferramenta de prototipação, conforme as Figuras 4.8 e 4.9.

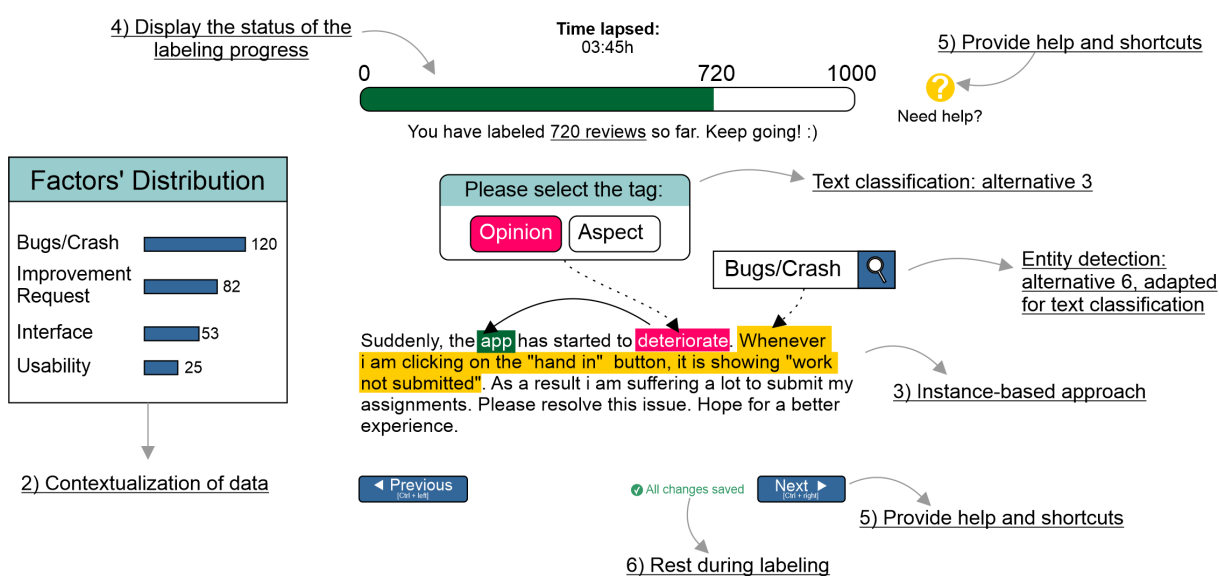


Figura 4.7 – Protótipo construído pelo participante P1.

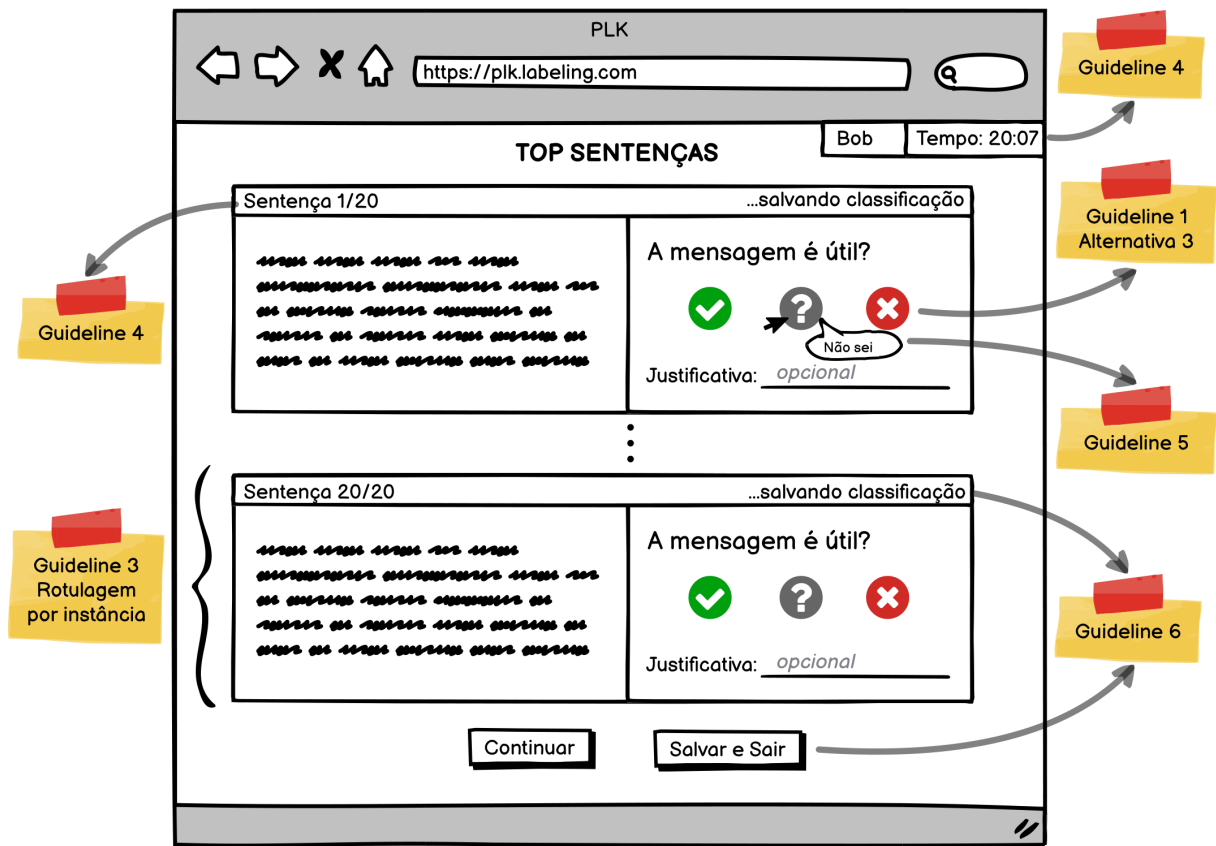


Figura 4.8 – Protótipo construído pela participante P2.

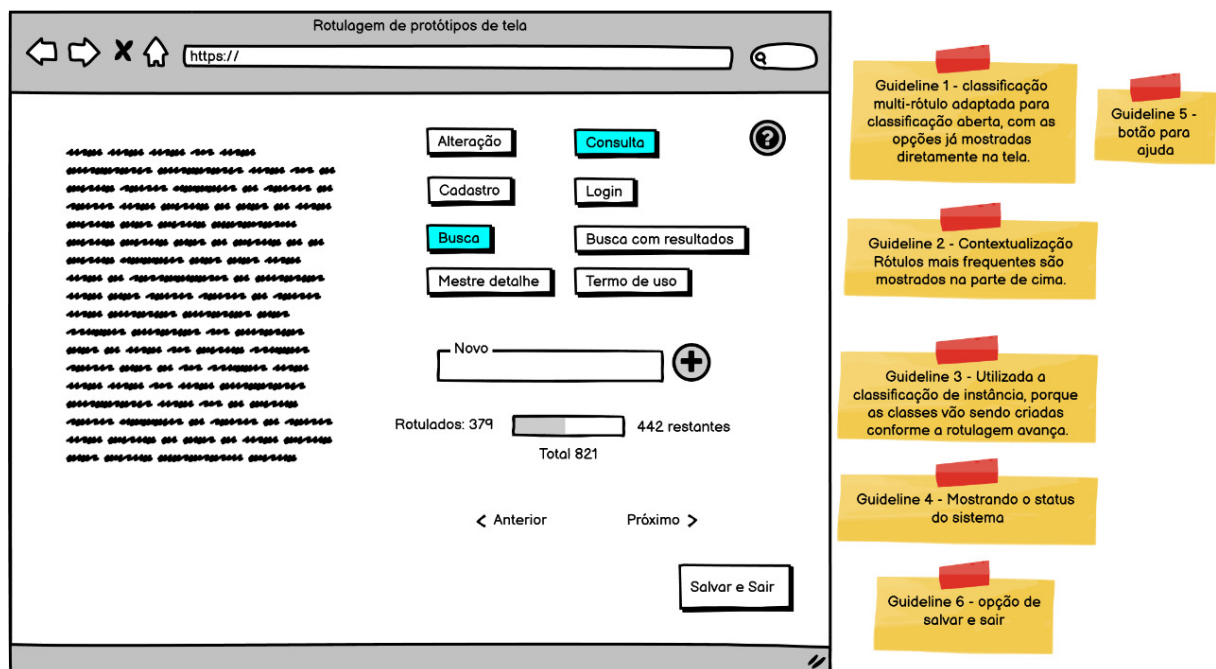


Figura 4.9 – Protótipo construído pela participante P3.

É possível observar nos protótipos desenvolvidos que todos os participantes utilizaram todas as *guidelines* propostas, com exceção da participante P2 que optou por não utilizar a *guideline* G2. Além disso, todos os participantes utilizaram a alternativa de rotulagem por instância em relação à *guideline* G3.

A participante P3 decidiu por utilizar uma técnica de rotulagem diferente para a tarefa de classificação multi-rótulo, como também adaptou a sugestão da *guideline* G2 para apresentar os rótulos mais frequentes primeiro. Estas diferenças foram anotadas para serem aderidas nas próximas versões das *guidelines*.

4.4.2.2 Perguntas finais

Na quinta etapa do estudo, os participantes responderam às perguntas finais referente ao uso das *guidelines*. Novamente, por questões de brevidade, a seguir são apresentadas as perguntas e trechos das respostas.

(1) **Quais *guidelines* foram mais úteis? Quais você considerou por primeiro?**

P1: *"Comecei pela primeira guideline da identificação da tarefa da rotulagem... Pensei nela primeiro porque ela é o core da rotulagem..."*

"...Inclusive a questão dos atalhos, eu não estava nem lembrando, mas a guideline me fez lembrar desse ponto..."

"...a parte do help foi o que eu coloquei por último."

P2: *"Eu considerei primeiro as guidelines 6, 1 e 4, porque elas foram as que eu considerei as mais importantes de estarem presentes na interface do protótipo..."*

P3: *"Considerei primeiro a guideline 1 e em seguida, a guideline 3..."*

O participante P1 relatou que considerou mais importante a *guideline* G1 porque era a atividade principal do sistema. Em seguida, considerou as *guidelines* G2, G3 e G4. Por fim, o participante informou que as *guidelines* G5 e G6 o lembraram de aspectos que no momento ele havia esquecido e, se não fossem por essas *guidelines*, ele não teria colocado atalhos e as alternativas de pausa para descanso.

A participante P2, entretanto, teve outra perspectiva, atentando-se mais para a experiência do usuário durante a tarefa de rotulagem. Por isso, utilizou por primeiro as *guidelines* G6, G1 e G4, pois as considerava como as mais importantes e úteis para o usuário durante a realização da rotulagem.

A participante P3, assim como o P1, considerou por primeiro a *guideline* G1 e depois como seria o formato da rotulagem, se seria por instância ou por classe, conforme a *guideline* G3. As demais *guidelines* foram consultadas depois, consideradas de igual importância após as *guidelines* G1 e G3.

(2) Quais *guidelines* não ficaram claras e que podiam melhorar?

P1: *"Não foi exatamente nas guidelines em si... Eu acabei invertendo a parte de detecção de identidade e a classificação de texto..."*

"..Sem a parte visual, sem as figuras, eu demoraria bem mais..."

P2: *"A guideline que menos ficou clara para mim foi o 2, pois eu ainda preciso entender melhor como aplicar a distribuição de termos frequentes de forma agradável assim como a visualização por tópicos"*

P3: *"A guideline de contextualização dos dados poderia ter mais exemplos."*

Para o participante P1, as *guidelines* foram claras. Além disso, optou por combinar técnicas, mesmo não estando explícita essa possibilidade nas *guidelines*. Estas técnicas, com as figuras, foram os que mais o ajudaram na hora de construir o protótipo.

Para a participante P2, não ficou claro a aplicação das soluções propostas na *guideline* G2, quanto a disposição delas na interface. Neste sentido, seria interessante incluir as soluções propostas em conjunto com os outros elementos de interface. A participante P3, assim como a P2, considerou que a *guideline* G2 poderia ter mais exemplos para deixar mais claro o seu uso no sistema.

Por fim, as considerações finais dos participantes foram: *"Não tive dificuldade, ajudou bastante, principalmente para ver o que precisa..."*, *"...Me fizeram repensar em conceitos que eu tinha antes, mas voltadas para sistemas de rotulagem"* e *"Foi útil consultá-las enquanto eu elaborava meu protótipo. Alguns aspectos, como a ajuda e atalhos, tendem a ser esquecidos,*

então consultar o guideline evitou isso.”. Os participantes, então, demonstraram estar satisfeitos com o apoio das *guidelines* para projetar sistemas de rotulagem.

O estudo evidenciou que as *guidelines* apoiam no projeto de sistemas de rotulagem de dados, visando uma interface com soluções para uma melhor experiência do usuário durante a atividade de rotulagem. Contudo, ainda é necessário realizar alguns ajustes para ter maior aderência e clareza de todas as *guidelines*, tais como incluir opção de justificativa no *guideline* G1 e sugerir mais soluções para a contextualização do dado (G2).

Os resultados indicaram que as *guidelines* foram úteis para os participantes, apresentando elementos essenciais da interface dos sistemas de rotulagem. As conclusões do estudo indicaram que as *guidelines* LabelUX ajudam os desenvolvedores a entender as necessidades do usuário durante o desenvolvimento desse tipo de sistema. Durante o uso, os três engenheiros de *software* também identificaram problemas nas *guidelines*. Este estudo foi publicado no Simpósio Brasileiro de Qualidade de Software (SBQS) 2021 (PASSOS et al., 2021b)

4.5 Estudo de avaliação de apoio à Usabilidade

Este estudo apresenta uma análise, com relação aos aspectos de usabilidade, de *guidelines* que apoiam o projeto de interfaces de sistemas de rotulagem de dados. Para isso, dois pesquisadores analisaram cada *guideline*, visando encontrar relações sobre o uso destas e as recomendações de usabilidade, através das heurísticas de Nielsen. Um terceiro pesquisador, especialista em IHC, revisou e validou as relações entre as *guidelines* e as heurísticas.

4.5.1 G1 - Identificação da tarefa de rotulagem

Esta *guideline* fornece alternativas de *design* para a tarefa de rotulagem de dados. Considerando que o tipo de dado é textual, a *guideline* é dividida em três tarefas de rotulagem: detecção de entidades, classificação de texto e classificação de texto multi-rótulo.

Conforme a Figura 4.1, cada tarefa de rotulagem apresenta alternativas de *design*, compostas por elementos de interface ligados à apresentação do dado e elementos de interface ligados a solicitação de *feedback* da rotulagem de dados.

Com o uso desta *guideline*, o *designer* de *software* é apresentado às alternativas de *design* de interface que oferecem múltiplas soluções. Dessa forma, o *designer* de *software* pode utilizar a solução mais adequado ao seu problema de ML. Na alternativa 3 da detecção de entidades (Figura 4.1 - (c)) e na alternativa 2 da classificação de texto (Figura 4.1 - (b)), são apresentados elementos na interface que fornecem uma liberdade ao usuário de incerteza sobre o dado, ou seja, quando o usuário não tem segurança quanto a rotulagem do dado. Além disso, outra solução que envolve a liberdade do usuário durante a rotulagem dos dados é a alternativa 4 da detecção de entidades (Figura 4.1 - (d)), onde é apresentado um elemento de interface em formato de escala, em que o usuário pode escalar em 4 graus a relevância de um dado.

Nas alternativas 5 e 6 de detecção de entidades (Figuras 4.1 - (e) e 4.1 - (f)), são apresentadas estratégias com destaques dos termos na cor de fonte ou de *background*, correspondendo a categoria que o usuário classificou o dado. Durante a tarefa de rotulagem, o usuário pode reconhecer os termos rotulados com o destaque deste, além disso, com as cores de fonte e *background* dos termos, o usuário pode lembrar qual a categoria esses dados pertencem, sem precisar consultar uma lista de categorias, por exemplo.

Portanto, o uso desta *guideline* pode ajudar o *designer* de *software* a atender duas heurísticas de Nielsen: Controle e liberdade do usuário e Reconhecer ao invés de relembrar.

4.5.2 G2 - Contextualização do dado

Nesta *guideline*, são apresentados componentes de interface que podem auxiliar o usuário a entender mais sobre os dados. A Figura 4.2 mostra, como exemplo, a frequência dos termos entre os dados e visualização por tópicos.

Na Figura 4.2 - (A), são apresentados ao desenvolvedor a distribuição de termos

frequentes para auxiliar o usuário durante a tarefa de rotulagem de dados textuais (FELIX; DASGUPTA; BERTINI, 2018). A frequência dos termos pode fornecer *insights* ao usuário, por exemplo, os termos mais frequentes, podem representar as palavras mais relevantes e as menos frequentes as palavras com menos relevância.

Na Figura 4.2 - (B), são apresentados ao desenvolvedor a visualização por tópicos para auxiliar o usuário na contextualização dos dados. Este componente é utilizado para contextualizar o usuário sobre os dados a serem rotulados, clicando nos termos da nuvem de palavras o usuário pode navegar até os trechos do texto original para entender o contexto e realizar rotulagens de alta qualidade (POURSABZI-SANGDEH et al., 2016).

As informações fornecidas pelos dois componentes da Figura 4.2 ajudam o usuário a ter noção sobre os dados já rotulados e o andamento geral da atividade. A partir disso, acredita-se que esta *guideline* está em conformidade à heurística de Visibilidade do *status* do sistema, por fornecer informações sobre o total de dados rotulados, as palavras mais frequentes e a visualização de tópicos.

4.5.3 G3 - Rotulagem por instância ou por classe

Esta *guideline* fornece duas abordagens sobre a tarefa de rotulagem. Conforme a Figura 4.3, a primeira é a rotulagem por instância que possibilita que o usuário consiga analisar e rotular o dado individualmente e desfazer erros de rotulagem. Na abordagem de rotulagem por classe, o usuário rotula um grupo de dados, analisando o conjunto inteiro. Além disso, nas duas abordagens, o usuário pode decidir se deseja pular um dado a ser rotulado, seja por dúvida ou insegurança sobre a rotulagem.

A rotulagem por classe torna a atividade mais produtiva pelo fato da análise ser feita sobre o conjunto de dados. Porém, a rotulagem por instância possibilita o usuário desfazer erros de rotulagem, voltando aos dados, de forma individual, e corrigindo os erros.

A análise feita sobre o conjunto de dados produz uma interface mais *clean* e minimalista por apresentar menos botões para o usuário rotular os dados, por exemplo,

caso um usuário precisasse rotular uma quantidade X de dados e a rotulagem fosse por instância, seria necessário no mínimo X cliques nos botões de rotulagem. Com a rotulagem por classe, o usuário precisa de apenas um clique para rotular todo o conjunto, caso exista algum dado que não seja de tal classe, o usuário pode retirar esse dado da classe. Este caso indica a conformidade com a heurística Estética e *design* minimalista.

Na rotulagem por classe, o usuário consegue voltar e corrigir rotulagens feitas, isso possibilita associar esta *guideline* a heurística Ajudar os usuários a reconhecer, diagnosticar e corrigir erros.

4.5.4 G4 - Exibir o *status* do progresso da rotulagem

Um dos motivos para analisar a interação do usuário com sistemas de rotulagem é proveniente da sensação de cansaço e exaustão que o usuário pode ter durante a atividade. Neste sentido, esta *guideline* apresenta aos *designers* de *software* opções de componentes de interface para mostrar ao usuário informações sobre o seu andamento, como tempo total da atividade de rotulagem, quantidade de dados rotulados, barra de progresso, entre outros.

Nesta *guideline*, o progresso da rotulagem também pode ser feito por mensagens motivadoras ao usuário como: "Você já rotulou 100 palavras, continue assim!". Na Figura 4.4, são apresentados duas opções para esta *guideline*.

Esta *guideline* apoia a conformidade à heurística Visibilidade do *status* do sistema, pois fornece uma visão aos *designers* de *software* de como informar ao usuário o andamento da atividade de rotulagem.

4.5.5 G5 - Disponibilizar ajuda e atalhos

Objetivo desta *guideline* é apresentar aos *designers* de *software* alternativas para o usuário solicitar ajuda durante a rotulagem. Além disso, deixar claro ao usuário outras opções de interação, como atalhos no teclado.

Esta *guideline* segue os princípios das heurísticas Flexibilidade e eficiência de uso e Ajuda e documentação. A Figura 4.5 - (A) apresenta uma possibilidade de como apresentar ao usuário a flexibilidade de uso, este exemplo o usuário pode confirmar uma rotulagem clicando no botão "submit" ou através do teclado nas teclas "Ctrl + Enter". O botão de ajuda apresentado na Figura 4.5 - (B) fornece uma possibilidade ao *designer* de *software* de mostrar ao usuário a opção de ajuda sobre a interface ou informações sobre a tarefa de rotulagem, quais os dados que o usuário está rotulando entre outros.

4.5.6 G6 - Descanso durante a rotulagem

Com o foco na qualidade da rotulagem, esta *guideline* fornece aos *designers* de *software* alternativas de componentes de interface para que o usuário consiga pausar a atividade de rotulagem, informado pelo sistema o *status* da rotulagem, se foram salvas ou não (Figura 4.6 - (A)). Além de apresentar ao usuário a opção de sair da atividade de rotulagem (Figura 4.6 - (B)). A Figura 4.6 apresenta alguns componentes que podem ser utilizados pelo *designer* de *software* ao seguir esta *guideline*.

Esta *guideline* auxilia o *designer* de *software* a informar ao usuário o *status* da rotulagem, proporcionando aos usuários *feedback* sobre suas ações realizadas no sistema. Por isso, associamos esta *guideline* com a heurística Visibilidade do status do sistema. Além disso, devido à possibilidade do usuário cancelar a qualquer momento a atividade de rotulagem, esta *guideline* também foi associada à heurística controle e liberdade do usuário.

4.5.7 Considerações finais

As *guidelines* ajudaram os *designers* de *software* a projetar as interfaces de sistemas de rotulagem de dados, fornecendo uma visão sobre aspectos da própria atividade de rotulagem de dados e aspectos de usabilidade presentes nas heurísticas de Nielsen.

Na Figura 4.10, são apresentados as relações entre as *guidelines* e as heurísticas de

Nielsen (NIELSEN, 1994). A heurística H1 - Visibilidade do *status* do sistema é a mais associada nas *guidelines*, as heurísticas H2 - Compatibilidade do sistema com o mundo real, H4 - Consistência e padrões e a H5 - Prevenção de erros não foram vinculadas à nenhuma *guideline*.



Figura 4.10 – Análise da conformidade das *Guidelines* às heurísticas de Nielsen.

A cobertura da *guideline* G1 - Identificação da tarefa de rotulagem abrange duas heurísticas, H3 - Controle e liberdade e H6 - Reconhecer ao invés de lembrar. Nesta *guideline*, é importante fornecer visualizações úteis aos usuários para que durante a análise dos dados este tenha *insights* sobre os dados. Além disso, o usuário pode precisar explorar mais sobre os dados para decidir a rotulagem do dado.

A *guideline* G2 - Contextualização do dado foi associada à heurística H1 - Visibilidade do *status* do sistema, pois o objetivo desta é apresentar ao *designer* de *software* soluções de interface que proporcionem ao usuário informações sobre o contexto dos dados no sistema de rotulagem, por exemplo, a frequência dos termos. A frequência dos termos, além de mostrar ao usuário os termos e o número de vezes que se repetem, também fornece uma visão geral da tarefa de rotulagem sobre os possíveis termos mais relevantes.

Cobrindo as heurísticas H8 - Estética e *design* minimalista e H9 - Ajudar os usuários a reconhecer, diagnosticar e corrigir erros, a *guideline* G3 - Rotulagem por instância ou por classe, fornece aos *designers* de *software* duas opções de como projetar a interface. O *designer* de *software* deve escolher a estratégia de rotulagem que mais se adapta ao problema de rotulagem. Por um lado, a estratégia de fazer o usuário rotular

os dados individualmente (por instância), fornece mais possibilidades ao usuário, por exemplo, desfazer rotulagens já feitas. Por outro lado, a rotulagem por classe fornece mais agilidade ao usuário por proporcionar uma categorização em lote, apresentando ao usuário apenas componentes essenciais na interface que a deixa mais limpa, com menos botões.

A *guideline* G4 - Exibir o *status* do progresso da rotulagem é utilizada para fornecer meios ao *designer* de *software* de informar ao usuário o andamento da atividade de rotulagem, através de componentes de progresso, mostrando quantas rotulagens foram feitas, o tempo gasto na atividade, além de mensagens motivadoras para estimular o usuário durante a atividade. Neste caso, esta *guideline* foi associada à heurística H1 - Visibilidade do *status* do sistema, pois proporciona ao usuário visões gerais sobre a atividade de rotulagem.

Com a *guideline* G5 - Disponibilizar ajuda e atalhos, são apresentados aos *designers* de *software* alternativas para o usuário solicitar ajuda para entender mais sobre a própria interface ou sobre os dados que estão sendo rotulados. Além disso, essa *guideline* também apresenta uma alternativa quanto a flexibilidade de uso do usuário, para que este tenha mais possibilidade de interação com a interface. Portanto, esta *guideline* foi associada às heurísticas H7 - Flexibilidade e eficiência de uso, e H10 - Ajuda e documentação.

