



VISUALIZAÇÃO DE DADOS DA MEDIATECA FAU-UFRJ

ALUNA: CÍNTIA MECHLER DE CARVALHO
ORIENTADORES: RODRIGO CURY PARAIZO E GONÇALO CASTRO HENRIQUES

ESTUDO FINAL | TFG2 | 2022



AGRADECIMENTOS

O desenvolvimento deste trabalho e também de toda a minha formação como aluna, arquiteta e pesquisadora se devem - primeiramente - aos meus orientadores Rodrigo Cury e Gonçalo Henrique. Me ensinaram a importância do compartilhamento do conhecimento e sempre me apoiaram para além dos muros da universidade.

Quero agradecer aos laboratórios LAMO 3d e LAURD por ter