Finalmente, a *guideline* G6 - Descanso durante a rotulagem sugere aos *designers* de *software* soluções para deixar o usuário ciente sobre as ações do sistema de rotulagem, como o *status* do sistema, se a rotulagem está sendo processada ou se já foi salva. Também, a *guideline* norteia o *designer* de *software* para informar ao usuário que a atividade de rotulagem pode ser pausada e que o processo da atividade será salvo. Por isso, esta *guideline* foi associada às heurísticas H1 - Visibilidade do *status* do sistema e H3 - Controle e liberdade do usuário.

As *guidelines* servem como guia para o *designers* de *software* em projetos de sistemas de rotulagem de dados. Além disso, as *guidelines* apoiam na construção de interfaces com maior qualidade, visto que nesses contextos os *designers* de *software* podem não ser especialistas em IHC ou ter pouco domínio na área.

Os resultados mostraram a conformidade das *guidelines* com as heurísticas de Nielsen. Dois pesquisadores inspecionaram as *guidelines* individualmente, construindo a relação entre as *guidelines* e as heurísticas. Em seguida, um especialista em IHC julgou as relações criadas. Os resultados mostraram que as *guidelines* podem ser úteis para desenvolvedores aplicarem sete das dez heurísticas no contexto da interface de sistemas de rotulagem. As heurísticas que não foram associadas são compatibilidade do sistema com o mundo real, consistência e padrões e prevenção de erros.

Os resultados desta Seção foram registrados em um artigo com o título *Applying Usability Heuristics in the Context of Data Labeling Systems*. Este artigo foi submetido e aceito no Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais (IHC 2021) na Trilha de Ideias Inovadoras e Resultados Emergentes (VIANA et al., 2021).

4.6 Estudo *In Vitro*

O estudo *In Vitro* teve como objetivo avaliar a facilidade, utilidade e pretensão de uso das *guidelines* durante a construção de um projeto de interface de um sistema de rotulagem de dados na perspectiva de alunos de graduação. Além disso, este estudo permitiu avaliar estas percepções conforme o conhecimento prévio informado pelos participantes sobre ML e sistemas de rotulagem. O estudo foi conduzido em uma turma de Interação Humano-Computador da Universidade Federal do Amazonas - UFAM. Esta turma continha 39 alunos, sendo 14 alunos de pós-graduação em informática e 25 alunos de graduação em ciência da computação. Os 39 participantes formaram equipes entre si de forma voluntária (com quantidade máxima de 6 alunos), totalizando nove equipes.

Dos 39 participantes, 6 possuíam conhecimento e experiência em ML e em sistemas de rotulagem (15,4%), 22 não possuíam nenhum conhecimento ou experiência em ML (56,4%), e os demais (28,2%) possuíam algum conhecimento ou experiência em ML. A maioria dos alunos nunca tinha usado a ferramenta de prototipagem *Balsamiq* (84,61%) e 27 alunos nunca tinham desenvolvidos protótipos navegáveis (69,23%). Após

a análise do perfil dos participantes, as equipes foram agrupadas em 3 perfis:

- A. Possui integrantes com conhecimento e experiência em ML e em sistemas de rotulagem;
- B. Possui integrantes com conhecimento e experiência apenas em ML;
- C. Não possui integrantes com conhecimento e experiência em ML e nem sistemas de rotulagem.

Das nove equipes formadas, quatro eram do perfil A (equipes 1, 2, 3 e 5), duas eram do perfil B (equipes 4 e 9) e três eram do perfil C (equipes 6, 7 e 8). Uma das equipes do perfil A (equipe 3) fez o protótipo com uma base de dados de notícias devido ao trabalho de mestrado de uma das integrantes. Por fim, com essa divisão das equipes, foi possível avaliar o entendimento sobre as *guidelines* para cada perfil de equipe.

4.6.1 Construção dos protótipos

Todas as nove equipes desenvolveram os protótipos na ferramenta *Balsamiq* e definiram os problemas de ML conforme descritos na Tabela 4.3. Cada problema foi categorizado pelos pesquisadores em uma tarefa de ML conforme a Tabela 4.2 da *guideline* 1.

Todos os protótipos navegáveis produzidos pelas equipes encontram-se em um repositório *online*². Cada equipe teve uma semana para construir os protótipos. O total de telas produzidas e as *guidelines* utilizadas pelas equipes de cada perfil pode ser observado na Tabela 4.4.

Ao analisar os protótipos produzidos pelas equipes de perfil A e B, foi possível notar que foram os que mais utilizaram as sugestões das *guidelines*. As próximas subseções detalham os protótipos produzidos por cada perfil. Para cada equipe, são descritos quais *guidelines* foram utilizadas e como cada *guideline* foram aplicadas nos protótipos.

² <<https://doi.org/10.6084/m9.figshare.16543812.v1>>

Equipe	Problema	Tarefa de ML
1	Identificação de datas comemorativas em <i>tweets</i>	Detecção de entidades
2	Classificar <i>tweets</i> de acordo com categorias de conteúdo	Classificação multi-rótulo
3	Detecção de sexismo em comentários de portais de notícia	Classificação binária
4	Classificar os <i>tweets</i> que falam sobre o Brasil em positivos ou negativos	Classificação binária
5	Detecção de conteúdo publicitário no <i>tweet</i>	Detecção de entidades
6	Identificar os tipos de saudações utilizados por cada país em <i>tweets</i>	Detecção de entidades
7	Classificação de <i>tweets</i> de ataques aos governos	Classificação binária
8	Identificação de <i>tweets</i> de propaganda/divulgação de terceiros	Detecção de entidades
9	Classificação de <i>tweets</i> no polo político	Classificação binária

Tabela 4.3 – Problema de ML definidos pelas equipes.

Perfil	Equipe	Quantidade de telas produzidas	Guidelines utilizadas
A	1	11	G1, G2, G3, G4, G5 e G6
	2	23	G1, G3, G4 e G5
	3	16	G1, G2, G3, G4 e G6
	5	14	G1, G3, G4, G5 e G6
B	4	49	G1, G3, G4, G5 e G6
	9	11	G1, G3, G4, G5 e G6
C	6	7	Não utilizaram as <i>guidelines</i>
	7	19	G1, G3 e G5
	8	22	G1 e G3

Tabela 4.4 – Total de telas produzidas e *guidelines* utilizadas por cada equipe de cada perfil.

4.6.2 Protótipos de perfil A

As equipes de perfil A possuíam integrantes com conhecimento e experiência em ML, e que também utilizaram ou projetaram sistemas de rotulagem de dados. Assim, era esperado que os protótipos fossem os mais completos por conta do conhecimento e experiência dos integrantes.

O protótipo da equipe 1 foi o que mais utilizou as sugestões das *guidelines*, como pode ser observado na Figura 4.11. Do perfil A, apenas esta equipe utilizou a estratégia de rotulagem por classe (G3). A equipe justificou o uso da *guideline* de contextualização do dado (G2) a partir dos gráficos disponíveis em outra aba da tela

principal da rotulagem.

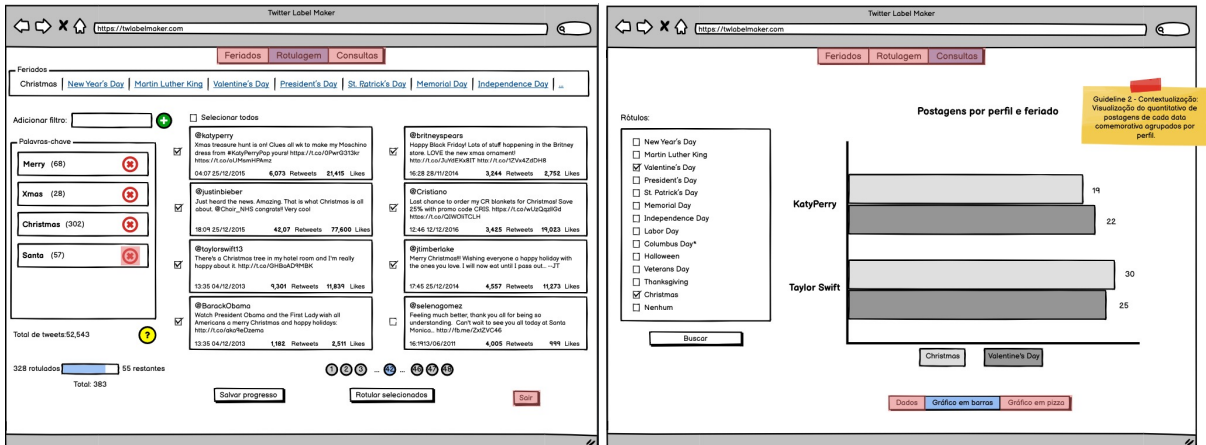


Figura 4.11 – Protótipo da equipe 1.

Dentre os protótipos das equipes de perfil A, o protótipo da equipe 2 foi o que menos utilizou as *guidelines*, como pode ser observado na Figura 4.12. A sugestão de barra de progresso (G4) foi substituída por um gráfico de *pizza*. Além disso, não foi utilizada a sugestão de indicar atalhos (G5), apesar de ter incluído o ícone de ajuda. Neste protótipo, não há como pausar ou cancelar a tarefa de rotulagem como proposto pela *guideline* de descanso durante a rotulagem (G6).



Figura 4.12 – Protótipo da equipe 2.

No protótipo da Figura 4.13, foi utilizada uma base de notícias para detectar sexismo e nele foram o que mais utilizaram as sugestões das *guidelines*. Não ficou claro o uso da *guideline* de ajuda e atalhos (G5). A equipe havia justificado o uso da

guideline 5 através dos exemplos de comentários sexistas. No entanto, esses exemplos se aplicam ao uso da *guideline* de contextualização do dado (G2). Este sistema já havia sido implementado anteriormente por uma das integrantes da equipe, e sua observação foi que gostaria de ter tido acesso essas *guidelines* antes de desenvolver o que está lançado atualmente.

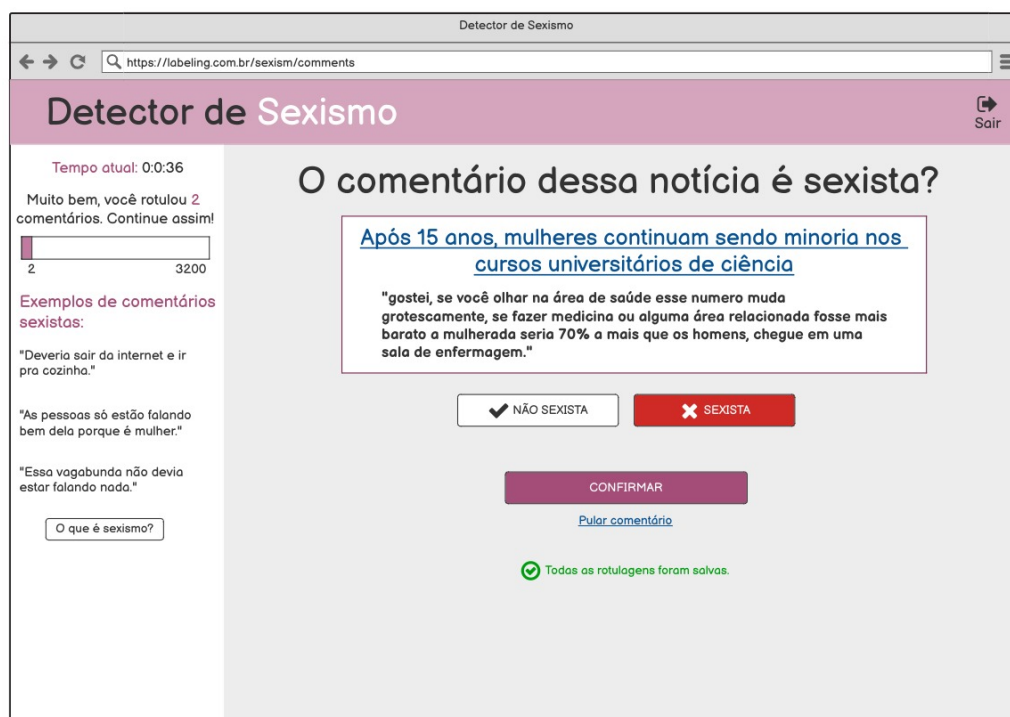


Figura 4.13 – Protótipo da equipe 3.

A Figura 4.14 apresenta as telas do protótipo da equipe 5. Esta equipe foi a única que fez um protótipo de sistema de rotulagem em uma versão *mobile*. A equipe justificou que a versão *mobile* poderia ter uma melhor interação na tarefa de detecção de entidades por conta do *touchscreen* dos *smartphones*. Por conta desta versão *mobile*, não houve necessidade de sugerir atalhos (G5) no protótipo. Apesar de haver uma barra de progresso na *home*, não é possível saber o andamento da rotulagem (G4) nas outras telas. Além disso, não foi utilizada a *guideline* de contextualização do dado (G2).

Por fim, apesar da experiência e conhecimento dos integrantes, apenas uma equipe utilizou todas as *guidelines*. A *guideline* de contextualização do dado (G2) foi a menos utilizada.



Figura 4.14 – Protótipo da equipe 5.

4.6.3 Protótipos de perfil B

As equipes de perfil B possuíam integrantes com conhecimento e experiência apenas em ML. Estes integrantes nunca utilizaram ou projetaram sistemas de rotulagem de dados. É esperado que este perfil tenha resultados semelhantes ao perfil A por entenderem conceitos de ML e a importância da etapa de rotulagem de dados.

Na Figura 4.17, é possível notar apesar de terem utilizado a sugestão de barra de progresso do *guideline* de *status* da rotulagem (G4), não é possível identificar numericamente de como está o andamento da tarefa. Além disso, o protótipo conta com o ícone de ajuda, porém não indica o uso de atalhos como proposto no *guideline* de ajuda e atalhos (G5). A equipe justificou o uso da *guideline* de contextualização do dado (G2) com a utilização de exemplos (texto em azul). Entretanto, os exemplos citados são classes e não instâncias que contextualizem o dado.

Na Figura 4.16, foi possível notar que a equipe decidiu por apresentar quatro questões binárias para um único *tweet*. Apesar de a equipe ter utilizado boa parte das *guidelines*, essa rotulagem multi-propósito pode confundir o usuário do que deve ser feito e tornar a rotulagem cansativa. Além disso, nele não foi detectado o uso da *guideline* de contextualização do dado (G2).

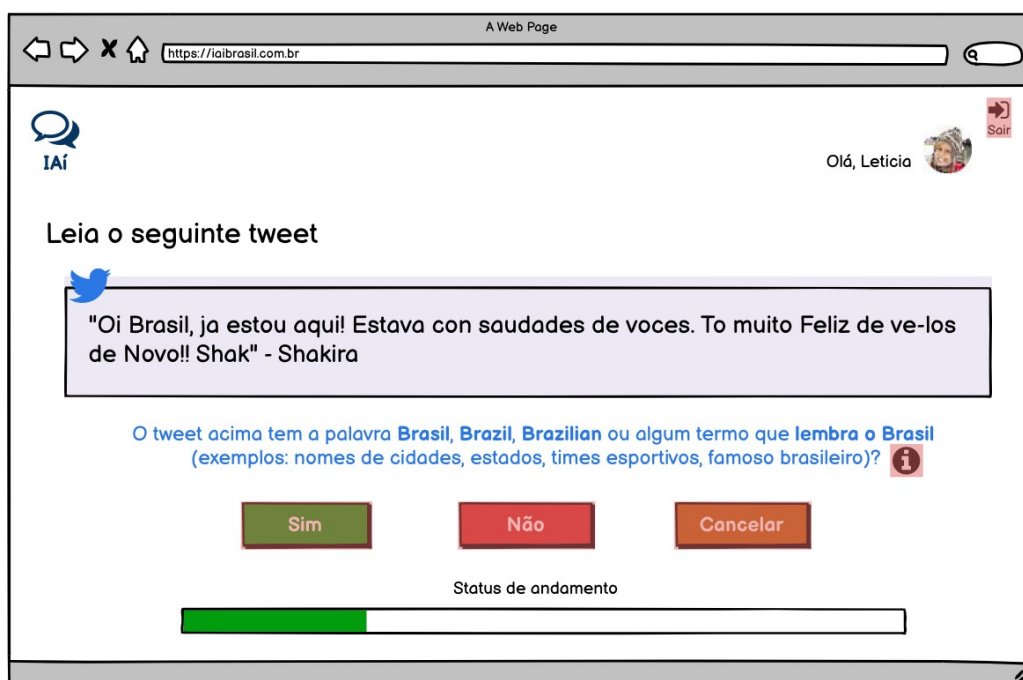


Figura 4.15 – Protótipo da equipe 4.

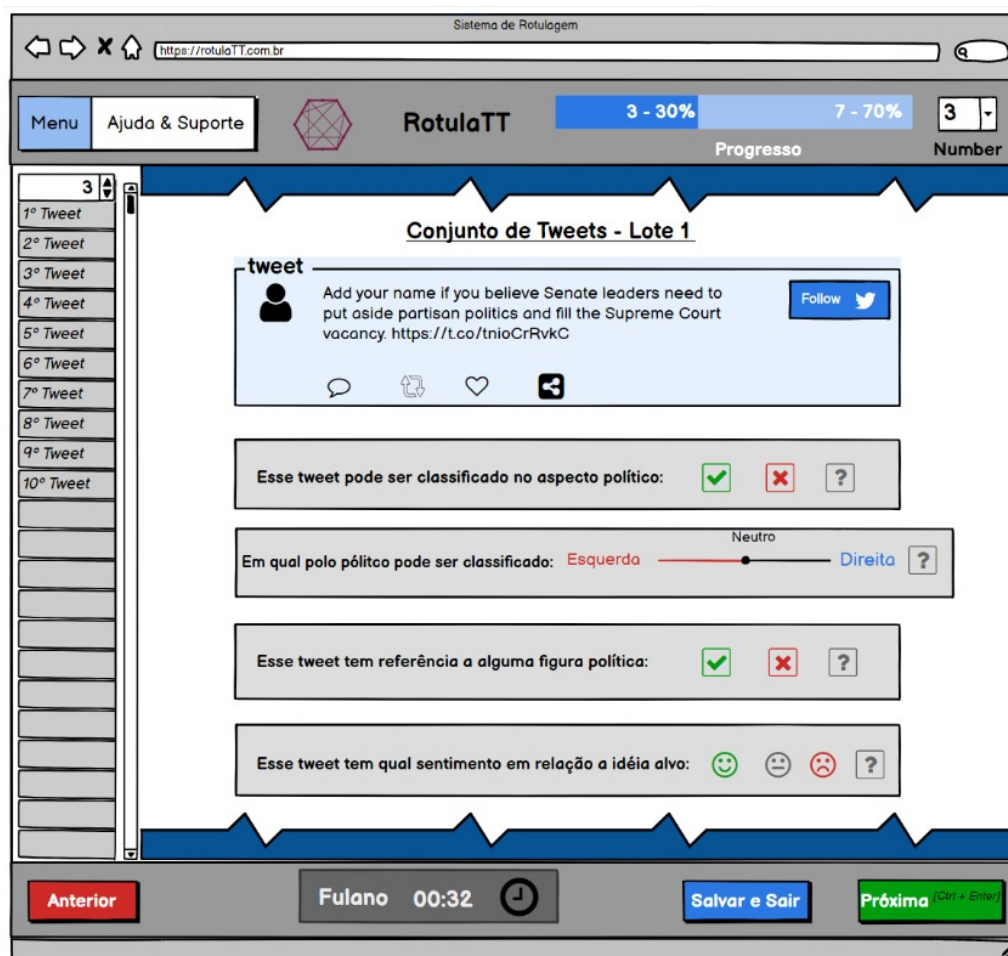


Figura 4.16 – Protótipo da equipe 9.

Por fim, como esperado, os resultados dos protótipos das equipes de perfil B foram semelhantes aos protótipos de perfil A. Novamente, a *guideline* de contextualização do dado (G2) foi a menos utilizada.

4.6.4 Protótipos de perfil C

As equipes de perfil C não possuíam integrantes com conhecimento e experiência em ML e nem sistemas de rotulagem de dados. Essa limitação foi um fator importante para avaliar se as *guidelines* estão apropriadas para apoiar profissionais neste perfil.

Os protótipos da equipe 7 (Figura 4.17) e equipe 8 (Figura 4.18) tratavam de problemas de classificação binária. Ambos protótipos não utilizaram as *guidelines* de contextualização do dado (G2), *status* da rotulagem (G4) e nem de descanso durante a rotulagem (G6).



Figura 4.17 – Protótipo da equipe 7.

Quanto à *guideline* de ajuda e atalhos (G5), apenas a equipe 8 a utilizou, aplicando somente o ícone de ajuda e sem sugestões de atalhos. Um ponto interessante notado foi que a equipe 8, além da equipe 1 (perfil A), utilizou a estratégia de rotulagem por classe (G3). Conceitualmente, a estratégia de rotulagem por instância é mais fácil por ser mais intuitiva.

Por fim, foi possível notar que os protótipos das equipes do perfil C foram os

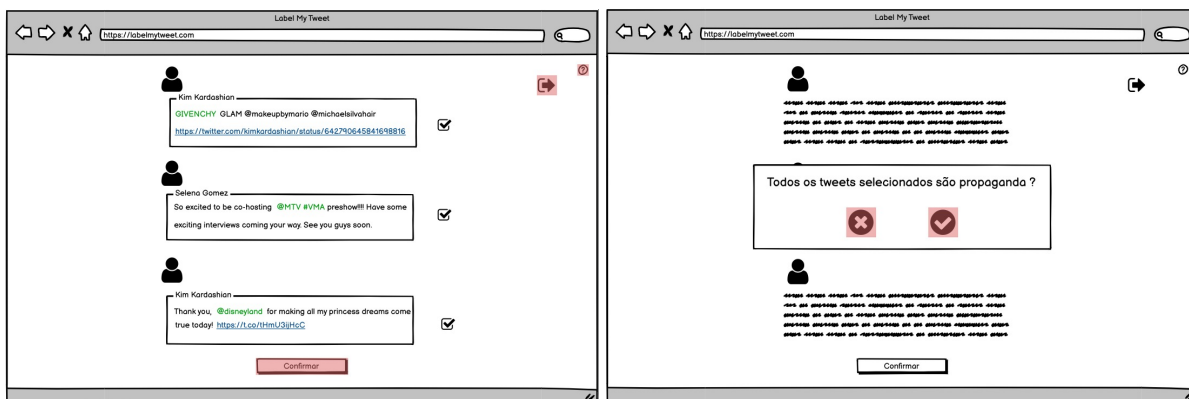


Figura 4.18 – Protótipo da equipe 8.

que menos utilizaram as *guidelines*. Este fato foi presente em todos os protótipos das equipes de perfil C. Além disto, a equipe 6 não compreendeu o problema proposto e não desenvolveu um sistema de rotulagem, como pode ser observado na Figura 4.19.

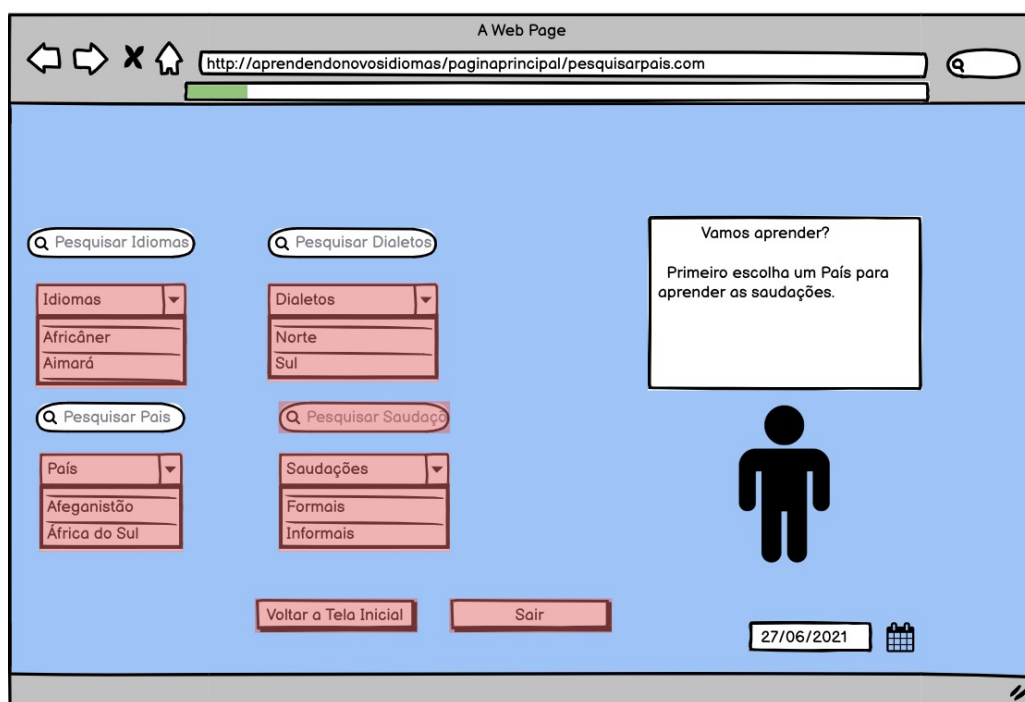


Figura 4.19 – Protótipo da equipe 6.

A equipe 6 tinha como problema identificar os tipos de saudações utilizados por cada país em *tweets*, porém foi projetado um sistema de aprendizado de idiomas. Ainda assim, não foram utilizadas as *guidelines* G4, G5 e G6 que poderiam ter sido aplicadas neste contexto.

4.6.5 Percepção de uso das *guidelines*

Após a construção dos protótipos, 14 integrantes das equipes, sendo pelo menos 1 de cada equipe, responderam um questionário sobre a percepção deles sobre as *guidelines*. Para cada *guideline*, foi perguntado:

1. *Qual o nível de dificuldade desta guideline?* Esta pergunta era no formato de escala *likert*, em que 1 representava “*muito fácil*” e 5 representava “*muito difícil*”;
2. *Você utilizaria esta guideline?* Esta pergunta era no formato de escala *likert*, em que 1 representava “*sim, sempre*” e 5 representava “*de jeito nenhum*”;
3. *Qual a sua percepção sobre esta guideline? O que mais lhe chamou atenção? O que você faria diferente?* Esta pergunta era aberta e obrigatória.

A seguir, estão os resultados obtidos das perguntas de cada *guideline*. São apresentados os gráficos com os resultados quantitativos das perguntas 1 e 2. Em seguida, são apresentados alguns trechos das respostas referente à pergunta 3. Para esta parte do estudo não foram considerados os perfis dos participantes dado que nem todos estavam presentes nesta fase do estudo.

G1 - Identificação da tarefa da rotulagem

Na Figura 4.20, é possível notar que apesar desta *guideline* não ser considerada muito fácil, ainda assim ela seria utilizada.

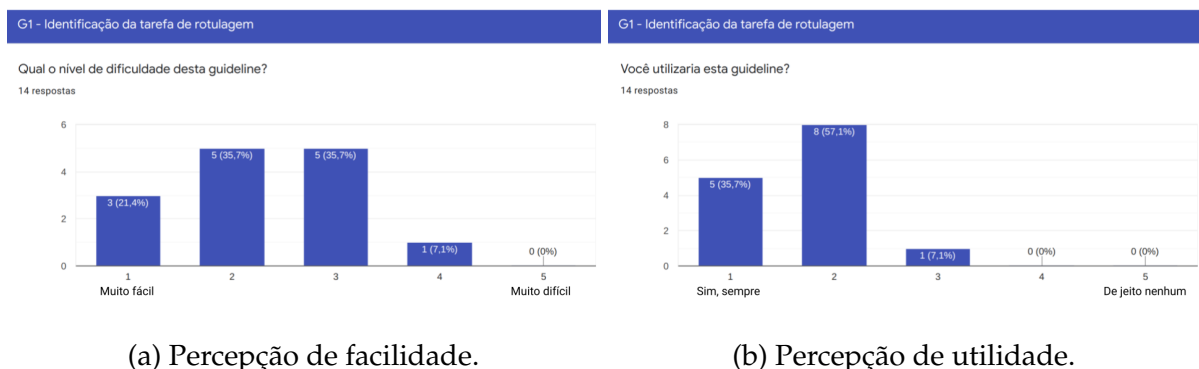


Figura 4.20 – Percepções de facilidade e utilidade consideradas pelos participantes sobre a *guideline* 1.

Quanto às respostas da pergunta 3 para esta *guideline*, foram destacadas as seguintes respostas:

- *Poderia descrever um exemplo real e ir aplicando os passos para classificação: Passo 1 - Leitura do Texto pelo usuário; Passo 2 - é solicitado que o usuário escolha as opções "Relevante" ou "Não Relevante"; Passo 3 - Usuário confirma a seleção e passa para a próxima classificação ou finaliza o acesso.*
- *Gostei muito da apresentação, mas talvez ficasse melhor num contexto de representação em tela.*
- *Adicionaria esses exemplos e mais outros nesta guideline.*

É possível notar nas respostas a necessidade dos participantes de mais exemplos visuais e dentro de um contexto real.

G2 - Contextualização do dado

Na Figura 4.21, é possível notar que esta *guideline* foi difícil de entender e não seria tão utilizada pelos participantes.

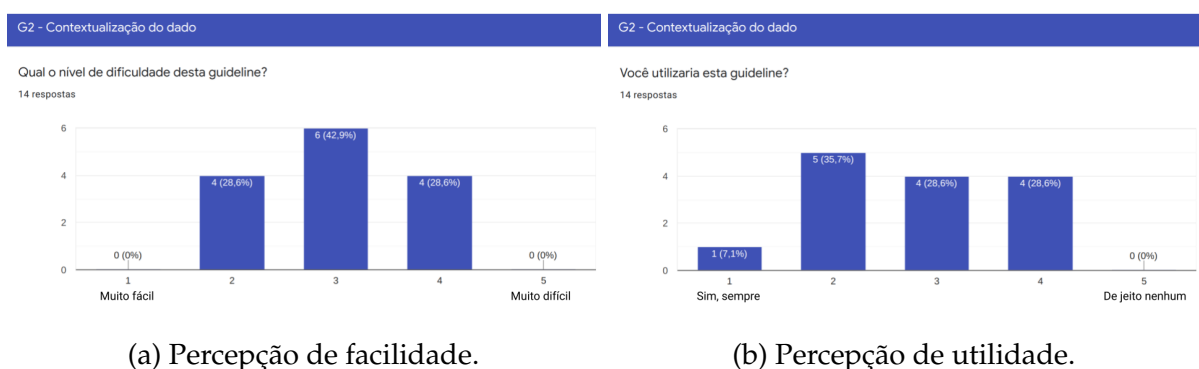


Figura 4.21 – Percepções de facilidade e utilidade consideradas pelos participantes sobre a *guideline* 2.

Quanto às respostas da pergunta 3 para esta *guideline*, foram destacadas as seguintes respostas:

- *A visualização por tópicos poderia ser um pouco mais visual.*

- Não entendi a segunda imagem, adicionaria mais dicas.
- Tentar avaliar novas oportunidades de problemas para trazer mais exemplos.
- Ao ler, não identifiquei de imediato como essa guideline poderia ser usada. Acredito que se for apresentado o uso em um exemplo de uma aplicação real, ficará mais simples para o leitor entender. Um exemplo: essa guideline será utilizada quando houver uma necessidade de visualização de termos frequentes, por exemplo: para o Problema de ML XYZ foi utilizado a Contextualização do dado em forma de "Distribuição de termos frequentes" para facilitar a resolução do problema XYZ (usa uma imagem ou um fluxo para demonstrar, esse processo).

Novamente, é possível notar nas respostas a necessidade dos participantes de mais exemplos visuais e dentro de um contexto real.

G3 - Rotulagem por instância ou por classe

Na Figura 4.22, é possível notar que esta *guideline* foi considerada parcialmente difícil de entender, mas ainda seria mais utilizada do que a *guideline* anterior.

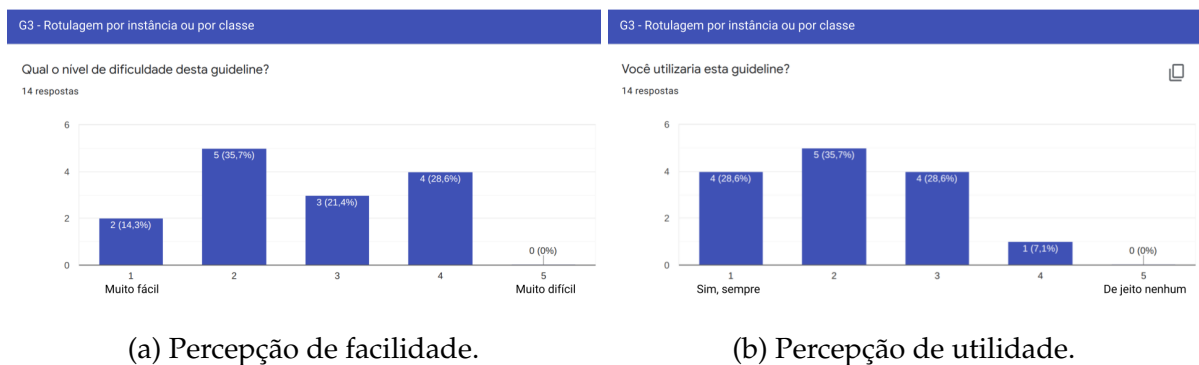


Figura 4.22 – Percepções de facilidade e utilidade consideradas pelos participantes sobre a *guideline* 3.

Quantos às respostas da pergunta 3 para esta *guideline*, foram destacadas as seguintes respostas:

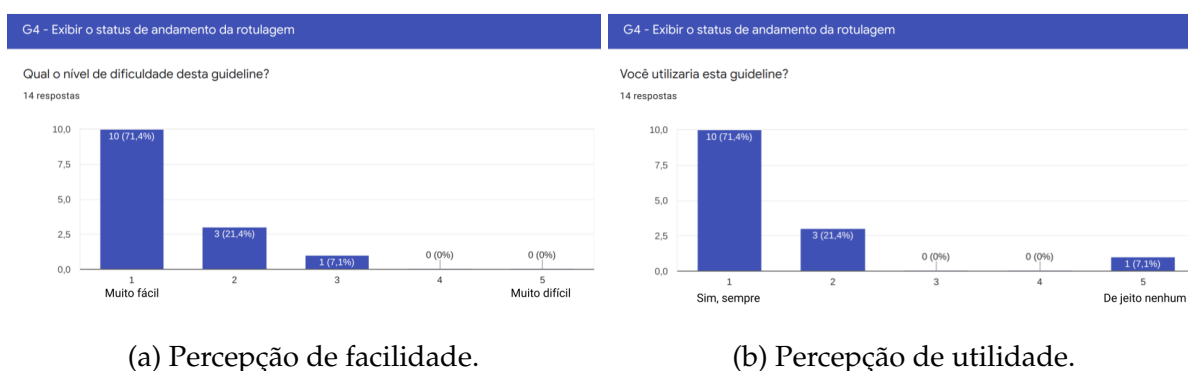
- Descobriria mais tipos de problemas associados na *guideline*.
- A escolha de qualquer uma das opções depende da natureza da tarefa.

- Como ponto de melhoria acredito que falta contextualizar a guideline, descrever o que é classe e o que é instância, em seguida explicar o exemplo proposto na imagem.

Novamente, é possível notar nas respostas a necessidade dos participantes de mais exemplos visuais e dentro de um contexto real. Além disso, foi sugerido que as opções também dependessem do tipo de tarefa de ML, não somente pelo tipo de dado. Também foi apontado a necessidade de explicar os termos de ML utilizados nesta *guideline*.

G4 - Exibir o status da rotulagem

Na Figura 4.23, é possível notar que esta *guideline* foi considerada muito fácil de entender e que também seria muito utilizada pelos participantes. Houve apenas uma resposta apontando que não utilizaria esta *guideline* “de jeito nenhum”, porém foi verificado que este participante respondeu negativamente para todas as *guidelines*. Porém, as respostas das perguntas abertas deste participante foram positivas (“...simples de aplicar e muito necessário”), o que indica que houve alguma confusão ao preencher o questionário.



(a) Percepção de facilidade.

(b) Percepção de utilidade.

Figura 4.23 – Percepções de facilidade e utilidade consideradas pelos participantes sobre a *guideline* 4.

Quanto às respostas da pergunta 3 para esta *guideline*, foram destacadas as seguintes respostas:

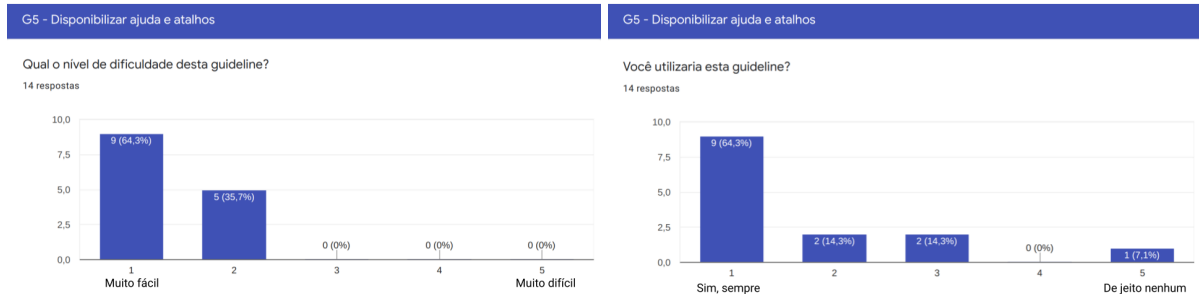
- Muito importante mostrar o andamento para o usuário, no entanto, pode dar uma impressão que falta muito para finalizar, usaria mais com porcentagem.

- *Esta guideline me pareceu ótima para verificar o progresso, mas a barra de progresso poderia ser por percentual, acredito que a visualização seria melhor.*
- *Interessante. Colocaria mais exemplos,*

Nestas respostas, foi verificada a sugestão de alterar a sugestão da barra de progresso com valores reais para porcentagem. Além disso, alguns participantes gostariam de ter mais exemplos.

G5 - Disponibilizar ajudas e atalhos

Na Figura 4.24, é possível notar que esta *guideline* foi considerada muito fácil de entender e que também seria muito utilizada pelos participantes. Novamente, houve apenas uma resposta apontando que não utilizaria esta *guideline* “*de jeito nenhum*”, e foi dada pelo participante que respondeu negativamente para todas as *guidelines*.



(a) Percepção de facilidade.

(b) Percepção de utilidade.

Figura 4.24 – Percepções de facilidade e utilidade consideradas pelos participantes sobre a *guideline* 5.

Quanto às respostas da pergunta 3 para esta *guideline*, foram destacadas as seguintes respostas:

- *Uma sugestão é colocar algo relacionado à prevenção de erros de quem está rotulando.*
- *Adicionaria mais dicas de atalhos, acho que pode ser mesclada em outra guideline.*
- *Usaria botões e ícones de ajuda, mas creio que os atalhos possam levar o usuário a apertar sem querer e acabar rotulando algo errado. Passa mais segurança clicar em um botão confirmando minha escolha, a qual deve estar claramente visível.*

Nestas respostas, foi novamente observada a necessidade de mais exemplos. Uma observação anotada foi referente a associação com a heurística de prevenção de erros e a possibilidade de desativar os atalhos para evitar erros acidentais do teclado.

G6 - Descanso durante a rotulagem

Na Figura 4.25, é possível notar que esta *guideline* foi considerada muito fácil de entender e que também seria muito utilizada pelos participantes. Novamente, houve apenas uma resposta apontando que não utilizaria esta *guideline* “de jeito nenhum”, e foi dada pelo participante que respondeu negativamente para todas as *guidelines*.

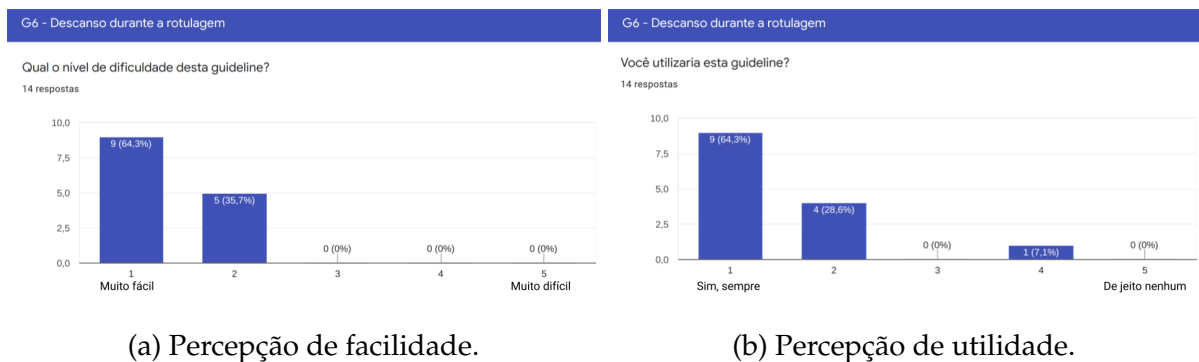


Figura 4.25 – Percepções de facilidade e utilidade consideradas pelos participantes sobre a *guideline* 6.

Quanto às respostas da pergunta 3 para esta *guideline*, foram destacadas as seguintes respostas:

- *Acho importante ter a opção de salvar, mas acho que o ideal é salvar automaticamente depois de tantas rotulagens.*
- *Tive a impressão que a guideline anterior poderia ser incorporada nessa.*
- *Acho que o botão de salvar e sair serve mais ao princípio do de descanso. No salvar automático, muito provável que eu teria me esquecido de salvar e saia pela própria aba.*

Nestas respostas, foi sugerido que o salvar automático fosse padrão. Além disso, foi observado que esta *guideline* poderia incluir a anterior. Também há uma observação quanto a associação da sugestão proposta e o nome da *guideline*.

Após a finalização das perguntas sobre cada *guideline*, os participantes foram solicitados que selecionassem as *guidelines* que foram fáceis de entender, as que foram úteis e quais eles adotariam em cenários futuros. As Figuras 4.26, 4.27 e 4.28 apresentam os resultados obtidos destas três perguntas.

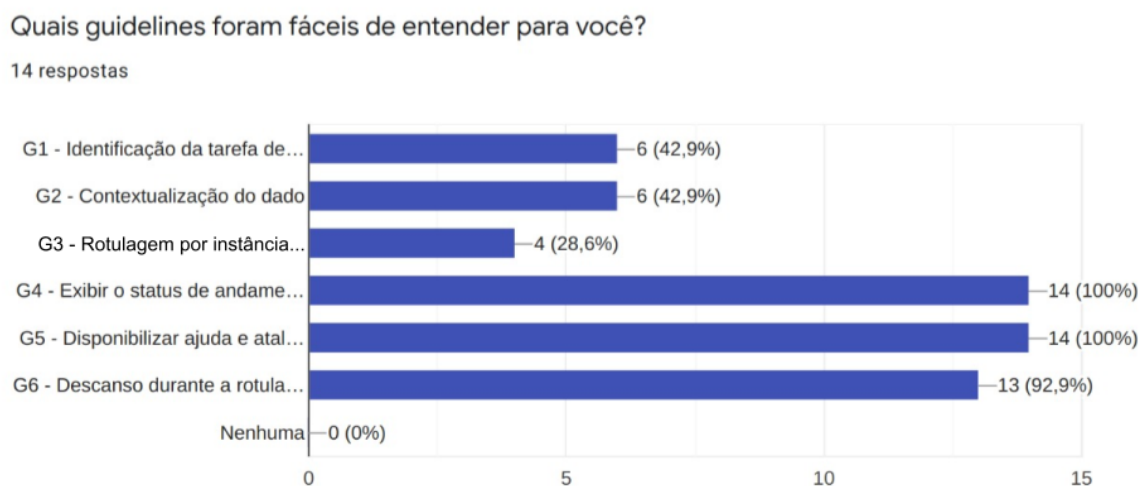


Figura 4.26 – Distribuição de *Guidelines* que foram consideradas fáceis de entender pelos participantes.

É possível notar na Figura 4.26 que as *guidelines* 4, 5 e 6 foram as consideradas mais fáceis de entender. Contudo, a *guideline* 3 foi a menos votada como fácil de entender. Além disso, as *guidelines* 2 e 3 não são consideradas por mais da metade dos participantes que responderam o questionário.

Na Figura 4.27, os resultados foram um pouco diferentes, apesar das *guidelines* 4, 5 e 6 serem as mais votadas. Nesta pergunta, a *guideline* 2 foi a menos votada como útil. A *guideline* 1 é útil por mais da metade dos participantes.

Na Figura 4.28, as *guidelines* 1, 4, 5 e 6 são as favoritas para serem adotadas em cenários futuros. No entanto, as *guidelines* 2 e 3 foram as menos votadas pelos participantes.

Quais guidelines foram úteis para você?

14 respostas

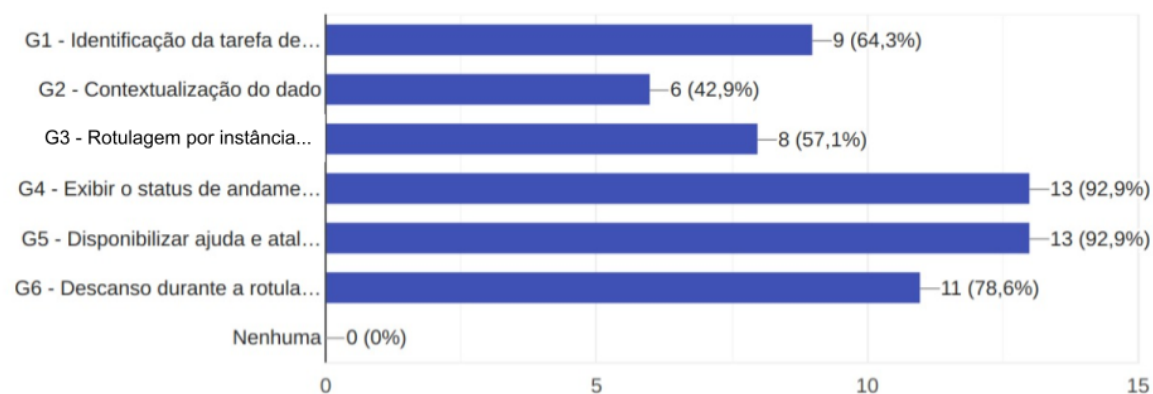


Figura 4.27 – Distribuição de *Guidelines* que foram consideradas úteis pelos participantes.

Quais guidelines você adotaria em cenários futuros?

14 respostas

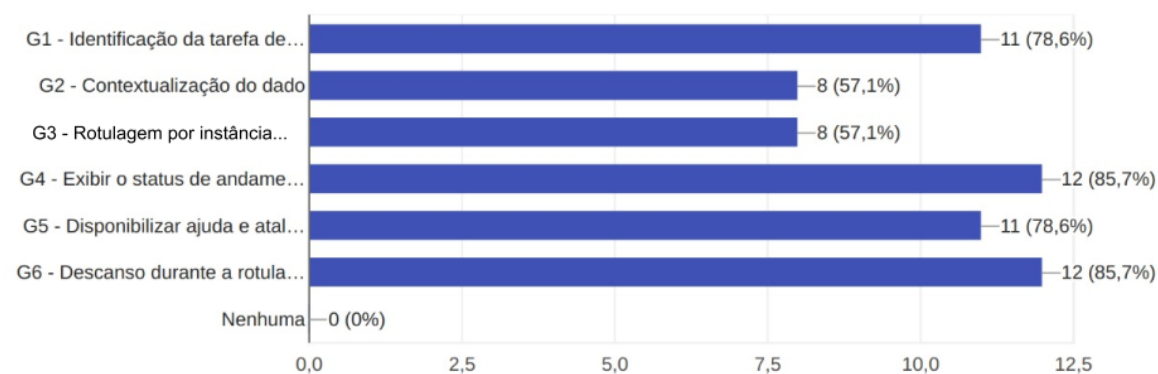


Figura 4.28 – Distribuição de *Guidelines* que poderiam ser adotadas em cenários futuros pelos participantes.

4.6.6 Considerações finais

Neste estudo com uma turma de Interação Humano-Computador, foi possível analisar que projetar com as *guidelines* ainda dependem de um conhecimento prévio de ML para obter êxito. Além disso, as *guidelines* 2 e 3 não estão claras e estão propensas em não serem adotadas em cenários futuros. As *guidelines* 4, 5 e 6 independem do tipo de dado e foram as que mais receberam *feedback* positivo dos participantes.

Um ponto abordado pelos participantes é a possibilidade de ter mais exemplos em um contexto real, apresentando não somente o componente isolado, mas contido numa interface de sistema de rotulagem de dados. Outro ponto, é a proposta de mais sugestões para a *guideline* 2 e 3 para cada tipo de problema de ML.

Por fim, a avaliação foi feita para contemplar projeto de sistema de rotulagem de dados textuais. Contudo, no geral, as *guidelines* cumpriram a finalidade de apoiar no projeto desses sistemas, apesar das limitações apontadas pelos participantes. No entanto, foi dada a necessidade de evoluir para uma solução mais eficiente em termos de soluções e uso.

A partir das observações apontadas, foi dada continuidade a pesquisa com uma análise dos sistemas de rotulagem do *GitHub* para a concepção dos padrões de interfaces *AnnotationUI*.

5

ANÁLISE DOS SISTEMAS DE ROTULAGEM DE TEXTO DO *GitHub*

Este capítulo descreve uma análise de sistemas de rotulagem de textos do *GitHub*. A análise faz parte do conjunto de estudos executados nesta pesquisa para a concepção e criação dos padrões de interface *AnnotationUI*. Durante a análise, foram verificadas as características dos sistemas de rotulagem e como elas suportam a atividade de rotulagem e a interação do usuário. As *guidelines LabelUX!* serviram de base para analisar as características dos sistemas de rotulagem.

5.1 Definindo a análise

A análise nas interfaces dos sistemas de rotulagem do *GitHub* seguiu algumas etapas. Essas etapas foram necessárias para garantir uma ampla cobertura dos sistemas de rotulagem de código aberto do *GitHub*. Foram utilizadas etapas inspiradas no protocolo de RSL de [Kitchenham, Budgen e Brereton \(2015\)](#). Ter etapas definidas permitem a reprodução da análise por outros pesquisadores, tornando-a auditável e reproduzível na maioria dos sistemas de rotulagem presentes no *GitHub*.

Primeiramente, foi definido a seguinte Questão de Pesquisa (QP):

- QP - *Quais as funcionalidades existem nas interfaces dos sistemas de rotulagem de dados textuais dos repositórios GitHub para dar suporte ao usuário durante a tarefa de rotulagem*

de dados?

A principal QP é ampla. Por isso, as Sub-Questões (SQ) foram definidas para dividir a principal questão de pesquisa por tipo de funcionalidade dos sistemas de rotulagem. As subquestões de pesquisa estão relacionadas a: *I) elementos de interface para apresentar dados textuais e solicitar feedback do usuário sobre rotulagem*: identificar os elementos de interface para apresentar dados textuais e feedback de rotulagem; *II) formas de interação*: identificar formas de interação em sistemas de rotulagem; *III) granularidade de rotulagem*: identificar estratégias de rotulagem quanto à granularidade dos dados; *IV) contextualização dos dados*: como apresentar informações sobre o modelo de ML aos usuários; *V) andamento da atividade de rotulagem*: identificar como apresentar o andamento da atividade de rotulagem aos usuários; *VI) opções de ajuda*: identificar as possibilidades de ajuda ao usuário; *VII) opções de atalho*: para identificar opções de atalho para as tarefas de rotulagem; e *VIII) feedback do status da rotulagem*: identificar a estratégia de feedback para mostrar o *status* do sistema de rotulagem. A Tabela 5.1 apresenta cada sub-questão de pesquisa.

Sub-Questões	Descrição
SQ1	Quais são os elementos de interface usados para apresentar os dados e solicitar feedback do usuário sobre rotulagem?
SQ2	Quais são as formas de interação na tarefa de rotulagem nos sistemas de rotulagem?
SQ3	Quais são as estratégias em relação à granularidade de dados nas tarefas de rotulagem?
SQ4	Como os sistemas apresentam as informações sobre o modelo ML para os usuários?
SQ5	Como os sistemas apresentam o progresso da atividade de rotulagem para os usuários?
SQ6	Como os sistemas apresentam opções de ajuda aos usuários durante as tarefas de rotulagem?
SQ7	Como os sistemas apresentam opções de atalhos para os usuários durante as tarefas de rotulagem?
SQ8	Como os sistemas apresentam o <i>status</i> da rotulagem?

Tabela 5.1 – Sub-questões de pesquisa.

Em seguida, as seguintes etapas foram realizadas: estratégia de busca, processo de seleção de sistemas de rotulagem e estratégia de extração de dados.

5.2 Estratégia de busca

Foi realizada uma busca exploratória no *GitHub* e encontrado nove repositórios de sistemas de rotulagem. Esses repositórios foram utilizados como referência para iniciar a busca por outros repositórios. Para isso, foram analisadas as *hashtags* dos repositórios de referência para descobrir outros sistemas de rotulagem. O conjunto de referência dos repositórios foi composto por: *Label Studio* (TKACHENKO et al., 2020-2021), *Brat* (STENETORP et al., 2011), *Doccano* (NAKAYAMA et al., 2018), *Inception* (KLIE et al., 2018), *Neuroner* (DERNONCOURT; LEE; SZOLOVITS, 2017), *YEDDA* (YANG et al., 2018), *Webanno* (CASTILHO et al., 2016), *Tag Editor* (D5555, 2021) e *ML Annotate* (IO, 2017).

O *GitHub* tem duas opções de pesquisa de repositório disponíveis. A primeira opção de pesquisa é por *hashtag* e a segunda é por uma *string* de busca. Foram testadas ambas opções de pesquisa com os repositórios de controle. A opção de busca escolhida foi por *hashtag* devido às restrições e limitações encontradas na busca por *string*. Para verificar essas restrições e limitações das buscas foram comparadas a completude de ambas.

5.2.1 Busca no *GitHub*

Kitchenham, Budgen e Brereton (2015) utilizam a definição de completude para verificar a cobertura da estratégia de busca em relação aos artigos encontrados. Esse conceito de completude foi utilizado para verificar a cobertura das duas opções de buscas do *GitHub*. Foram aplicados os critérios de *recall* para avaliar a integridade de ambas as opções de pesquisa no *GitHub*.

O *recall* da busca por *hashtag* foi de 77,78%, enquanto o *recall* da *string* de busca foi de 22,22%. Devido à instabilidade da busca por *strings* e imprecisão, foi decidido prosseguir com a busca por *hashtag*. A próxima sub-seção detalha os problemas encontrados na busca por *string*.

5.2.1.1 Problemas na busca por *string*

Na prática, o *GitHub* não permite o uso dos conectores OR, AND, NOT na busca de *hashtag*, apesar do site do *GitHub* informar ser possível utilizar os conectores nesse tipo de busca ¹.

Então, cada *hashtag* dos repositórios de controle foi testada separadamente. O conjunto de *hashtags* foi composto da seguinte forma:

- *annotation-tool; text-annotation; data-labeling; tagging-tool; labeling-tool; text-tagging; labeling.*

A busca por *hashtag* procura as palavras no nome do repositório e a descrição "sobre" em cada repositório do *Github*. Assim, após ser verificado as principais palavras-chave dos projetos de controle, foi decidido utilizar a seguinte consulta de pesquisa:

- *(text OR data) AND (annotation OR label) AND (tool OR system)*

No entanto, foram encontrados apenas dois de sistemas de controle nesta primeira busca, apesar de retornar mais de 1000 projetos. Foi analisado novamente o *GitHub Docs* e foi concluído que não é possível usar caracteres especiais na consulta de pesquisa². Assim, a consulta de pesquisa foi dividida em pequenas consultas usando apenas o operador AND (consulte a Tabela 5.2). Porém, mesmo após coletar todos os resultados da pesquisa, apenas dois projetos de controle retornaram.

Por fim, para entender mais sobre a busca no *GitHub*, foi realizada uma pesquisa na internet para verificar o que os praticantes dizem à respeito em uma comunidade de desenvolvedores³. Foi então verificado que não é possível usar uma consulta de pesquisa de maneira útil, pois a pesquisa do *GitHub* é instável e sua documentação parece estar desatualizada. Assim, a busca continuou somente por meio de *hashtags* individuais, realizando várias buscas (cada *hashtag* tinha uma busca individual).

¹ <<https://docs.github.com/en/search-github/getting-started-with-searching-on-github/troubleshooting-search-queries>>

² <<https://docs.github.com/en/search-github/searching-on-github/searching-code#considerations-for-code-search>>

³ <<https://stackoverflow.com/questions/29136057/can-i-search-github-labels-with-logical-operator-or/61618255#61618255>><<https://stackoverflow.com/questions/29136057/can-i-search-github-labels-with-logical-operator-or/61618255#61618255>>

String de busca	quantidade	Total encontrado	Projetos de controle encontrado	Recall
text AND annotation AND tool	148	1	Yedda	11,11
text AND label AND tool	105	0		0,00
data AND annotation AND tool	272	1	Label Studio	11,11
data AND label AND tool	338	1	Label Studio	11,11
text AND annotation AND system	38	0		0,00
text AND annotation AND software	21	0		0,00
text AND label AND system	81	0		0,00
text AND label AND software	34	0		0,00
data AND annotation AND system	58	0		0,00
data AND annotation AND software	42	0		0,00
data AND label AND system	194	0		0,00
data AND label AND software	87	0		0,00
Total	1418	2		22,22
Total sem duplicados	1121			

Tabela 5.2 – Consultas de pesquisa usando apenas o operador AND.

Para auxiliar no processo de busca, foi desenvolvido um *script* para fazer um *scraping* em algumas informações de cada repositório encontrado nas buscas de *hashtags*. Foi coletado o nome e o link do repositório, a descrição "sobre" e suas *hashtags*. Durante a execução do *script*, algumas instabilidades foram encontradas no servidor do *GitHub*, o que adicionou um atraso de três segundos a cada solicitação do repositório para garantir a coleta das informações necessárias.

5.2.2 Processo de Seleção dos sistemas de rotulagem

Foram utilizados dois filtros durante o processo de seleção dos sistemas de rotulagem. Primeiro, foram verificados alguns atributos dos repositórios, como o título, a descrição "sobre" como também o arquivo *README* dos repositórios. Foi analisado se o repositório se referia a um sistema de rotulagem de dados textuais, aplicando os critérios de inclusão e exclusão (Tabela 5.3). No segundo filtro, os repositórios foram baixados e instalados, aplicando os critérios de seleção novamente (Tabela 5.3).

Sobre os critérios de exclusão, existem algumas considerações. Não foram considerados os repositórios com sistemas de rotulagem em que o usuário tinha que interagir por meio de linhas de código, ou seja, sistemas que não possuem interfaces gráficas (CE2). Não foram considerados os repositórios que não puderam ser usados e testados, seja porque não possuem instruções claras de instalação do sistema ou porque foram arquivados (CE6).

5.2.3 Estratégia de extração de dados

Esta seção detalha o processo de extração dos sistemas de rotulagem. As *guidelines* criadas anteriormente (Seção 4.2) foi utilizada como base para a extração. Dois pesquisadores instalaram e testaram cada um dos sistemas de rotulagem para realizar a extração. Dois outros pesquisadores revisaram as extrações. Os dados extraídos foram os seguintes:

ID	Critérios de Seleção
CI1	O repositório deve conter uma ferramenta para auxiliar os usuários na etapa de rotulagem de dados do tipo texto (o sistema de rotulagem pode conter rotulagem para outros tipos de dados).
CI2	O repositório não possui um sistema de rotulagem textual, mas possui links para um ou mais sistemas de rotulagem.
ID	Critérios de exclusão
CE1	Não atende aos critérios de inclusão.
CE2	O sistema de rotulagem disponível no repositório não possui interface gráfica.
CE3	O sistema de rotulagem do repositório não está disponível publicamente para uso (seja na versão web ou desktop).
CE4	O repositório está duplicado ou é uma cópia de outro repositório incluído.
CE5	O repositório ou o sistema não possui a opção de idioma em inglês.
CE6	O repositório não tem instruções claras sobre como instalar/usar o sistema ou o repositório foi arquivado.
CE7	O repositório possui um sistema que não é pública e/ou não possui versões de demonstração para teste.
CE8	O repositório deve ter pelo menos uma <i>star</i> ou um <i>forks</i> .

Tabela 5.3 – Critérios de seleção

- **Nome do sistema, ano da primeira e última versão da ferramenta (Releases), link do repositório GitHub, palavras-chave, número de estrelas recebidas no repositório, número de forks do repositório, tipo de instalação necessária para testar o sistema de rotulagem.** Esses dados foram usados para caracterizar e identificar sistemas de rotulagem de texto.
- **Orientações e ajuda sobre como usar o sistema de rotulagem.** Esses dados apresentam informações se o repositório fornecer diretrizes para o uso de sistemas de rotulagem.
- **QP1 - elementos de interface são usados para apresentar dados textuais e solicitar feedback do usuário sobre rotulagem.** Identificar as estratégias de apresenta-

ção de texto na interface e as opções de rotulagem do usuário.

- **QP2 - formas de interação.** Identificar formas de interação na tarefa de rotulagem dos sistemas.
- **QP3 - granularidade de rotulagem.** Identificação de estratégias utilizadas em sistemas de rotulagem aplicados à rotulagem de dados múltiplos.
- **QP4 - contextualização do dado.** Esses dados representam como as informações estatísticas do modelo ML que está sendo treinado são apresentadas ao usuário durante a rotulagem.
- **QP5 - progresso da atividade de rotulagem.** Identificação dos elementos de interface usados para indicar o andamento da tarefa de rotulagem.
- **QP6 - opções de ajuda.** Identificação dos elementos de interface usados para permitir que os usuários solicitem ajuda.
- **QP7 - opções de atalhos.** Esses dados apresentam as opções de atalhos dos sistemas de rotulagem.
- **QP8 - *feedback* de *status* de rotulagem.** Representa soluções de interface usadas em sistemas de rotulagem para fornecer informações ao usuário sobre o *status* do sistema.

5.2.4 Resultados

Esta seção relata os resultados da extração da interface dos sistemas de rotulagem de dados do GitHub.

Foi realizado o teste de concordância estatística *Kappa* de [Cohen \(1960\)](#) para diminuir o viés do pesquisadores. Foram selecionados 20 repositórios do *GitHub* aleatoriamente, foram aplicados os critérios de seleção (Tabela [5.3](#)), testados e extraídos os sistemas de rotulagem. O valor de *Kappa* obtido foi de 0,82, que significa que existe uma força de concordância substancial entre os dois pesquisadores.

5.2.4.1 Repositórios Selecionados

O conjunto de *hashtags* retornou um total de 553 repositórios do *GitHub* (Figura 5.1). Foram encontrados repositórios com o mesmo sistema de rotulagem, estes foram excluídos por duplicidade (CE4). Foram aplicados os critérios de seleção no primeiro filtro, excluindo 463 repositórios. Na etapa seguinte, foi realizado a *backward snowballing*. Sete repositórios foram incluídos pelo critério de inclusão CI2, pois não são repositórios do sistema, mas possuem *links* para outros repositórios ou sistemas de rotulagem. Foram buscados os sistemas de rotulagem referenciados nos repositórios que foram analisados no primeiro filtro e encontrados 44 outros sistemas de rotulagem, dos quais 28 já estavam incluídos no primeiro filtro. Alguns repositórios têm *links* para sistemas de rotulagem que não estão no *GitHub*. Esses sistemas fora do *GitHub* eram os seguintes: Ubiai (UBIAI, 2021), Prodigy (PRODIGY, 2021) e Tagtog (CEJUELA; ROST, 2011). Não foi possível aplicar os critérios de exclusão do CE8 para esses sistemas, pois eles estão fora do *GitHub* e não possuem informações de *forks* e *stars* do repositório. Apesar disso, os três sistemas foram incluídos na análise.

Por fim, no segundo filtro foram aplicados novamente os critérios de seleção nos repositórios. Além disso, os sistemas de rotulagem foram extraídos. Nesta última etapa, foram excluídos 70 repositórios e extraídos 25.



Figura 5.1 – Processo de seleção dos sistemas de rotulagem.

Os 25 sistemas de rotulagem extraídos foram publicados no *GitHub* entre 2011 e 2021 (Figura 5.2). Além disso, os sistemas disponíveis nos repositórios possuem versões com correções e atualizações ao longo do tempo.

Cinco dos 25 sistemas de rotulagem têm muitos *forks* e *stars*, de 785 a 7.000 *forks* e de 263 a 795 *stars* (Figura 5.3). O número de *stars* significa o quanto outros usuários do *GitHub* gostaram do repositório. O número de *forks* representa o número de vezes

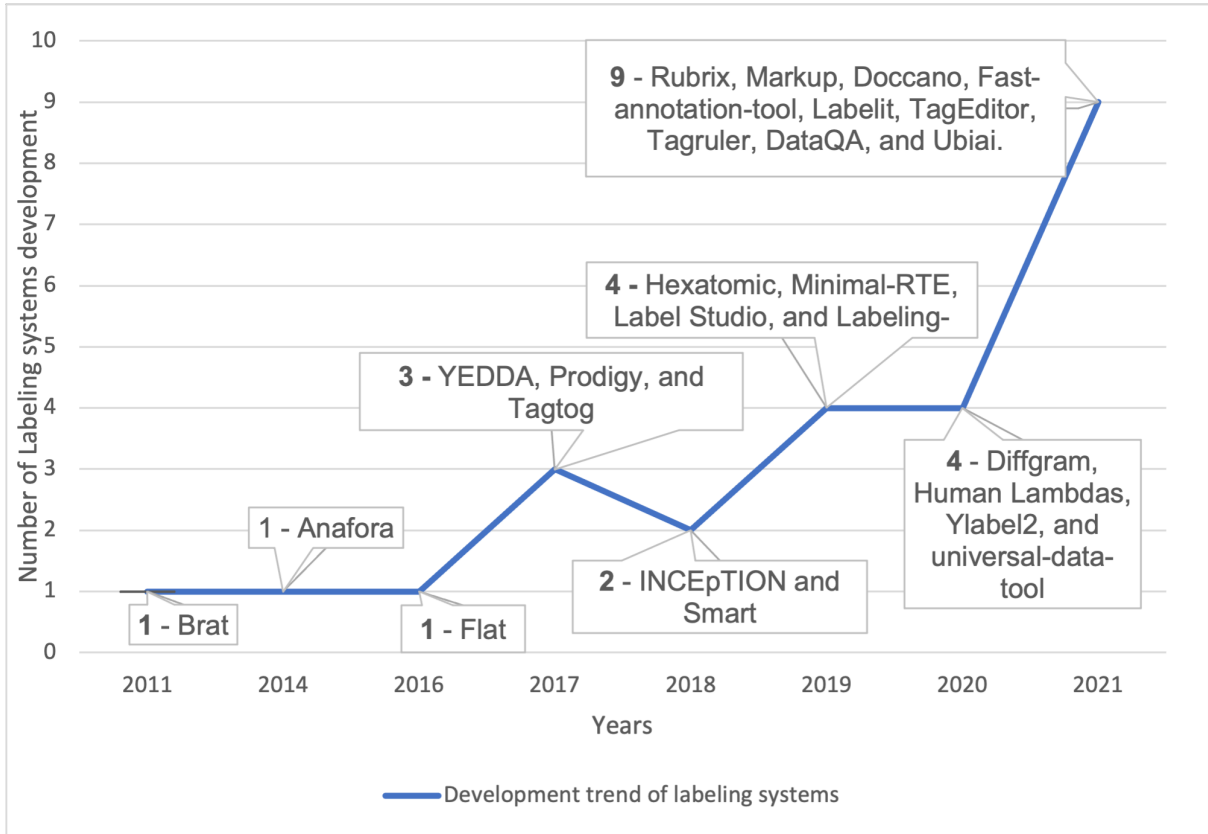


Figura 5.2 – Tendência de desenvolvimento de sistemas de rotulagem.

que outros usuários do *GitHub* duplicaram o repositório. Isso mostra a importância dos sistemas para os usuários que procuram este tipo de sistema.

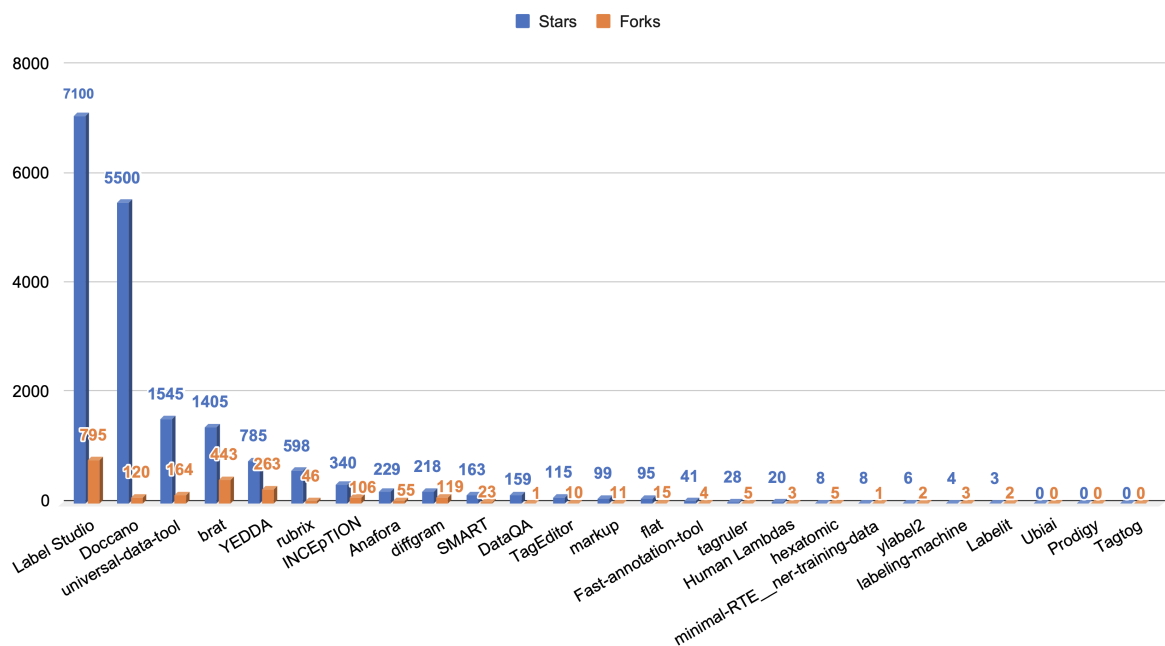


Figura 5.3 – A figura apresenta a quantidade de estrelas e bifurcações de cada repositório.

Apenas quatro sistemas (YEDDA (YANG et al., 2018), INCEpTION (KLIE et al., 2018), ylabel2 (LIGHTTAG, 2021) e labeling-machine (AGHAJANI et al., 2019)) dos 25 sistemas de rotulagem não orientaram o uso da ferramenta ou não continham informações completas sobre o sistema. A ausência de informações de uso foi um desafio para utilizar as ferramentas e, conseqüentemente, extrair as características dos sistemas.

O problema de NLP mais abordado em sistemas de rotulagem é o Reconhecimento de Entidade Nomeada (Figura 5.4). Vinte dos 25 sistemas de rotulagem possuem recursos que permitem aos usuários fazer esse tipo de rotulagem. Apenas dois sistemas de rotulagem (rubrix (RECOGNAI, 2021) e Human Lambdas (HUMAN-LAMBDAS, 2021)) contemplam a rotulagem para o problema de análise de sentimento. As informações são essenciais para entender como os desenvolvedores projetam as interfaces para um tipo específico de problema em NLP. Além disso, foi possível verificar que as interfaces de sistema que tratam de outros problemas de NLP podem usar as estratégias usadas nessas interfaces de sistema.

Também foi identificado que um sistema pode apresentar um ou mais problemas de NLP. Por exemplo, o sistema *Label Studio* (TKACHENKO et al., 2020-2021) funciona em Keyphrase Tagging, POS Tagging, classificação de texto, reconhecimento de entidade nomeada (NER) e extração de relação.

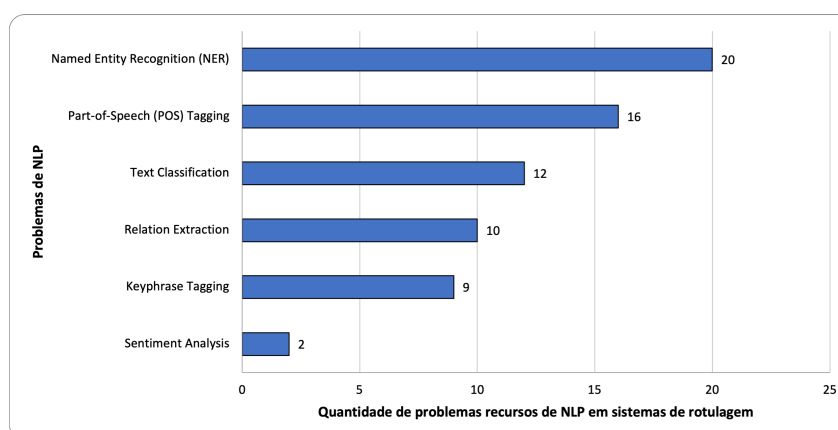


Figura 5.4 – A figura apresenta a cobertura dos sistemas de rotulagem para cada problema de NLP.

5.2.4.2 Respondendo à questão de pesquisa

Nesta seção é respondida a questão de pesquisa e discutido os resultados das subquestões. A QP foi respondida através das funcionalidades encontradas nas interfaces dos sistemas de rotulagem.

SQ1 - Quais são os elementos de interface usados para mostrar dados e obter *feedback* do usuário sobre a rotulagem?

As interfaces dos sistemas de rotulagem foram analisadas considerando como elas apresentam o texto aos usuários e como elas solicitam o *feedback* de rotulagem do usuário. Todas as interfaces oferecem apenas visualização de texto aos usuários, como uma exceção ao sistema de rotulagem `minimal-RTE__ner-training-data` (HOUZZIO, 2021). O sistema de rotulagem `minimal-RTE__ner-training-data` permite ao usuário editar o texto a ser rotulado.

Uma funcionalidade importante das interfaces é o destaque de palavras já rotuladas. O usuário pode observar quais palavras já foram rotuladas olhando o texto e observando as palavras destacadas. Treze das 25 interfaces fornecem destaque de palavras. As estratégias utilizadas para isso são: alterar a cor de fundo do texto rotulado e alterar a cor do texto. Cinco dos 12 sistemas restantes não usam realce de texto rotulado porque a tarefa de rotulagem é de classificação de texto, em que o usuário deve rotular todo o texto e não partes dele.

Em resumo, foram encontrados três atributos dos sistemas de rotulagem para a funcionalidade de apresentação de dados. 1) a própria apresentação dos dados por meio de blocos de texto; 2) Edição dos dados - o usuário pode alterar o texto durante a tarefa de rotulagem; e 3) destaque de palavras - o sistema apresenta determinadas palavras destacadas ao usuário.

Além disso, foram identificados um conjunto de elementos de interface usados para solicitar *feedback* de rotulagem dos usuários (Tabela 5.4). Dezesesseis sistemas utilizaram a lista de botões, que foi o componente mais utilizado nas interfaces dos sistemas de rotulagem. A lista de botões refere-se a todos os botões que permitem ao usuário rotular um texto, sejam botões binários (sim ou não), botões verticais ou botões horizontais. Foram classificados como uma lista de botões independentemente do posicionamento

dos botões na interface. A análise dos sistemas permitiu verificar que a lista de botões pode ser utilizada em qualquer sistema de rotulagem independente do problema de NLP.

Foi encontrado o componente de interface clicar e arrastar em sistemas que solicitam que o usuário clique e selecione duas partes do texto, criando assim uma relação entre as partes. Em seguida, o usuário atribui um rótulo ao relacionamento criado. Dez interfaces de sistema de rotulagem têm esse componente. Todos esses sistemas se concentram no problema de rotulagem de NLP de *relation extraction*. Não foram encontrados nenhum sistema que utilizasse outro componente para resolver o problema de rotulagem de *relation extraction*.

O componente *radio button* foi usado em duas interfaces do sistema para solicitar *feedback* de rotulagem do usuário. Uma característica deste componente é categorizar apenas uma classe. Nenhum dos sistemas analisados utilizou o componente *radio button* para categorização multi-rótulo. Esta solução resolveu os seguintes problemas de NLP: análise de sentimento, Keyphrase Tagging, POS Tagging, classificação de texto e NER.

Em duas interfaces, foi observado a utilização do componente *pop-up*. Ao clicar e selecionar uma parte do texto, uma janela *pop-up* é aberta com botões para rotulagem. Ao mostrar o *pop-up* para o usuário, outros componentes podem ser mostrados e assim compor um conjunto de componentes para solicitar o *feedback* da rotulagem. Esse componente atendeu aos seguintes problemas de NLP: Keyphrase Tagging, marcação POS, classificação de texto e NER. Embora poucos sistemas usem esse componente, não foi identificada nenhuma limitação ao seu uso para outros problemas de rotulagem de NLP.

Foi observado o uso do componente *input* em três sistemas de rotulagem. Este componente possibilita aos usuários criar novos rótulos e associá-los a uma parte do texto. Este componente é essencial quando todas as categorias não são conhecidas. O componente *input* foi encontrado em sistemas que resolvem problemas de Keyphrase Tagging, POS Tagging, NER e *relation extraction*.

Três sistemas usaram o componente *drop-down menu* em suas interfaces. Quando o usuário seleciona a parte do texto, a interface mostra o *drop-down menu* com uma lista

de valores, então o usuário navega na lista e escolhe um dos valores. O valor selecionado é o rótulo correto. O componente resolveu os problemas de Keyphrase Tagging, POS Tagging, classificação de texto, NER e *relation extraction*.

Dois sistemas de rotulagem usaram o componente *checkbox* para solicitar *feedback* de rotulagem do usuário. Este componente permite que o usuário marque uma ou mais categorias que julgue corretas sobre um texto. O *checkbox* pode ser útil para problemas de NLP multi-catagóricos. O componente resolveu os problemas de rotulagem para análise de sentimento, Keyphrase Tagging, POS Tagging, classificação de texto, NER e *relation extraction*. Ademais, apenas um sistema de rotulagem usou o componente de interface *icon* para solicitar *feedback* de rotulagem do usuário.

Por fim, apenas um sistema de rotulagem usou o componente *Emoji* para solicitar *feedback* de rotulagem do usuário. Cada rótulo possui um *Emoji* associado. A palavra ou frase rotulada tem um *Emoji* ao lado para mostrar seu rótulo. Este componente resolveu os seguintes problemas de NLP: Keyphrase Tagging, POS Tagging e NER. Além disso, não foi observado nenhuma limitação do componente que possa impedir que ele seja usado em outros problemas de rotulagem.

SQ2 - Quais são as formas de interação na atividade de rotulagem nos sistemas?

Analisando os sistemas foram encontradas duas formas principais de interação para rotulagem. Ambas as formas de interação são funcionalidades diferentes nos sistemas. O primeiro seleciona uma parte do texto e depois a associa a um rótulo selecionando-o. Segundo, o usuário primeiro escolhe o rótulo e depois a parte do texto que deseja rotular. Na segunda forma de interação, os sistemas de rotulagem, cobrem o problema de NLP de *relation extraction* (Inception (KLIE et al., 2018), universal-data-tool (TOOL, 2021), Label Studio (TKACHENKO et al., 2020-2021), TagEditor (D5555, 2021), DataQA (DATAQA, 2021), brat (STENETORP et al., 2012), Ubiai (UBIAI, 2021), Anafora (CHEN; STYLER, 2013), Prodigy (PRODIGY, 2021) e Tagtog citartagtog). Esses sistemas de rotulagem permitem que os usuários rotulem partes do texto e façam relacionamentos entre essas partes, o sistema Figura 5.5 é um exemplo desse sistema.

SQ3 - Quais são as estratégias em relação à granularidade de dados nas tarefas de rotulagem?

Formas de solicitar o <i>feedback</i>	Qtd. de sistemas	Sistemas de rotulagem
Lista de botões	16	Rubrix (RECOGNAI, 2021), YEDDA (YANG et al., 2018), INCEpTION (KLIE et al., 2018), Ylabel2 (LIGHTTAG, 2021), Universal-data-tool (TOOL, 2021), Label Studio (TKACHENKO et al., 2020-2021), Doccano (NAKAYAMA et al., 2018), Fast-annotation-tool (FAST..., 2021), Labelit (TKACHENKO et al., 2020-2021), TagEditor (D5555, 2021), Tagruler (MEGAGONLABS, 2021), DataQA (DATAQA, 2021), Ubiai (UBIAI, 2021), Anafora (CHEN; STYLER, 2013), Prodigy (PRODIGY, 2021), and Tagtog (CEJUELA; ROST, 2011).
Click and drag	10	INCEpTION (KLIE et al., 2018), Universal-data-tool (TOOL, 2021), Label Studio (TKACHENKO et al., 2020-2021), Fast-annotation-tool (FAST..., 2021), TagEditor (D5555, 2021), DataQA (DATAQA, 2021), Brat (STENETORP et al., 2012), Ubiai (UBIAI, 2021), Anafora (CHEN; STYLER, 2013), and Prodigy (PRODIGY, 2021).
Radio Button	2	Human Lambdas (HUMAN-LAMBIDAS, 2021) and Markup (SAMUELDOLBI, 2021).
Pop-up	2	Diffgram (DIFFGRAM, 2021) and Flat (PROYCON, 2021).
Input	3	INCEpTION (KLIE et al., 2018), Hexatomic (HEXATOMIC, 2021), and Doccano (NAKAYAMA et al., 2018).
Menu drop-down	3	Labeling-machine (AGHAJANI et al., 2019), DataQA (DATAQA, 2021), and Smart (CHEW et al., 2019).
Check-box	2	Human Lambdas (HUMAN-LAMBIDAS, 2021) and Prodigy (PRODIGY, 2021).
Emoji	1	Minimal-RTE__ner-training-data (HOU2ZIO, 2021).

Tabela 5.4 – Componentes de interfaces utilizadas para solicitar o *feedback* de rotulagem.

Foram identificados dois tipos de estratégias de granularidade de dados nas tarefas de rotulagem. Os sistemas aplicam as estratégias através das funcionalidades: rotulagem de unidades e rotulagem de lotes. A estratégia de rotulagem de unidades acontece quando o sistema apresenta uma instância por vez para rotular. A rotulagem em lote é quando o sistema oferece mais de uma instância para o usuário rotular, isso acelera o trabalho de rotulagem e economiza recursos e tempo do usuário. Apenas o

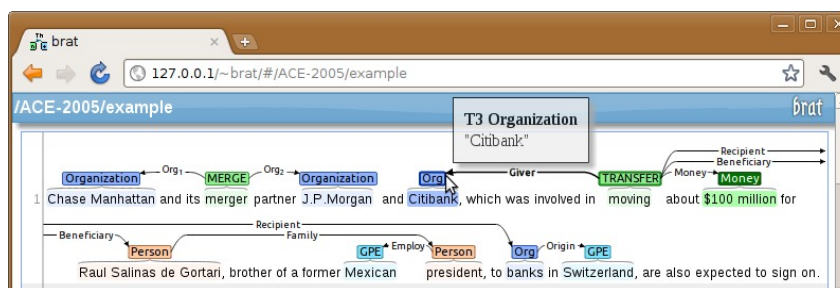
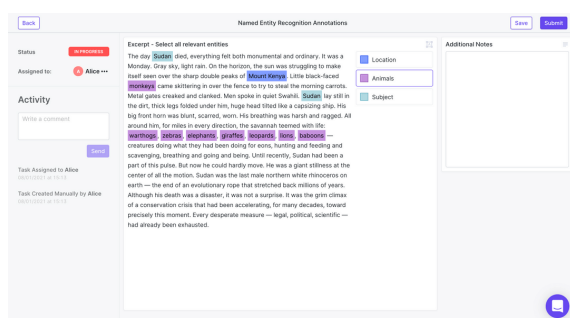
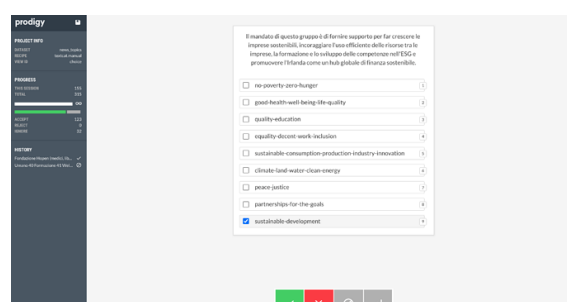


Figura 5.5 – Brat - sistema de rotulagem. A tarefa de rotulagem do usuário é a *relation extraction*. O usuário deve relacionar partes do texto entre si. A interação ocorre com o usuário clicando em uma das partes do texto e arrastando para a outra parte.



(a) Human Lambdas.



(b) Prodigy.



(c) TagEditor.



(d) Universal Data Tool.

Figura 5.6 – Interfaces de sistemas de rotulagem de dados.

sistema Rubrix (RECOGNAI, 2021) utiliza a estratégia de rotulagem de lotes, os demais sistemas utilizam a rotulagem de unidades.

SQ4 - Como os sistemas apresentam as informações sobre o modelo ML aos usuários?

Em relação à funcionalidade dos sistemas de rotulagem para indicar as informações sobre o modelo ML, as seguintes estratégias para implementar essa funcionalidade foram encontradas: Frequência de palavras (RECOGNAI, 2021), Destaque de palavras-chave (LIGHTTAG, 2021), Matriz de confusão (FAST..., 2021), ML Métricas de Modelo (MEGAGONLABS, 2021) e Dados de Fundo (CHEW et al., 2019).

A frequência das palavras é a informação de quantas palavras foram rotuladas. Essas informações ajudam o usuário a entender o quanto já foi rotulado em um texto. O destaque de palavras-chave é uma informação pré-processada pelo sistema de rotulagem que indica possíveis palavras-chave sobre o texto para o usuário. Ajuda o usuário a pensar sobre o tema central do texto.

A matriz de confusão é informação sobre o desempenho do ML. Para usar a matriz de confusão, é necessário o pré-processamento com os dados rotulados. Essas informações podem estar disponíveis sempre que o usuário rotular um dado e quiser ver como o modelo foi executado. As métricas *F1-score*, *coverage*, *precision*, and *recall* também são informações sobre o desempenho do modelo. Essas informações podem existir a partir de um conhecimento específico em ML de um usuário.

Por fim, foi identificado a estratégia de dados adicionais, que são informações extras sobre o texto. Esses dados fornecem informações para ajudar o usuário a entender o texto.

SQ5 - Como os sistemas apresentam o progresso da atividade de rotulagem para os usuários?

Em relação às funcionalidades utilizadas na interface para apresentar ao usuário o andamento da tarefa de rotulagem, foram identificadas oito tipos (Tabela 5.5) em 11 dos 25 sistemas de rotulagem. Todos os oito tipos destinam-se a transmitir o progresso da tarefa de rotulagem ao usuário. Esses tipos podem ser apresentados na interface apenas com uma mensagem de texto ou com o uso de elementos visuais.

SQ6 - Como os sistemas apresentam opções de ajuda aos usuários durante as tarefas de rotulagem?

Foram identificados três funcionalidades principais que auxiliam os usuários durante a tarefa de rotulagem. Dez dos 25 sistemas de rotulagem incluíam essas estratégias. A primeira funcionalidade de ajuda é um botão que contém informações sobre a tarefa de rotulagem (KLIE et al., 2018; SAMUELDORBI, 2021; TKACHENKO et al., 2020-2021; NAKAYAMA et al., 2018; D5555, 2021; CHEW et al., 2019; UBIAI, 2021; PRODIGY, 2021; CEJUELA; ROST, 2011). A segunda funcionalidade é quando o usuário passa o mouse sobre um elemento da interface, apresentando um pop-up

Elementos da interface de progresso da atividade de rotulagem	Quantidade de sistemas
Quantidade de textos rotulados	6
Barra de progresso	5
Quantidade de textos não rotulados	4
Quantidade de textos rotulados por rótulo	3
Linhas de texto rotuladas	1
Porcentagem de listas de texto rotuladas	1
Quais textos foram rotulados	1
Última linha de texto rotulada	1

Tabela 5.5 – Os componentes de interface usados para apresentar o progresso da rotulagem ao usuário

com informações (PROYCON, 2021). Por fim, a terceira funcionalidade de ajuda é um *chat-bot* onde o usuário pode fazer suas perguntas (CEJUELA; ROST, 2011).

SQ7 - Como os sistemas apresentam opções de atalhos para os usuários durante as tarefas de rotulagem?

Sete dos 25 sistemas de rotulagem possuem funcionalidades de atalho em suas interfaces. Foram identificadas duas estratégias de atalho. O primeiro são os atalhos de teclado para as ações principais do sistema, por exemplo, botão salvar (entrar), voltar (voltar) e cancelar (ESC). A segunda estratégia de atalho é para rótulos, onde cada rótulo possui um atalho de teclado. Embora poucas ferramentas contenham essa funcionalidade, ela pode economizar tempo do usuário durante a rotulagem.

SQ8 - Como os sistemas apresentam o status do sistema de rotulagem?

Oito dos 25 sistemas de rotulagem apresentam seis funcionalidades para comunicar o status do sistema e proporcionar liberdade de uso ao usuário. As funcionalidades são diversificadas e não seguem um padrão. O primeiro são os botões salvar, voltar e avançar que dão ao usuário a liberdade de escolher suas ações no sistema. O curioso é que os botões salvar, voltar e avançar nem sempre são usados juntos nas interfaces do sistema.

A segunda funcionalidade é a confirmação de rotulagem. Após o usuário clicar em salvar rotulagem, o sistema apresenta um pop-up para o usuário confirmar a rotulagem.

A terceira funcionalidade é informar ao usuário quando o último registro foi salvo. O usuário pode entender que o que foi rotulado está sendo capturado e salvo pelo sistema. A quarta estratégia é apresentar o status de rotulagem. Quando o usuário finaliza a rotulagem, o sistema mostra seu status, informando se a rotulagem está em andamento, se foi salva ou se ocorreu algum erro. A quinta estratégia permite que o usuário volte à página principal do sistema. Essa estratégia é importante quando o usuário está perdido e deseja voltar para casa. Por fim, a sexta estratégia refere-se à liberdade do usuário em sair do sistema.

6

ARTEFATO FINAL: PADRÕES

AnnotationUI

Neste capítulo são apresentados os padrões *AnnotationUI*, um conjunto de soluções para apoiar desenvolvedores a projetar a interface de sistemas de rotulagem de texto. Esses padrões foram resultados de um compilado de estudos na literatura, indústria, academia e estado-da-prática. Os padrões contêm soluções e exemplos de componentes de interface importantes para sistemas de rotulagem de texto.

6.1 *AnnotationUI*

Todos os padrões incluem cinco seções: problema, uso, solução, justificativa e exemplos. A seção do problema mostra qual é a tarefa do usuário. A seção de uso descreve quando o desenvolvedor pode usar o padrão. A seção de solução apresenta a solução imediata para o problema por meio de componentes de interface. A justificativa explica por que é essencial usar. Finalmente, os exemplos mostram componentes de interface adaptados dos sistemas encontrados na análise do estado da prática. Esses exemplos foram prototipados no Balsamiq¹ para torná-los mais genéricos e não vinculados a um sistema de rotulagem específico.

Na Tabela 6.1, são apresentadas as referências de cada padrão. Existem 19 padrões, e estes foram divididos em dois grupos principais:

¹ <<https://balsamiq.cloud/>>

- **Padrões de Tarefas de Rotulagem** - Cinco padrões relacionados a um tipo específico de rotulagem. Eles mostram como apresentar os dados e coletar feedback do usuário.
- **Padrões Interface UX** - 14 padrões gerais relacionados à experiência do usuário no sistema de rotulagem. Eles complementam a interface com soluções para trazer uma experiência positiva para o usuário durante a tarefa de rotulagem. Além disso, esses padrões foram divididos em três subcategorias:
 - *Usuário precisa de ajuda*: estes padrões estão relacionados para auxiliar o usuário na utilização da ferramenta e para auxiliar durante a atividade de etiquetagem;
 - *Usuário precisa de liberdade de uso*: esses padrões visam dar ao usuário o controle e a liberdade de fazer o que quiser no sistema;
 - *Usuário precisa de informações*: esses padrões visam informar o usuário sobre o que está acontecendo no sistema.

O desenvolvedor deve escolher um dos padrões da categoria *Labeling Tasks*. Cada padrão está relacionado a um problema de ML. Em seguida, o desenvolvedor deve explorar os padrões de Interface UX para usar o máximo possível. Portanto, esses padrões visam ajudar os usuários nas tarefas de rotulagem.

Nesta seção, são descritos apenas um padrão de cada categoria (Tabelas 6.2, 6.3, 6.4 e 6.5). No entanto, todos eles estão disponíveis no Apêndice A.1 e também na página da web dos padrões *AnnotationUI* ².

² <<https://sites.google.com/view/annotationui>>

Categoria dos padrões	Padrões	Rastreabilidade
Padrões Labeling Task	Text or Document Labeling for a Category	(TKACHENKO et al., 2020-2021), (RECOGNAI, 2021), (DATAQA, 2021), (UBIAI, 2021), (AGHAJANI et al., 2019), (NAKAYAMA et al., 2018), (CHEW et al., 2019), (HUMAN-LAMBDAS, 2021), (FAST..., 2021)
	Text or Document Labeling for Multi-category	(HUMAN-LAMBDAS, 2021), (PRODIGY, 2021), (TKACHENKO et al., 2020-2021), (RECOGNAI, 2021), (DATAQA, 2021), (UBIAI, 2021)
	Entity Labeling Task	(MIROBIT, 2021), (HOU2ZIO, 2021), (RECOGNAI, 2021), (HUMAN-LAMBDAS, 2021), (SAMUELDOBBI, 2021), (TOOL, 2021), (TKACHENKO et al., 2020-2021), (NAKAYAMA et al., 2018), (D5555, 2021), (MEGAGONLABS, 2021), (UBIAI, 2021), (CHEN; STYLER, 2013), (PRODIGY, 2021), (CEJUELA; ROST, 2011), (DATAQA, 2021), (DIFFGRAM, 2021), (YANG et al., 2018), (PROYCON, 2021), (STENETORP et al., 2012)
	Relation Extraction	(KLIE et al., 2018), (TOOL, 2021), (TKACHENKO et al., 2020-2021), (FAST..., 2021), (D5555, 2021), (DATAQA, 2021), (STENETORP et al., 2012), (UBIAI, 2021), (CHEN; STYLER, 2013), (PRODIGY, 2021)
	Batch Labeling	(RECOGNAI, 2021)
UX Interface: User needs help	Tool help	(HEXATOMIC, 2021), (CEJUELA; ROST, 2011), (KLIE et al., 2018), (TKACHENKO et al., 2020-2021), (D5555, 2021), (PRODIGY, 2021), (PROYCON, 2021), (UBIAI, 2021), (CHEW et al., 2019), (PROYCON, 2021), (UBIAI, 2021), (RECOGNAI, 2021)
	External help	(STENETORP et al., 2012), (CEJUELA; ROST, 2011), (TKACHENKO et al., 2020-2021), (SAMUELDOBBI, 2021)
	Word search	(PROYCON, 2021), (DIFFGRAM, 2021)
UX Interface: User needs freedom of use	Undo a task	(CHEN; STYLER, 2013), (CHEW et al., 2019)
	Navigation Between Instances	(MEGAGONLABS, 2021), (MIROBIT, 2021), (AGHAJANI et al., 2019), (TOOL, 2021), (CEJUELA; ROST, 2011), (SAMUELDOBBI, 2021)
	Add category	(PROYCON, 2021; UBIAI, 2021; TKACHENKO et al., 2020-2021)
	Feedback of Lack of Data Domain	(MEGAGONLABS, 2021), (AGHAJANI et al., 2019), (CHEW et al., 2019), (TOOL, 2021), (PRODIGY, 2021)
	Save	(TKACHENKO et al., 2020-2021), (HUMAN-LAMBDAS, 2021), (RECOGNAI, 2021), (DATAQA, 2021), (UBIAI, 2021), (PRODIGY, 2021)
	Cancel	(TKACHENKO et al., 2020-2021), (HUMAN-LAMBDAS, 2021), (RECOGNAI, 2021), (DATAQA, 2021), (UBIAI, 2021), (PRODIGY, 2021)
	Shortcuts	(CEJUELA; ROST, 2011), (D5555, 2021), (PRODIGY, 2021), (AGHAJANI et al., 2019), (TKACHENKO et al., 2020-2021)
UX Interface: User needs information	Tool status	(CEJUELA; ROST, 2011), (DIFFGRAM, 2021), (HUMAN-LAMBDAS, 2021), (RECOGNAI, 2021)
	Labeling Task Progress	(CHEW et al., 2019), (MIROBIT, 2021), (AGHAJANI et al., 2019), (RECOGNAI, 2021)
	Labeling History Overview	(HUMAN-LAMBDAS, 2021), (NAKAYAMA et al., 2018)
	Statistics	(AGHAJANI et al., 2019), (LIGHTTAG, 2021), (FAST..., 2021)

Tabela 6.1 – Categorias dos padrões *AnnotationUI* e as suas rastreabilidades.

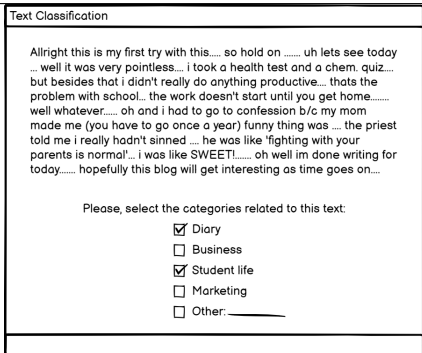
Padrão	Text or Document Labeling for Multi-category
Categoria	Labeling Task
Problema	O usuário precisa visualizar e rotular um documento ou texto em mais de uma categoria.
Uso	Use quando uma tarefa de rotulagem do usuário for visualizar e categorizar um documento ou texto em várias categorias. O usuário pode escolher mais de uma categoria para rotular um único texto.
Solução	<ul style="list-style-type: none"> • A interface deve mostrar o texto ou documento a ser rotulado. • A visualização de texto ou documento pode mostrar as palavras mais importantes ou mais frequentes. • O usuário pode escolher mais de uma categoria para rotular um único texto (instância), e a interface deve exibir as categorias nas possíveis formas: <ul style="list-style-type: none"> – Buttons; – Checkbox; – Dropdown menu. • Opcionalmente, a interface pode mostrar opções para indicar dúvidas sobre os dados de rotulagem; • Outra possibilidade é que a interface mostre uma opção para o usuário indicar uma nova categoria.
Justificativa	O desenvolvedor precisa que o usuário veja o texto ou documento e dê o feedback de rotulagem. O texto categorizado servirá para construir uma base de dados e assim seguir as demais etapas do processo de criação do modelo de ML.
Exemplo	 <p>Exemplo de rotulagem de texto ou documento para padrão de várias categorias. A interface mostra a instância de texto e solicita que o usuário selecione as categorias relacionadas a ela. A interface exibe as classes em um componente <i>checkbox</i> onde o usuário pode escolher mais de um rótulo. Além disso, o usuário pode selecionar a opção "outro" e inserir uma nova categoria.</p>

Tabela 6.2 – Um exemplo de padrão de tarefa de rotulagem: classificação de texto para uma multi-categoria.

A Tabela 6.2 mostra um exemplo de um padrão de tarefa de rotulagem chamado

Text or Document Labeling for Multi-category. Este padrão visa fornecer uma solução quando o usuário precisa visualizar e rotular um documento ou um texto em multi-categoria. Isso significa que o usuário pode escolher mais de uma categoria para rotular um único texto. Os padrões *AnnotationUI* apresentam sugestões de interface, como botões e caixa de seleção, para coletar feedback do usuário. Opcionalmente, o desenvolvedor pode adicionar outros padrões para indicar dúvidas sobre a instância de dados ou indicar uma nova categoria. A lógica por trás desse padrão é desenvolver um conjunto de dados de treinamento para construir um modelo de ML posteriormente. É possível encontrar exemplos deste padrão nos sistemas ([HUMAN-LAMBDAS, 2021](#)), ([PRODIGY, 2021](#)), ([TKACHENKO et al., 2020-2021](#)), ([RECOGNAI, 2021](#)), ([DATAQA, 2021](#)), ([UBIAI, 2021](#)).

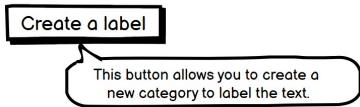
Padrão	Tool Help
Categoria	UX Interface: O usuário precisa de ajuda
Problema	O usuário tem dúvidas sobre as funcionalidades da ferramenta.
Uso	Use quando o usuário precisar de instruções sobre a tarefa de rotulagem ou dúvidas sobre a funcionalidade da ferramenta.
Solução	Deixe as instruções acessíveis e visíveis na interface da ferramenta. A interface deve mostrar uma destas opções: <ul style="list-style-type: none"> • Ícone de ajuda para abrir instruções em um modal ou pop-up; • Dicas de ferramentas em todos os botões principais; • Instruções de rotulagem por um texto ou vídeo;
Justificativa	Ajude os usuários a usar a ferramenta inteiramente sem dúvidas sobre a funcionalidade e ações.
Exemplo	 <p>Dicas de ferramentas em botões com instruções.</p>

Tabela 6.3 – Um exemplo de padrão de interface UX - User needs help: Tool Help.

A Tabela 6.3 apresenta um exemplo de um padrão de Interface UX da subcategoria "O usuário precisa de ajuda", chamada *Tool Help*. Este padrão fornece soluções quando o usuário tem dúvidas sobre as funcionalidades da ferramenta. Assim, é recomendado que a interface mostre pelo menos uma forma de fornecer instruções, como ícones de ajuda ou dicas de ferramentas. A lógica por trás disso é ajudar o usuário a usar a ferramenta sem dúvidas. É possível encontrar exemplos desse padrão nos sistemas (HEXATOMIC, 2021), (CEJUELA; ROST, 2011), (KLIE et al., 2018), (TKACHENKO et al., 2020-2021), (D5555, 2021), (PRODIGY, 2021), (PROYCON, 2021), (UBIAI, 2021), (CHEW et al., 2019), (PROYCON, 2021), (UBIAI, 2021), (RECOGNAI, 2021).

A Tabela 6.4 apresenta um exemplo de um padrão de Interface UX da subcategoria "Usuário precisa de liberdade", chamada *Atalhos*. Este padrão fornece soluções quando o usuário deseja realizar as principais funções do sistema de forma fácil e em menor tempo. Então, é recomendado que a interface forneça atalhos para as principais ações da ferramenta através da barra de atalhos ou atalhos de teclado. A lógica por trás disso é permitir que os usuários rotulem mais rapidamente em menos tempo. É possível encontrar exemplos deste padrão nos sistemas (CEJUELA; ROST, 2011), (D5555, 2021), (PRODIGY, 2021), (AGHAJANI et al., 2019), (TKACHENKO et al., 2020-2021).


Padrão	Shortcuts
Categoria	UX Interface: O usuário precisa de liberdade de uso
Problema	O usuário precisa de atalhos para acessar as principais funções em menos tempo.
Uso	Use quando o usuário quiser fazer as principais ações da ferramenta com facilidade.
Solução	Permitir atalhos para as principais ações do sistema, como: <ul style="list-style-type: none"> • Barra de atalhos com as principais ações da ferramenta; • Atalhos de teclado ou dicas de ferramentas visíveis nos botões principais;
Justificativa	Ajude os usuários a encontrar atalhos para funcionalidades essenciais da ferramenta. Além disso, o usuário pode realizar a rotulagem mais rapidamente em menos tempo.
Exemplo	<p>Please, select one of these categories to label the text:</p>  <p>Atalhos de teclado visíveis para botões de rótulo.</p>

Tabela 6.4 – Um exemplo de padrão de interface UX - User needs freedom of use: Shortcuts.

Na tabela 6.5, é apresentado um exemplo de um padrão de interface UX da subcategoria de informações de necessidades do usuário chamada *Progresso da tarefa de rotulagem*. Esse padrão fornece soluções quando o usuário deseja saber o *status* da tarefa de rotulagem, identificando quantos dados precisam ser rotulados ou quantos dados foram rotulados, ou quanto tempo eles gastam rotulando. É recomendado que a interface forneça mensagens ou imagens para informar o usuário sobre o *status* da rotulagem. A lógica por trás disso é informar aos usuários quanto esforço eles realizaram durante a rotulagem. É possível encontrar exemplos desse padrão nos sistemas (CHEW et al., 2019), (MIROBIT, 2021), (AGHAJANI et al., 2019), (RECOGNAI, 2021).

6.2 Avaliação dos padrões AnnotationUI

Foi realizado um estudo empírico para avaliar a utilidade, facilidade e intenção de usar os padrões AnnotationUI. Os participantes foram profissionais de software com experiência em projetos de desenvolvimento de ML. O objetivo do estudo foi verificar

Padrão	Labeling Task Progress						
Categoria	UX Interface: O usuário precisa de informações						
Problema	O usuário precisa saber o status da tarefa de rotulagem.						
Usage	Use este padrão quando o usuário quiser: <ul style="list-style-type: none"> • Identificar quantos dados ainda precisam ser rotulados • Identificar quantos dados já foram rotulados. • Saber quanto tempo levou na tarefa de rotulagem. 						
Solução	Apresente mensagens ou imagens que informem o usuário sobre a atividade de rotulagem. Os componentes da interface devem relatar uma ou mais destas opções: <ul style="list-style-type: none"> • A quantidade de rotulagem realizada; • O tempo gasto; • A quantidade de dados que precisam ser rotulados; 						
Justificativa	É essencial informar aos usuários quanto esforço eles fizeram durante a rotulagem. Adicionando elementos de interface que notificam o status de andamento da tarefa, o usuário recebe esse feedback.						
Exemplo	<table border="1" data-bbox="464 1081 863 1144"> <thead> <tr> <th>Total instances</th> <th>Instances labeled</th> <th>Instances left</th> </tr> </thead> <tbody> <tr> <td>100</td> <td>47 (47%)</td> <td>53 (53%)</td> </tr> </tbody> </table> <p>Tabela com informações de atividade de rotulagem.</p>	Total instances	Instances labeled	Instances left	100	47 (47%)	53 (53%)
Total instances	Instances labeled	Instances left					
100	47 (47%)	53 (53%)					

Tabela 6.5 – Um exemplo de padrão de interface UX - User needs information: Labeling Task Progress.

como foi a experiência do desenvolvedor ao prototipar uma interface de sistema de rotulagem usando os padrões *AnnotationUI*. Os resultados podem contribuir para aprimorar os padrões *AnnotationUI*.

O estudo teve quatro etapas: a apresentação dos padrões *AnnotationUI*, a explicação de um cenário de uso, a prototipagem da interface e um questionário. Na primeira etapa, foram apresentados os desafios de projetar um sistema de rotulagem de dados e os padrões *AnnotationUI*. Foi explicado como foi o processo de identificação dos padrões e como usá-los na página da *Web* do *AnnotationUI* ³.

Na segunda etapa, um cenário de uso foi apresentado para ajudar o participante a entender como usar os padrões *AnnotationUI*. A apresentação deste cenário também é

³ <<https://sites.google.com/view/annotationui>>

necessária para possibilitar a execução da etapa de prototipagem. O cenário apresentado aos participantes foi o seguinte:

Você conseguiu um emprego! Você passou por um processo de seleção de uma empresa BigTech para uma posição de engenheiro de front-end em uma equipe de soluções de ML. Essa equipe é responsável por desenvolver sistemas para apoiar o desenvolvimento de projetos de ML, como sistemas de manipulação de dados, monitoramento de modelos e rotulagem de dados. No seu primeiro dia de trabalho, seu supervisor o desafiou a ajudar a resolver um problema para um novo projeto de ML. Este projeto consistiu em construir um modelo de ML para classificar automaticamente os comentários de vídeo em três classes: ódio, felicidade e opinião. No entanto, antes de criar o modelo, é necessário rotular manualmente uma base de comentários nessas três classes. Portanto, seu desafio é desenvolver a interface para um sistema de rotulagem de comentários em vídeo. Além disso, para te ajudar, você terá o suporte dos padrões AnnotationUI.

Após descrever e apresentar o cenário de uso, foi solicitado aos participantes que compartilhasse sua tela com a página da Web AnnotationUI aberta. Eles teriam que começar a explorar os padrões da Tarefa de Rotulagem para decidir qual era o padrão que poderia resolver o problema do cenário. Após escolher o padrão correto conforme o problema de dado no cenário, os participantes puderam começar a prototipagem na ferramenta Balsamiq.

Foi solicitado que os participantes se expressassem pensando em voz alta enquanto prototipavam no Balsamiq e exploravam a página da AnnotationUI. Foi solicitado aos participantes que explorassem cada padrão enquanto prototipavam a interface. Eles eram livres para escolher que padrão queriam. Além disso, foi ressaltado que a avaliação seria dos padrões AnnotationUI e não dos participantes.

Depois que os participantes declararam que terminaram seus protótipos, foi iniciado a etapa final, o *feedback* dos padrões por meio de um questionário. Foi solicitado aos participantes que respondessem honestamente a cada pergunta porque isso serviria como base para aprimorar os padrões AnnotationUI. O questionário foi respondido mediante uma entrevista. O questionário respondido pelos participantes foi o seguinte:

- *Os padrões foram úteis ou não? Justifique*
- *Os padrões são fáceis ou difíceis de usar? Justificar.*
- *Você usaria ou não padrões em um cenário futuro? Justificar.*
- *O que pode ser melhorado nos padrões?*
- *O que você não entendeu?*
- *Como foi sua experiência com os padrões AnnotationUI?*

Com o questionário acima foi visado identificar a perspectiva do participante sobre a facilidade, utilidade e pretensão de uso em cenário futuro. Além disso, foram realizadas perguntas diretamente aos participantes sobre suas dificuldades, experiência e percepção do que pode ser melhorado nos padrões.

6.2.1 Perfil dos participantes

Quatro praticantes de ML foram convidados para realizar uma sessão virtual de uma hora com cada um. Antes de iniciar uma seção com um participante, foi enviado um formulário para pedir consentimento e sua experiência com ML e prototipagem. Além disso, foi enviado a eles um *link* para o *Balsamiq Cloud* e um *link* para um pequeno tutorial para saber como usá-lo. As perguntas realizadas sobre a experiência em ML dos participantes foram as seguintes:

- *Você já desenvolveu protótipos navegáveis?*
- *Você já usou a ferramenta Balsamiq?*
- *Há quanto tempo você trabalha com ML?*
- *De quantos projetos de ML você participou?*
- *Quais etapas você trabalhou nesses projetos (compreensão do problema, coleta de dados, rotulagem de dados, manipulação de dados, desenvolvimento de modelos, monitoramento de modelos)?*

ID	Experiência em prototipação	Experiência no Balsamiq	Experiência em ML (em anos)	Projetos ML (participação)	Etapas de desenvolvimento ML (experiência)
P1	Sim	Não	Mais do que 5	4	Compreensão do Problema, Coleta de Dados, Rotulagem de Dados, Manipulação de Dados, Desenvolvimento de Modelo e Monitoramento de Modelo
P2	Sim	Não	4	8	Compreensão do Problema, Coleta de Dados, Rotulagem de Dados, Manipulação de Dados, Desenvolvimento de Modelo e Monitoramento de Modelo
P3	Sim	Não	2	3	Coleta de Dados, Rotulagem de Dados, Manipulação de Dados e Monitoramento de Modelo
P4	Sim	Não	4	3	Coleta de Dados Rotulagem de Dados Manipulação de Dados

Tabela 6.6 – Perfil dos participantes

Analisando as suas respostas (consulte a Tabela 6.6) observa-se que todos os participantes têm experiência em rotulagem de dados em projetos reais de ML e prototipagem. No entanto, todos eles nunca usaram a ferramenta Balsamiq. Os participantes P1 e P2 foram os mais experientes. Ambos trabalharam em todas as etapas de um projeto de desenvolvimento de ML. P1 tinha mais de cinco anos de experiência em projetos de ML e P2 trabalhou em oito projetos de ML em diferentes domínios. P3 e P4 trabalharam em três projetos de ML em dois e quatro anos.

6.2.2 Resultados e Discussão

Foi realizado uma análise qualitativa com base nas gravações dos *feedbacks* dos participantes. Foi analisado a facilidade de uso e o uso pretendido dos padrões, seus aspectos a serem melhorados e a experiência do desenvolvedor ao usá-los.

Em relação à facilidade de uso e intenção de uso, foi observado que os padrões ajudam os participantes a projetar a interface dos sistemas de rotulagem. Alguns comentários sobre os aspectos positivos do uso dos padrões são:

P1 - "Posso expor os padrões da maneira que quiser; eles são intuitivos."

P1- "Eles foram 100% úteis. Não vejo como não usar alguns dos padrões aqui."

P3- "Sempre gosto quando há essa barra de progresso porque aconteceu de eu estar preenchendo um formulário no Google com muitas abas e não conseguir saber quanto tempo vai demorar."

P4 - "Achei fácil de usar por causa de sua documentação - informando como os padrões funcionam e o propósito. Acho que a documentação o torna mais acessível."

Com as falas dos participantes, foram observados pontos positivos quanto ao uso e a intenção de uso. O participante P1 destaca que não consegue imaginar a não utilização de alguns padrões nas interfaces desses sistemas; ele mostra como eles fornecem contexto para a especificidade dos sistemas de rotulagem.

Também foram identificados aspectos negativos do uso dos padrões:

P1 - Gosto da ideia" (sobre o status da ferramenta padrão - exemplo 1), "mas não gosto na prática porque o usuário terá que OK todos os dados."

No exemplo 1 citado pelo participante P1 do padrão de *status* da ferramenta, é apresentada uma solução onde a interação do usuário é confirmar a rotulagem por meio de um *pop-up*. O participante P1 destaca que essa atividade pode se tornar chata para o usuário na prática. É importante destacar esse aspecto negativo, pois o principal objetivo dos padrões é ajudar os desenvolvedores a projetar interfaces mais utilizáveis que proporcionem uma UX positiva para os usuários. Portanto, uma nota das informações nos padrões foi adicionada, informando essa limitação dos mesmos.

P2 - "Vou colocar uma lista de palavras e um botão para rotular automaticamente todas as palavras exatas no texto."

Nesse caso, o participante P2 estava usando os atalhos e o padrão de pesquisa de palavras quando teve a ideia de criar uma funcionalidade para pesquisar e rotular as palavras exatas em um texto. Essa funcionalidade ajuda o usuário a economizar tempo durante a tarefa de rotulagem. No entanto, pode ser um recurso que gera rotulagem incorreta se o usuário não revisar a instância de texto completamente.

O participante P3 comentou:

P3 - "Eu colocaria um texto dinâmico como se o usuário tivesse no início da rotulagem, eu colocaria um "continue". Se o usuário estava na metade "você já está na metade" e quando estava quase terminando "não é longo", "vai dar certo", uma mistura dos exemplos de padrões."

O participante P3 conseguiu criar sua solução a partir dos exemplos dos padrões, o que mostra que usar exemplos por imagens pode ajudar a estimular a criatividade do desenvolvedor.

Os padrões também ajudam os participantes a pensar em componentes de interfaces que podem ser úteis para o usuário:

P1 - "Eu colocaria botões. Um botão salvar (enter) e outro anterior (backspace). Colocar um checkbox para o usuário marcar para salvar a ação - acho desnecessário, mas é melhor colocar."

P4 - "Eu usaria os padrões para criar um sistema de rotulagem, mas se eu tivesse algo muito específico que o padrão não cobrisse, eu criaria um novo baseado no que já existe."

Embora P1 não ache importante apresentar opções de atalho na interface, o padrão ajudou o desenvolvedor a se sentir como se fosse o usuário e decidiu colocar opções de atalho na interface. O desenvolvedor deve se colocar no lugar do usuário, pois o usuário é quem vai utilizar o sistema.

Além dos aspectos positivos, os participantes também falaram sobre os aspectos negativos e confusos dos padrões:

P2 - "No exemplo 1 do padrão de rotulagem de texto ou documento para um padrão de categoria há palavras destacadas em verde, não sei como ele decidiu isso."

Apesar dos aspectos negativos, os padrões *AnnotationUI* podem ajudar os desenvolvedores durante o *design* de uma interface de sistema de rotulagem textual. Os praticantes também afirmaram que os padrões economizam tempo durante a concepção do projeto porque não precisam gastar horas procurando aplicativos semelhantes. Além disso, os padrões ajudam na UX da interface por causa dos exemplos baseados em sistemas de rotulagem de dados reais.

6.3 Ameaças à Validade

A análise das ameaças à validade deste estudo seguiram a recomendação de Wohlin (WOHLIN et al., 2012). Uma ameaça crítica de validade nesta pesquisa seria a possibilidade da criação mal projetada dos padrões *AnnotationUI*. Para corroborar isso, os padrões foram submetidos a duas avaliações de diferentes perspectivas para mitigar essa ameaça. Antes de executar cada avaliação, um estudo piloto foi conduzido para solucionar problemas e aprimorar os padrões *AnnotationUI*.

Sobre a avaliação dos padrões *AnnotationUI*, uma possível ameaça seria que a avaliação dos padrões não fosse realizada por participantes alvos. Para isso, o estudo foi conduzido com participantes que tinham contato, de alguma forma, com o desenvolvimento de modelos ML. Os participantes do estudo de avaliação dos padrões tinham contato com mais de três projetos de ML diferentes em todas as etapas de desenvolvimento, o que significa que devem conhecer os desafios da equipe de ML em diferentes contextos. Além disso, todos os participantes conseguiram utilizar os padrões e construir seus protótipos. Adicionalmente, os participantes concordaram que os padrões são úteis para ajudar os desenvolvedores a projetar sistemas de rotulagem de dados textuais.

Outra ameaça refere-se aos dados qualitativos como medida para verificar se os padrões suportam a questão de pesquisa. No entanto, foi descoberto por experiência anterior que estudos empíricos com questões abertas agregaram mais conhecimento a esta pesquisa. Assim, uma hora de sessão foi adicionada a cada praticante para obter dados mais qualitativos sobre sua perspectiva sobre os padrões.

O *GitHub* foi escolhido como a principal fonte desta análise do estado da prática por motivos fundamentais. A primeira razão é que ele é o repositório de código-fonte mais popular e significativo na comunidade de praticantes. Além disso, os projetos são de código aberto, portanto, estão disponíveis para download e execução gratuitamente. Não foram considerados os sistemas pagos, pois a maioria deles necessita de acesso corporativo com alto custo financeiro.

Outras ameaças à validade estão relacionadas ao uso limitado dos padrões *AnnotationUI*, uma vez que o estudo limitou os participantes no projeto de sistemas de rotulagem de texto em um cenário específico. No entanto, todos os participantes são praticantes experientes que trouxeram boas percepções sobre o uso dos padrões.

7

CONSIDERAÇÕES FINAIS

Neste capítulo são apresentadas as conclusões desta dissertação, resumindo sua motivação, proposta e apresentando as suas contribuições, limitações e trabalhos futuros.

Dado o contexto do grande volume de dados sendo gerado e o avanço do ML para aprender a automatizar tarefas a partir desses dados, houve um aumento na demanda por sistemas de apoio às etapas de desenvolvimento e monitoramento dos modelos de ML. Em alguns problemas de ML supervisionado, há a necessidade da etapa de rotulagem de dados por usuários especialistas e assim, surge o desafio de projetar sistemas de rotulagem para esses usuários.

Ao projetar sistemas de rotulagem, é necessário considerar qual é o tipo de dado a ser rotulado e qual problema de ML o sistema irá atender. Para cada tipo de dado e para cada problema de ML relacionado, há diferentes formas de interação e visualização de dados.

Este trabalho apresenta os padrões *AnnotationUI* para apoiar o projeto de interface de sistemas de rotulagem de texto. Os padrões foram resultados de uma pesquisa, tendo como primeiro artefato as guidelines *LabelUX*.

A motivação dessa proposta surgiu a partir da necessidade de ter instruções para projetar os sistemas de rotulagem de texto, visando construir uma interface com boa usabilidade para os rotuladores.

Os resultados obtidos indicam que os padrões respondem a QP (Seção 1.3) deste trabalho. Contudo, as limitações desta solução estão relacionadas a avaliações aplicadas somente na perspectiva dos desenvolvedores e não de rotuladores, utilizando o produto

final. Além disso, este trabalho é limitado a sistemas de rotulagem de texto *open-source*.

7.1 Contribuições

As principais contribuições desta pesquisa foram:

- Construção dos padrões *AnnotationUI* para projetos de interface de sistemas de rotulagem de texto. Os padrões estão disponíveis publicamente para visualização via página WEB¹.
- Disseminação dos resultados obtidos nos estudos. Os resultados obtidos descritos nesta qualificação foram submetidos para publicação em eventos. A seguir é apresentada a lista de artigos publicados:
 1. O primeiro artigo, referente ao Capítulo 3, foi submetido e aceito no *Workshop de Engenharia de Requisitos (WER 2021)* com o título *Rotule-me! Uma experiência de engenharia de requisitos para um sistema de rotulagem*.
 2. O segundo artigo, referente ao Capítulo 4, foi submetido e aceito no *Simpósio Brasileiro de Qualidade de Software (SBQS 2021)* na Trilha de Trabalhos Técnicos com o título *LabelUX! Guidelines to support software engineers to design data labeling systems*.
 3. O terceiro artigo, referente ao estudo de avaliação de apoio à Usabilidade do Capítulo 4, foi submetido e aceito no *Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais (IHC 2021)* na Trilha de Ideias Inovadoras e Resultados Emergentes com o título *Applying Usability Heuristics in the Context of Data Labeling Systems*.

7.2 Trabalhos Futuros

Com relação aos trabalhos futuros, esta pesquisa pode servir como base de metodologia para a criação de novos padrões de interface de sistema de rotulagem para os tipos de

¹ <<https://sites.google.com/view/annotationui>>

dados de imagem, áudio e vídeo.

Além disso, futuras pesquisas podem ser realizadas para avaliação de UX de sistemas construídos com os padrões *AnnotationUI*. Estas avaliações colocariam o rotulador como papel principal nos estudos, podendo analisar tanto a UX quanto a produtividade em comparação com outros sistemas.

Outro trabalho que poderia ser estendido seria a análise de sistemas de rotulagem que não são *open-source*. A análise de sistemas pagos poderia contribuir para atualização dos padrões ou até mesmo a criação de novos padrões.

Por fim, esse trabalho pode contribuir em pesquisas de avaliação de projetos de interface tendo como base os padrões *AnnotationUI*. Assim, os padrões poderiam ser usados para averiguar a qualidade de outros sistemas de rotulagem de texto.

REFERÊNCIAS

- ACUNA, D. et al. Efficient interactive annotation of segmentation datasets with polygon-rnn++. In: *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2018. p. 859–868. 36
- AGHAJANI, E. et al. Software documentation issues unveiled. In: IEEE. *2019 IEEE/ACM 41st International Conference on Software Engineering (ICSE)*. [S.l.], 2019. p. 1199–1210. 109, 113, 120, 123, 124
- ALPAYDIN, E. *Introduction to Machine Learning*. [S.l.]: MIT Press, 2020. 22, 42
- AMERSHI, S. et al. Software engineering for machine learning: A case study. In: IEEE. *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*. [S.l.], 2019. p. 291–300. 21, 22
- AMERSHI, S. et al. Power to the people: The role of humans in interactive machine learning. *Ai Magazine*, v. 35, n. 4, p. 105–120, 2014. 21, 22, 34, 38, 43
- AMERSHI, S. et al. Guidelines for human-ai interaction. In: *Proceedings of the 2019 chi conference on human factors in computing systems*. [S.l.: s.n.], 2019. p. 1–13. 21
- AWS, A. W. S. *Amazon SageMaker Ground Truth*. 2021. Acesso em 03 de maio de 2021. Disponível em: <<https://docs.aws.amazon.com/sagemaker/latest/dg/sms-task-types.html>>. 14, 21, 36, 58, 62
- BORCHERS, J. O. A pattern approach to interaction design. In: *Cognition, Communication and Interaction*. [S.l.]: Springer, 2008. p. 114–131. 29
- CAI, L.; ZHU, Y. The challenges of data quality and data quality assessment in the big data era. *Data science journal*, Ubiquity Press, v. 14, 2015. 20
- CASTILHO, R. Eckart de et al. A web-based tool for the integrated annotation of semantic and syntactic structures. In: *Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH)*. Osaka, Japan: The COLING 2016 Organizing Committee, 2016. p. 76–84. Disponível em: <<https://www.aclweb.org/anthology/W16-4011>>. 101
- CEJUELA, J. M.; ROST, B. tagtog: collaborative interactive semi-supervised learning and annotation web-based framework. 2011. 107, 113, 115, 116, 120, 123

- CHEN, C. et al. From lost to found: Discover missing ui design semantics through recovering missing tags. *Proceedings of the ACM on Human-Computer Interaction*, ACM New York, NY, USA, v. 4, n. CSCW2, p. 1–22, 2020. 30
- CHEN, W.-T.; STYLER, W. Anafora: A web-based general purpose annotation tool. In: *Proceedings of the 2013 NAACL HLT Demonstration Session*. Association for Computational Linguistics, 2013. p. 14–19. Disponível em: <<http://www.aclweb.org/anthology/N13-3004>>. 112, 113, 120
- CHEW, R. et al. Smart: an open source data labeling platform for supervised learning. *The Journal of Machine Learning Research*, JMLR.org, v. 20, n. 1, p. 2999–3003, 2019. 113, 114, 115, 120, 123, 124
- CHOI, M. et al. Aila: Attentive interactive labeling assistant for document classification through attention-based deep neural networks. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. [S.l.: s.n.], 2019. p. 1–12. 9, 34, 35, 62
- CHOPRA, A.; PRASHAR, A.; SAIN, C. Natural language processing. *International journal of technology enhancements and emerging engineering research*, Citeseer, v. 1, n. 4, p. 131–134, 2013. 36
- COHEN, J. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, Sage Publications Sage CA: Thousand Oaks, CA, v. 20, n. 1, p. 37–46, 1960. 106
- COSENTINO, V.; IZQUIERDO, J. L. C.; CABOT, J. A systematic mapping study of software development with github. *IEEE Access*, IEEE, v. 5, p. 7173–7192, 2017. 27
- COWLEY, N.; WESSON, J. L. An experiment to measure the usefulness of patterns in the interaction design process. In: SPRINGER. *IFIP Conference on Human-Computer Interaction*. [S.l.], 2005. p. 1142–1145. 29
- CUI, W. Visual analytics: A comprehensive overview. *IEEE Access*, IEEE, v. 7, p. 81555–81573, 2019. 35
- D5555. Tageditor - annotation tool for spacy. 2021. Disponível em: <<https://github.com/d5555/TagEditor>>. 101, 112, 113, 115, 120, 123
- DATAQA. *Labelling platform for text using weak supervision*. 2021. Disponível em: <<https://github.com/dataqa/dataqa>>. 112, 113, 120, 122
- DEARDEN, A.; FINLAY, J. Pattern languages in hci: A critical review. *Human-computer interaction*, Taylor & Francis, v. 21, n. 1, p. 49–102, 2006. 29
- DERNONCOURT, F.; LEE, J. Y.; SZOLOVITS, P. NeuroNER: an easy-to-use program for named-entity recognition based on neural networks. *Conference on Empirical Methods on Natural Language Processing (EMNLP)*, 2017. 101
- DIFFGRAM. *Complete training data platform for machine learning delivered as a single application*. 2021. Disponível em: <<https://github.com/diffgram/diffgram>>. 113, 120
- DIS, E. A. V. et al. Chatgpt: five priorities for research. *Nature*, Nature Publishing Group UK London, v. 614, n. 7947, p. 224–226, 2023. 20

DUDHABAWARE, R. S.; MADANKAR, M. S. Review on natural language processing tasks for text documents. In: IEEE. *2014 IEEE International Conference on Computational Intelligence and Computing Research*. [S.l.], 2014. p. 1–5. 37

DUDLEY, J. J.; KRISTENSSON, P. O. A review of user interface design for interactive machine learning. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, ACM New York, NY, USA, v. 8, n. 2, p. 1–37, 2018. 21, 62, 63

DUDLEY, J. J.; KRISTENSSON, P. O. A review of user interface design for interactive machine learning. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, ACM New York, NY, USA, v. 8, n. 2, p. 1–37, 2018. 22, 38, 39, 43

FAST: Fast Annotation tool for SmarT devices. 2021. Open source software available from <https://github.com/CyberAgent/fast-annotation-tool>. Disponível em: <<https://github.com/CyberAgent/fast-annotation-tool>>. 113, 114, 120

FELIX, C.; DASGUPTA, A.; BERTINI, E. The exploratory labeling assistant: Mixed-initiative label curation with large document collections. In: *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. [S.l.: s.n.], 2018. p. 153–164. 42, 62, 76

FLACH, P. *Machine learning: the art and science of algorithms that make sense of data*. Cambridge, United Kingdom: Cambridge University Press, 2012. 32, 33

GHANIM, T. M.; NABIL, A. M. Offline signature verification and forgery detection approach. In: IEEE. *2018 13th International Conference on Computer Engineering and Systems (ICCES)*. [S.l.], 2018. p. 293–298. 20

GONZALES, G. R. a.; HORITA, F. Supporting visual analytics in decision support system: A systematic mapping study. In: . New York, NY, USA: Association for Computing Machinery, 2020. ISBN 9781450381727. 22

Governo do Estado de São Paulo. *O que é e como é investido o ICMS*. 2021. <<https://portal.fazenda.sp.gov.br/Noticias/Paginas/O-que-%C3%A9-e-como-%C3%A9-investido-o-ICMS.aspx>>. Acesso em 23 de fevereiro de 2021. 41

GROSMAN, J. S. et al. Eras: Improving the quality control in the annotation process for natural language processing tasks. *Information Systems*, Elsevier, v. 93, p. 101553, 2020. 34, 36

HASSENZAHN, M. The thing and i: understanding the relationship between user and product. In: *Funology 2*. [S.l.]: Springer, 2018. p. 301–313. 30

HEBBALAGUPPE, R. et al. How interaction methods affect image segmentation: user experience in the task. In: IEEE. *2013 1st IEEE Workshop on User-Centered Computer Vision (UCCV)*. [S.l.], 2013. p. 19–24. 22, 34

HEXATOMIC. *Hexatomic is an extensible software for deep multi-layer annotation of linguistic corpora*. 2021. Disponível em: <<https://github.com/hexatomic/hexatomic>>. 113, 120, 123

HOU2ZIO. *Minimal customization of Quill.js Rich Text Editor for easy annotation of text snippets for NER model training with spaCy*. 2021. Disponível em: <https://github.com/hou2zi0/minimal-RTE__ner-training-data>. 110, 113, 120

- HUANG, S.-W. et al. Leveraging the crowd to improve feature-sentiment analysis of user reviews. In: *Proceedings of the 2013 international conference on Intelligent user interfaces*. [S.l.: s.n.], 2013. p. 3–14. [22](#)
- HUMAN-LAMBIDAS. *Open Source Human in the Loop platform for anyone to run their own private Mechanical Turk*. 2021. Disponível em: <https://github.com/Human-Lambdas/human-lambdas>. [109](#), [113](#), [120](#), [122](#)
- IO falcony. *Use ML-Annotate to label data for machine learning purpose*. 2017. Disponível em: <https://github.com/falcony-io/ml-annotate>. [101](#)
- ISO. *ISO 9241-210: 2019 Ergonomics of human-system interaction—Part 210: Human-centred design for interactive systems*. [S.l.]: AFNOR Paris, 2019. <https://www.iso.org/standard/77520.html>. [30](#)
- KIM, B. et al. ibcm: Interactive bayesian case model empowering humans via intuitive interaction. 2015. [43](#)
- KIM, B. et al. Scalable and interpretable data representation for high-dimensional complex data. 2015. [36](#)
- KITCHENHAM, B. A.; BUDGEN, D.; BRERETON, P. *Evidence-based software engineering and systematic reviews*. [S.l.]: CRC press, 2015. v. 4. [99](#), [101](#)
- KLIE, J.-C. et al. The INCEpTION platform: Machine-assisted and knowledge-oriented interactive annotation. In: *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*. Santa Fe, New Mexico: [s.n.], 2018. p. 5–9. Disponível em: <https://www.aclweb.org/anthology/C18-2002>. [101](#), [109](#), [112](#), [113](#), [115](#), [120](#), [123](#)
- KNAEBLE, M.; NADJ, M.; MAEDCHE, A. Oracle or teacher? a systematic overview of research on interactive labeling for machine learning. In: *Wirtschaftsinformatik (Zentrale Tracks)*. [S.l.: s.n.], 2020. p. 2–16. [21](#), [34](#)
- LEIVA, L. A.; HOTA, A.; OULASVIRTA, A. Enrico: A dataset for topic modeling of mobile ui designs. In: *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*. [S.l.: s.n.], 2020. p. 1–4. [30](#)
- LIGHTTAG. *In browser document classification with active learning and full text search*. 2021. Disponível em: <https://github.com/LightTag/ylabel2>. [109](#), [113](#), [114](#), [120](#)
- LIU, T. F. et al. Learning design semantics for mobile apps. In: *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. [S.l.: s.n.], 2018. p. 569–579. [30](#)
- MALIK, Z. H.; MUNIR, T.; ALI, M. Ui design patterns for flight reservation websites. In: SPRINGER. *Future of Information and Communication Conference*. [S.l.], 2020. p. 310–327. [30](#)
- MANNIO, M.; NIKULA, U. Requirements elicitation using a combination of prototypes and scenarios. In: CITESEER. *WER*. [S.l.], 2001. p. 283–296. [42](#)
- MARX, V. The big challenges of big data. *Nature*, Nature Publishing Group, v. 498, n. 7453, p. 255–260, 2013. [20](#)

- MEGAGONLABS. *Data programming by demonstration for information extraction and span annotation*. 2021. Disponível em: <<https://github.com/megagonlabs/tagruler>>. 113, 114, 120
- MICROSOFT. *Universal Human Relevance System (UHRS)*. 2021. <<https://prod.uhrs.playmsn.com/uhrs/>>. Acesso em 03 de maio de 2021. 63
- MIROBIT. *Anonymize and annotate text*. 2021. Disponível em: <<https://github.com/Mirobit/Labelit>>. 120, 124
- MITCHELL, T. et al. Machine learning. *Annual review of computer science*, Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA, v. 4, n. 1, p. 417–433, 1990. 32
- MONZÓN, I.; ANGELERI, P.; DÁVILA, A. Design techniques for usability in m-commerce context: A systematic literature review. In: SPRINGER. *International Conference on Software Process Improvement*. [S.l.], 2020. p. 305–322. 30
- MUNZNER, T. *Visualization analysis and design*. [S.l.]: CRC press, 2014. 22, 62
- NADJ, M. et al. Power to the oracle? design principles for interactive labeling systems in machine learning. *KI-Künstliche Intelligenz*, Springer, p. 1–12, 2020. 21, 22, 36, 38, 39, 43
- NAKAYAMA, H. et al. doccano: Text annotation tool for human. 2018. Software available from <https://github.com/doccano/doccano>. Disponível em: <<https://github.com/doccano/doccano>>. 9, 38, 101, 113, 115, 120
- NASCIMENTO, E. et al. Understanding development process of machine learning systems: Challenges and solutions. In: IEEE. *2019 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*. [S.l.], 2019. p. 1–6. 9, 33, 34
- NIELSEN, J. *Usability engineering*. [S.l.]: Morgan Kaufmann, 1994. 26, 60, 79
- NIELSEN, J.; LORANGER, H. *Prioritizing Web Usability*. USA: New Riders Publishing, 2006. ISBN 0321350316. 30
- OTTER, D. W.; MEDINA, J. R.; KALITA, J. K. A survey of the usages of deep learning for natural language processing. *IEEE Transactions on Neural Networks and Learning Systems*, IEEE, v. 32, n. 2, p. 604–624, 2020. 21, 37
- PAPAGIANNPOULOU, E.; TSOUMAKAS, G. A review of keyphrase extraction. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Wiley Online Library, v. 10, n. 2, p. e1339, 2020. 37
- PASSOS, L. et al. Rotule-me! uma experiência de engenharia de requisitos para um sistema de rotulagem. In: *24th Workshop on Requirements Engineering*. [S.l.: s.n.], 2021. 25, 40, 60
- PASSOS, L. C. et al. Labelux! guidelines to support software engineers to design data labeling systems. In: *XX Brazilian Symposium on Software Quality*. [S.l.: s.n.], 2021. p. 1–10. 26, 74

PLUMMER, B. A. et al. Give me a hint! navigating image databases using human-in-the-loop feedback. In: IEEE. *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. [S.l.], 2019. p. 2048–2057. 43

PORTER, R.; THEILER, J.; HUSH, D. Interactive machine learning in data exploitation. *Computing in Science & Engineering*, IEEE, v. 15, n. 5, p. 12–20, 2013. 36

POURSABZI-SANGDEH, F. et al. Alto: Active learning with topic overviews for speeding label induction and document labeling. In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. [S.l.: s.n.], 2016. p. 1158–1169. 62, 63, 76

PRODIGY. *Prodigy: Annotation Tool for AI, Machine Learning & NLP*. 2021. <<https://prodi.gy>>. Acesso em 03 de maio de 2021. 63, 107, 112, 113, 115, 120, 122, 123

PROYCON. *FoLiA Linguistic Annotation Tool - Flat is a web-based linguistic annotation environment based around the FoLiA*. 2021. Disponível em: <<https://github.com/proycon/flat>>. 113, 116, 120, 123

Receita Federal - Ministério da Economia. *NCM - Nomenclatura Comum do Mercosul*. 2021. <<https://receita.economia.gov.br/orientacao/aduaneira/classificacao-fiscal-de-mercadorias/ncm>>. Acesso em 23 de fevereiro de 2021. 41

RECOGNAI. *Rubrix*. 2021. Disponível em: <<https://github.com/recognai/rubrix>>. 109, 113, 114, 120, 122, 123, 124

REED, P. et al. User interface guidelines and standards: progress, issues, and prospects. *Interacting with Computers*, Oxford University Press Oxford, UK, v. 12, n. 2, p. 119–142, 1999. 29

ROH, Y.; HEO, G.; WHANG, S. E. A survey on data collection for machine learning: a big data-ai integration perspective. *IEEE Transactions on Knowledge and Data Engineering*, IEEE, 2019. 42

RUSSELL, S.; NORVIG, P. *Artificial intelligence: a modern approach*. 2002. 32

SAMUELDOBBI. *A web-based text annotation tool, powered by active learning*. 2021. Disponível em: <<https://github.com/samueldobbie/markup>>. 113, 115, 120

Secretaria da Fazenda. *Portal da Nota Fiscal Eletrônica - Informações sobre Estatísticas*. 2021. Disponível em: <<https://www.nfe.fazenda.gov.br/portal/infoEstatisticas.aspx>>. Acesso em 22 de fevereiro de 2021. 41

STENETORP, P. et al. Brat: a web-based tool for nlp-assisted text annotation. In: *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*. [S.l.: s.n.], 2012. p. 102–107. 36, 112, 113, 120

STENETORP, P. et al. Bionlp shared task 2011: Supporting resources. In: *Proceedings of BioNLP Shared Task 2011 Workshop*. Portland, Oregon, USA: Association for Computational Linguistics, 2011. p. 112–120. Disponível em: <<http://www.aclweb.org/anthology/W11-1816>>. 101

STONE, P.; VELOSO, M. Multiagent systems: A survey from a machine learning perspective. *Autonomous Robots*, Springer, v. 8, n. 3, p. 345–383, 2000. 33

THOMAS, J.; COOK, K. *Illuminating the Path: Research and Development Agenda for Visual Analytics*. 2005. 36

TKACHENKO, M. et al. *Label Studio: Data labeling software*. 2020–2021. <<https://labelstud.io/templates/>>. Open source software available from <https://github.com/heartexlabs/label-studio>. 14, 58, 63, 101, 109, 112, 113, 115, 120, 122, 123

TOOL, U. D. *Universal Data Tool*. 2021. Disponível em: <<https://github.com/UniversalDataTool/universal-data-tool>>. 112, 113, 120

UBIAI. *Create Training Data for Your NLP Model*. 2021. Disponível em: <<https://ubiai.tools/>>. 107, 112, 113, 115, 120, 122, 123

VIANA, L.; OLIVEIRA, E.; CONTE, T. An interface design catalog for interactive labeling systems. In: *Proceedings of the 23rd International Conference on Enterprise Information Systems (ICEIS 2021) - Volume 2*. [S.l.: s.n.], 2021. p. 483–494. 22, 25, 34

VIANA, L. et al. Applying usability heuristics in the context of labeling systems. In: *XX Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais*. [S.l.: s.n.], 2021. 26, 81

VICTORELLI, E. Z.; REIS, J. C. D. Human-data interaction design guidelines for visualization systems. In: *Proceedings of the 19th Brazilian Symposium on Human Factors in Computing Systems*. [S.l.: s.n.], 2020. p. 1–10. 21

WALLACE, B. C. et al. Deploying an interactive machine learning system in an evidence-based practice center: abstrackr. In: *Proceedings of the 2nd ACM SIGHIT international health informatics symposium*. [S.l.: s.n.], 2012. p. 819–824. 43

WARD, M. O.; GRINSTEIN, G.; KEIM, D. *Interactive data visualization: foundations, techniques, and applications*. [S.l.]: CRC press, 2010. 35

WOHLIN, C. et al. *Experimentation in software engineering*. [S.l.]: Springer Science & Business Media, 2012. 24, 131

YANG, J. et al. Yedda: A lightweight collaborative text span annotation tool. 2018. Disponível em: <<http://aclweb.org/anthology/P18-4006>>. 101, 109, 113, 120

YUAN, J. et al. A survey of visual analytics techniques for machine learning. *Computational Visual Media*, Springer, p. 1–34, 2020. 21, 34, 35

ZHOU, L. et al. Machine learning on big data: Opportunities and challenges. *Neurocomputing*, Elsevier, v. 237, p. 350–361, 2017. 20

A

Padrões AnnotationUI

Este capítulo descreve todos os padrões do *AnnotationUI*.

A.1 Padrões *AnnotationUI*

A.1.1 Padrão Text or Document Labeling for a Category

Problema. O usuário precisa visualizar e rotular um documento ou texto em uma categoria.

Uso. Use quando uma tarefa de rotulagem do usuário for visualizar e categorizar um documento ou texto em apenas uma categoria. Isso significa que o usuário pode escolher uma entre várias categorias para rotular um texto ou documento.

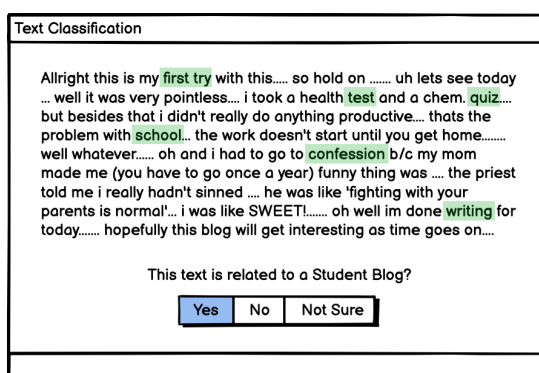
Solução.

- A interface deve mostrar o texto ou documento a ser rotulado.
- A visualização de texto ou documento pode mostrar as palavras mais importantes ou mais frequentes (Figura [A.1](#)).
- Para fornecer feedback sobre a rotulagem, o usuário pode escolher apenas uma categoria para rotular um único texto (instância), e a interface deve exibir as categorias de uma destas formas possíveis:
 - *Buttons* (Figura [A.1](#));
 - *Dropdown menu*;
 - *Radio button*;

- Drag to an option;
- Opcionalmente, a interface pode mostrar opções para indicar dúvidas sobre rotulagem de dados (botão *not sure* na Figura A.1 - mais detalhes em A.1.12);
- Outra possibilidade é que a interface mostre uma opção para o usuário indicar uma nova categoria (mais detalhes em A.1.11).

Justificativa. O desenvolvedor precisa que o usuário veja o texto ou documento e dê o *feedback* de rotulagem. O texto categorizado servirá para construir um banco de dados e assim seguir as demais etapas do processo de criação do modelo de ML.

Exemplos



Text Classification

Allright this is my **first** try with this..... so hold on uh lets see today ... well it was very pointless.... i took a health **test** and a chem. **quiz**.... but besides that i didn't really do anything productive... thats the problem with **school**... the work doesn't start until you get home..... well whatever..... oh and i had to go to **confession** b/c my mom made me (you have to go once a year) funny thing was ... the priest told me i really hadn't sinned he was like 'fighting with your parents is normal'... i was like SWEET!..... oh well im done **writing** for today..... hopefully this blog will get interesting as time goes on...

This text is related to a Student Blog?

Figura A.1 – Exemplo de *Padrão Text or Document Labeling for a Category*. A interface mostra uma instância de texto com as palavras mais essenciais ou mais frequentes destacadas. Há uma lista de botões para solicitar *feedback* sobre a categoria de texto. Além disso, há um botão "not sure" para indicar dúvidas sobre a rotulagem de dados.

A.1.2 Padrão Text or Document Labeling for Multi-category

Problema. O usuário precisa visualizar e rotular um documento ou texto em uma multi-categoria.

Uso. Use quando uma tarefa de rotulagem do usuário for visualizar e categorizar um documento ou texto em várias categorias. O usuário pode escolher mais de uma categoria para rotular um único texto.

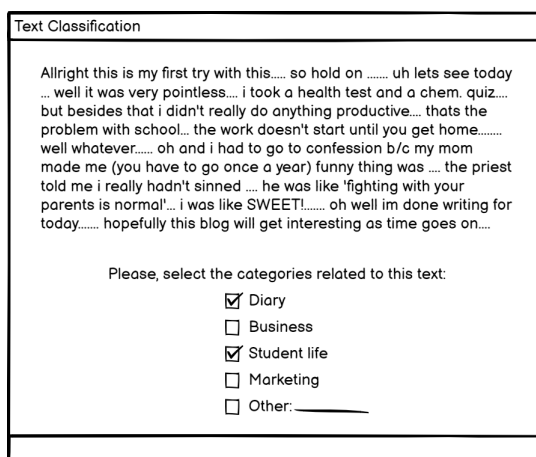
Solução.

- A interface deve mostrar o texto ou documento a ser rotulado.

- A visualização de texto ou documento pode mostrar as palavras mais importantes ou mais frequentes (Figura A.2).
- O usuário pode escolher mais de uma categoria para rotular um único texto (instância), e a interface deve exibir as categorias de uma destas formas possíveis:
 - *Buttons*;
 - *Checkbox* (Figura A.2);
 - *Dropdown* menu com várias seleções.
- Opcionalmente, a interface pode mostrar opções para indicar dúvidas sobre rotulagem de dados (mais detalhes em A.1.12);
- Outra possibilidade é que a interface mostre uma opção para o usuário indicar uma nova categoria (item “others” na Figura A.2 - mais detalhes em A.1.11).

Justificativa. O desenvolvedor precisa que o usuário veja o texto ou documento e dê o *feedback* de rotulagem. O texto categorizado servirá para construir um banco de dados e assim seguir as demais etapas do processo de criação do modelo de ML.

Exemplo.



The screenshot shows a window titled "Text Classification". Inside, there is a text area containing a paragraph of text. Below the text, there is a prompt: "Please, select the categories related to this text:". Underneath the prompt, there is a list of categories with checkboxes: "Diary" (checked), "Business" (unchecked), "Student life" (checked), "Marketing" (unchecked), and "Other:" followed by a blank line for text entry.

Figura A.2 – Exemplo de padrão *Text or Document Labeling for Multi-category*. A interface mostra a instância de texto e solicita que o usuário selecione as categorias relacionadas a ela. A interface exibe as classes em um componente *checkbox* onde o usuário pode escolher mais de um rótulo. Além disso, o usuário pode selecionar a opção "other" e inserir uma nova categoria.

A.1.3 Padrão Entity Labeling Task

Problema. O usuário precisa visualizar o texto e identificar entidades neste texto.

Uso. Use quando o problema de aprendizado de máquina for o seguinte: *Named Entity Recognition (NER)*, *Part-of-Speech (POS)*, *Keyphrase Tagging*. Todos esses problemas exigem que o usuário anote uma palavra em um texto para uma categoria.

Solução.

- Permite que o usuário visualize e rotule entidades em um texto para problemas de PNL. A interface deve mostrar as entidades disponíveis de uma dessas maneiras:
 - *Buttons*: o usuário clica em um botão de categoria e, em seguida, seleciona a palavra relacionada (Figura A.3);
 - *Dropdown menu*: o usuário seleciona uma palavra e, em seguida, clica na categoria correspondente disponível no *dropdown menu*;
- A interface pode mostrar que uma palavra anotada está relacionada a uma categoria por uma cor acessível no plano de fundo.
- A cor da palavra anotada deve corresponder à cor da categoria correspondente.
- A interface deve permitir que o usuário exclua a anotação (mais detalhes em A.1.9).
- Opcionalmente, a interface pode mostrar opções para indicar dúvidas sobre rotulagem de dados (mais detalhes em A.1.12);
- Outra possibilidade é que a interface exiba uma opção para o usuário indicar uma nova categoria (mais detalhes em A.1.11).

Justificativa. O desenvolvedor precisa que o usuário rotule partes ou entidades de um texto. As palavras rotuladas servirão para construir um banco de dados e assim seguir as demais etapas do processo de criação do modelo de ML. **Exemplo.**

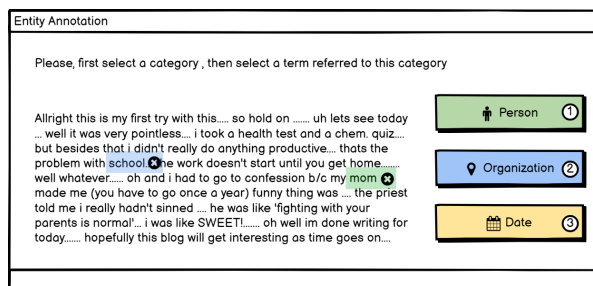


Figura A.3 – Exemplo de padrão *Entities Labeling*. A interface mostra a instância de texto e os rótulos por meio de botões. O usuário clica em um botão que representa um rótulo e seleciona a palavra relacionada. A palavra anotada tem uma cor de fundo que corresponde à mesma cor do rótulo correspondente.

A.1.4 Padrão Relation Extraction Task

Problema. O usuário precisa relacionar dois termos rotulados.

Uso. Use quando for necessário resolver um problema de extração de relações.

Solução.

- Permite que o usuário clique em dois termos rotulados do texto para criar uma relação entre eles. A interface deve mostrar uma lista de relações possíveis e a possibilidade de adicionar uma nova categoria de relação (veja a opção *other* na Figura A.4).
- A interface deve indicar a relação por meio de uma seta com o rótulo selecionado.
- A cor da seta pode estar relacionada à cor da categoria de relação correspondente.

Justificativa. O desenvolvedor precisa que o usuário relacione duas entidades rotuladas de um texto. Essas relações servirão para construir um banco de dados e assim seguir as demais etapas do processo de criação do modelo de ML.

Dependência Use este padrão junto com o padrão *Entity Labeling Task* A.1.3.

Exemplo.

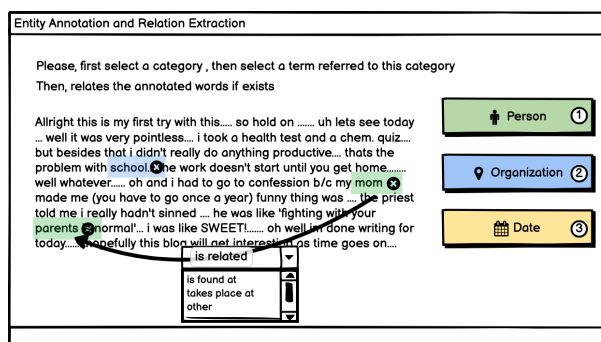


Figura A.4 – Exemplo do padrão *Relation Extraction Task Labeling*. A interface mostra uma instância de texto com algumas palavras rotuladas. O usuário deve associar duas palavras rotuladas clicando em ambas as palavras e selecionando a relação correta. A interface apresenta a nova relação por meio de uma seta com o rótulo.

A.1.5 Padrão Batch Labeling

Problema. O usuário pode precisar rotular ou validar mais de uma instância por vez.

Uso. Use somente quando houver um pré-processamento dos dados e uma categorização inicial, onde o usuário pode atuar como inspetor e validar os dados por lote.

Solução. A interface mostra os dados em diferentes visualizações, dependendo do cenário de rotulagem. Os dados podem passar por pré-processamento. Permite que o usuário rotule mais de um dado por vez. A interface deve incluir:

- Mostrando um grupo de instâncias e perguntando se estão corretas;
- Perguntar ao usuário quais instâncias estão corretas e pode indicar isso (Figura A.5);

Opcionalmente, use filtros para selecionar as instâncias que possuem as palavras mais frequentes (veja a tabela à direita na Figura A.5).

Justificativa. Maximize a análise de rotulagem do usuário. O usuário pode validar mais de um dado quando apresentado a um conjunto previamente processado pela ferramenta de rotulagem.

Exemplo.

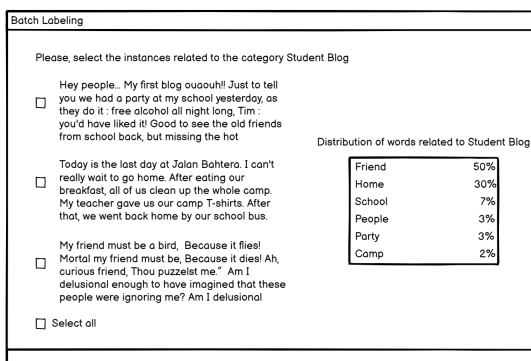


Figura A.5 – Exemplo do padrão *Batch Labeling*. A interface mostra um grupo de instâncias rotuladas e solicita que o usuário selecione as corretas. No lado direito, uma tabela apresenta as palavras mais frequentes dos casos.

A.1.6 Padrão Tool Help

Problema. O usuário tem dúvidas sobre as funcionalidades da ferramenta.

Uso. Use quando o usuário precisar de instruções sobre a tarefa de rotulagem ou dúvidas sobre a funcionalidade da ferramenta.

Solução. Deixe as instruções acessíveis e visíveis na interface da ferramenta. A interface deve mostrar uma destas opções:

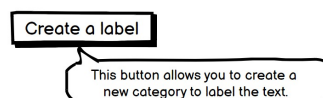
- Ícone de ajuda para abrir instruções em um modal ou pop-up (Figura A.6a);
- Dicas de ferramentas em todos os botões principais (Figura A.6b);
- Instruções de rotulagem por um texto ou vídeos;

Justificativa. Ajudar os usuários a usar a ferramenta inteiramente sem dúvidas sobre a funcionalidade e ações.

Exemplo



(a) Ícone de ajuda



(b) Dica de ferramenta no botão.

Figura A.6 – Exemplos de padrão de ajuda.

A.1.7 Padrão External Help

Problema. O usuário tem dúvidas sobre o domínio de dados a ser rotulado.

Uso. Use quando o usuário precisar entender melhor sobre os dados.

Solução. Forneça um link para a Web para pesquisar informações sobre os dados.

Justificativa. Ajude os usuários a entender mais sobre os dados por meio de ajuda externa.

Exemplo.



Figura A.7 – Exemplo de Ajuda Externa. A interface fornece um link para a web para pesquisar informações sobre os dados

A.1.8 Padrão Word Search

Problema. O usuário precisa encontrar uma palavra específica em um texto.

Uso. Use quando o usuário quiser encontrar uma palavra específica no texto.

Solução. Permite uma barra de busca de palavras na interface. Opcionalmente, a barra de pesquisa pode fazer:

- *Autocomplete*;
- Destaque no texto as palavras encontradas;
- Dê *feedback* de quantas palavras a pesquisa encontrou;
- Dê *feedback* se a pesquisa não retornou nenhuma palavra encontrada;

Justificativa. Ajudar os usuários a encontrar uma palavra específica no texto. Normalmente, um usuário tem uma ideia de uma palavra específica que pode ajudá-lo a rotular o texto.

Exemplo.

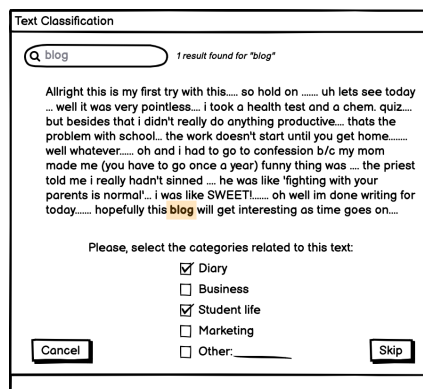


Figura A.8 – Exemplo do padrão *Word Search*. A interface fornece uma *search bar*, destacando as palavras encontradas e dando o *feedback* de quantas palavras a pesquisa encontrou.

A.1.9 Padrão Undo a labeling task

Problema. O usuário deseja desfazer uma tarefa ou limpar anotações.

Uso. Use quando o usuário quiser:

- Desfazer uma tarefa de rotulagem;
- Limpe uma anotação;

Solução. A interface deve permitir que o usuário desfaça uma anotação. Para isso, a interface deve mostrar:

- Excluir opção de uma anotação. Para rotulagem de entidades, a interface deve mostrar um ícone de exclusão ao lado da palavra anotada (Figura A.9);
- Botão Limpar tudo;
- Botão Desfazer (Figura A.9).

Justificativa. O usuário deve poder desfazer algumas anotações em caso de erro ou mal-entendido.

Dependência. Use este padrão com o padrão *Entity Labeling Task* A.1.3.

Exemplo.

Allright this is my first try with this..... so hold on uh lets see today ... well it was very pointless.... i took a health test and a chem. quiz... but besides that i didn't really do anything productive... thats the problem with school... the work doesn't start until you get home..... well whatever..... oh and i had to go to confession b/c my mom made me (you have to go once a year) funny thing was ... the priest told me i really hadn't sinned ... he was like 'fighting with your parents is normal'... i was like SWEET!..... oh well im done writing for today..... hopefully this blog will get interesting as time goes on...

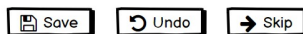


Figura A.9 – Exemplo do padrão *Undo*. A interface fornece um botão de desfazer e mostra um ícone de exclusão ao lado das palavras anotadas.

A.1.10 Padrão Navigation Between Instances

Problema. O usuário deve ter um caminho para ir para a instância anterior ou posterior.

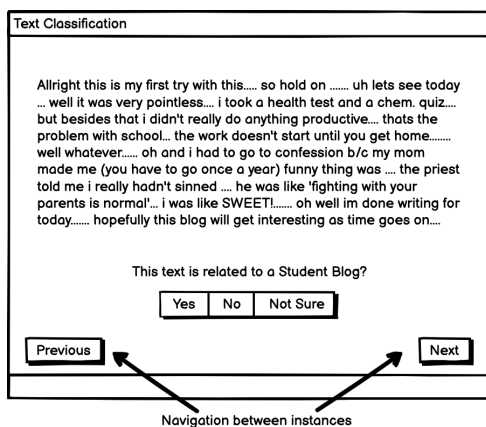
Uso. Use quando o usuário quiser:

- Volte para a rotulagem feita anteriormente;
- Vá para o próximo dado a ser rotulado;

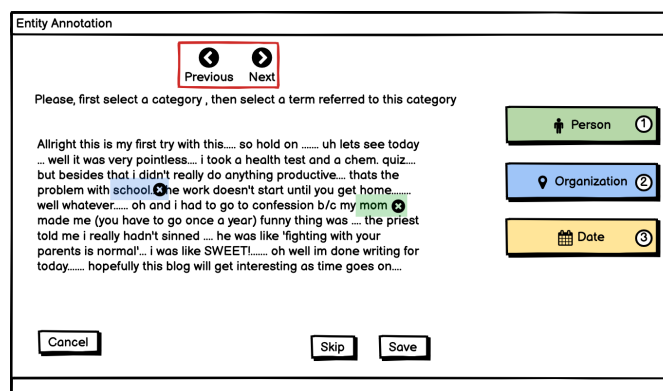
Solução. A interface deve apresentar botões para permitir que o usuário vá para a instância anterior ou seguinte, indicada por botões ou ícones (Figuras A.10a e A.10b).

Justificativa. O usuário precisa se sentir à vontade para escolher qual instância rotular.

Exemplo.



(a) Navegação através de botões.



(b) Navegação através ícones.

Figura A.10 – Exemplos do padrão *Navigation Between Instances*.

A.1.11 Padrão Add Category

Problema. O usuário deseja adicionar outra categoria.

Uso. Quando o usuário deseja indicar outra categoria que possa estar relacionada aos dados.

Solução. A interface deve mostrar opções aos usuários para adicionar outras opções. A interface pode exibir uma caixa de entrada para *other* (Figura A.11a) ou um botão para adicionar uma categoria (Figura A.11b).

Justificativa. O usuário pode ser um especialista no domínio de dados e contribuir para a tarefa de rotulagem adicionando novas categorias.

Dependência. Use este padrão junto com um padrão:

- *Text or Document Labeling for a Category* A.1.1;
- *Text or Document Labeling for Multi-category* A.1.2;
- *Entity Labeling Task* A.1.3.

Exemplo.

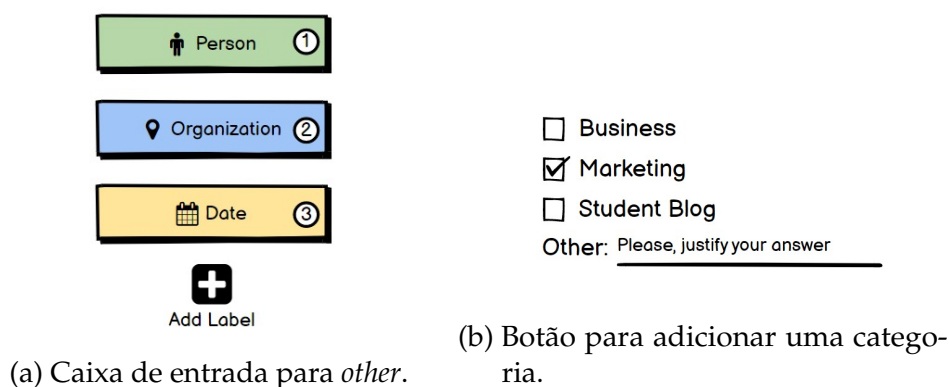


Figura A.11 – Exemplos do padrão *Add Category*.

A.1.12 Padrão Feedback of Lack of Data Domain

Problema. O usuário deseja indicar não saber sobre o domínio de dados.

Uso. Use quando o usuário não conhece ou não domina a rotulagem de determinados dados.

Solução. A interface deve mostrar opções aos usuários para indicar que eles não têm certeza sobre a categoria correta. A interface pode oferecer um botão *not sure* para permitir que o usuário exiba não saber sobre os dados (Figura A.12).

Justificativa. Alguns usuários podem não ter conhecimento sobre o domínio de dados. Para esses usuários, a ferramenta deve evitar a rotulagem errada.

Dependência. Use este padrão junto com um padrão:

- *Text or Document Labeling for a Category* A.1.1;
- *Padrão Text or Document Labeling for Multi-category* A.1.2;
- *Entity Labeling Task* A.1.3.

Exemplo.

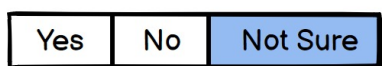


Figura A.12 – Exemplo do padrão *Feedback of Lack of Data Domain*. A interface fornece um botão *not sure*.

A.1.13 Padrão Save

Problema. O usuário deseja salvar o progresso.

Uso. Quando o usuário precisa salvar a tarefa de rotulagem.

Solução. A interface deve permitir que o usuário salve e veja seu feedback. A interface deve mostrar um botão salvar (Figura A.13a). A ferramenta pode ter um recurso de salvamento automático (Figura A.13b), mas certifique-se de que o usuário saiba disso (mais detalhes em A.1.16).

Justificativa. O usuário deve ter certeza de que todo o trabalho será salvo e não perderá nenhum trabalho realizado.

Exemplos.

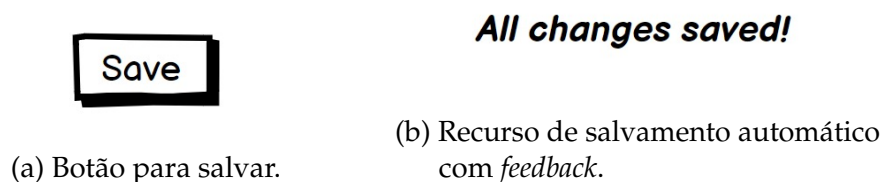


Figura A.13 – Exemplos do padrão *Save*.

A.1.14 Padrão Cancel

Problema. O usuário deseja cancelar a tarefa de rotulagem e sair da ferramenta.

Uso. Use quando o usuário precisar cancelar a tarefa de rotulagem e sair da ferramenta.

Solução. A interface deve permitir que o usuário cancele através de um botão, mas certifique-se de que o sistema salvou todo o trabalho antes (consulte [A.1.16](#) e [A.1.13](#)).

Justificativa. O usuário deve ter certeza de deixar a ferramenta sem perder nenhum trabalho feito.

Exemplo.

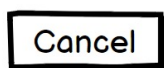


Figura A.14 – Exemplo do padrão *Cancel*.

A.1.15 Padrão Shortcuts

Problema. O usuário precisa de atalhos para acessar as principais funções em menos tempo.

Uso. Use quando o usuário quiser fazer as principais ações da ferramenta com facilidade.

Solução. Permitir atalhos para as principais ações do sistema, como:

- Barra de atalhos com as principais ações da ferramenta (Figura [A.15a](#));

- Atalhos de teclado visíveis ou dicas de ferramentas nos botões principais (Figura A.15b);

Justificativa. Ajude os usuários a encontrar atalhos para funcionalidades essenciais da ferramenta. Além disso, o usuário pode realizar a rotulagem mais rapidamente em menos tempo.

Exemplos.



- (a) Barra de atalhos com as principais ações da ferramenta. (b) Atalhos de teclado visíveis para botões de rótulo.

Figura A.15 – Exemplos do padrão *Shortcuts*.

A.1.16 Padrão Tool status

Problema. O usuário precisa saber sobre o status da tarefa de rotulagem em processamento na ferramenta.

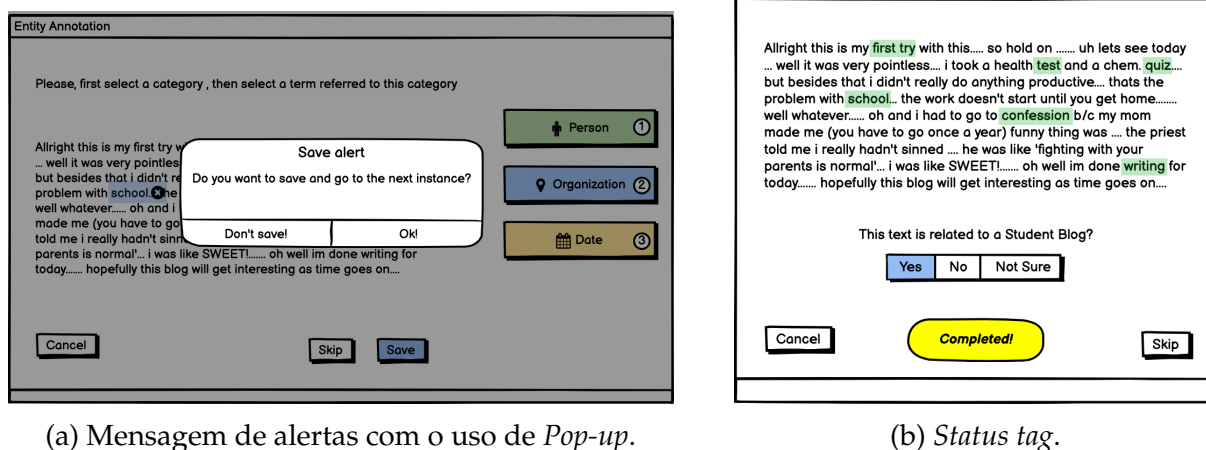
Uso. Use esse padrão quando o usuário precisar aguardar o processamento de uma ação pela ferramenta.

Solução. Comunique ao usuário se as ações estiverem em andamento ou registradas. Após cada ação do usuário, a interface pode mostrar:

- Mensagem de alertas com o uso de *Pop-up* (Figura A.16a);
- Mensagem de texto na interface;
- *Status tag* (Figura A.16b);

Justificativa. É útil para o usuário ter feedback sobre o que está acontecendo na ferramenta. Esse feedback torna a ferramenta mais confiável para o usuário.

Exemplos.

(a) Mensagem de alertas com o uso de *Pop-up*.(b) *Status tag*.Figura A.16 – Exemplos do padrão *Tool Status*.

A.1.17 Padrão Labeling Task Progress

Problema. O usuário precisa saber o status da tarefa de rotulagem.

Uso. Use este padrão quando o usuário quiser:

- Identifique quantos dados ainda precisam ser rotulados.
- Identifique quantos dados já foram rotulados.
- Saiba quanto tempo levou na tarefa de rotulagem.

Solução. Apresente mensagens ou imagens que informem o usuário sobre a atividade de rotulagem. Os componentes da interface devem relatar uma ou mais destas opções:

- A quantidade de marcação realizada (Figuras A.17a, A.17c e A.17d);
- O tempo gasto (figura A.17b);
- A quantidade de dados que precisa ser rotulada (Figura A.17d);

Justificativa. É essencial informar aos usuários quanto esforço eles fizeram durante a rotulagem. Adicionando elementos de interface que notificam o status de andamento da tarefa, o usuário recebe esse feedback.

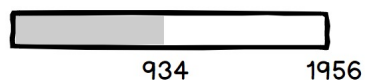
Exemplos.

Great! You already labeled 13 emails! Keep going!

Current Time: 0:00:56

(a) Mensagem sobre a quantidade de rotulagem realizada.

(b) Message about the time spent.



Total instances	Instances labeled	Instances left
100	47 (47%)	53 (53%)

(c) Barra de progresso.

(d) Tabela com informações sobre a atividade de rotulagem.

Figura A.17 – Exemplos do padrão *Labeling Task Progress Status*.

A.1.18 Padrão Labeling History Overview

Problema. O usuário precisa ver um histórico das atividades realizadas na ferramenta.

Uso. Use esse padrão quando o usuário quiser saber quais instâncias foram rotuladas.

Solução. A interface deve mostrar quais atividades o usuário realizou na ferramenta. É opcional mostrar:

- A atividade atual do usuário é rotular;
- Nome do usuário;
- Carimbo de hora da atividade;

Justificativa. Verificar o histórico de ações do que o usuário fez.

Exemplo.

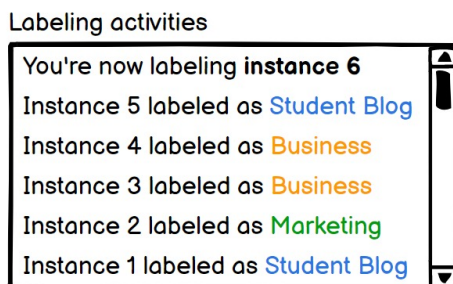


Figura A.18 – Exemplo do padrão *Labeling History Overview*. O componente de interface fornece o histórico de atividades do usuário que está rotulando.

A.1.19 Padrão Statistics

Problema. O usuário precisa de uma visão geral da tarefa de rotulagem de classificação para entender os dados e o modelo de ML em construção.

Uso.

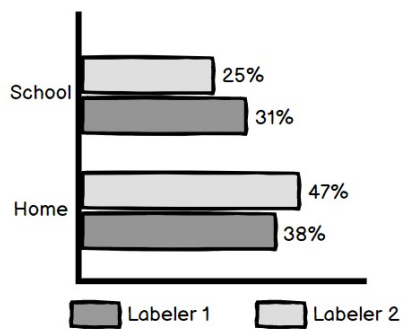
- Use quando um usuário curador, em um processo de validação, precisa conhecer as estatísticas gerais de palavras, categorias e classificações;
- Use se a ferramenta tiver um modelo de ML online em construção, a interface pode mostrar os resultados atuais do modelo.

Solução. TA interface deve conter componentes que ajudem o usuário a entender melhor os dados rotulados. Dados estatísticos podem ajudar nisso. Além disso, as técnicas de Visual Analytics podem ajudar o usuário a entender o contexto dos dados. A interface pode mostrar:

- Distribuição de instâncias rotuladas por usuários em um gráfico de barras (Figura [A.19a](#));
- Frequência de palavras existentes em uma categoria (Figura [A.19b](#));
- Matriz de confusão com os resultados do modelo ML.

Justificativa. Ele ajuda o usuário a entender os dados e o modelo de ML por meio de uma visão geral da tarefa de rotulagem de classificação.

Exemplos.



(a) Gráfico de barras com distribuição de instâncias rotuladas por usuários.

Word	Frequency	Progress Bar
School	50/100	<div style="width: 50%;"></div>
Work	48/100	<div style="width: 48%;"></div>
Quiz	40/100	<div style="width: 40%;"></div>
Exam	35/100	<div style="width: 35%;"></div>
Test	20/100	<div style="width: 20%;"></div>

(b) Componente de tabela com frequência de palavras existentes em uma categoria.

Figura A.19 – Exemplos do padrão *Statistics* .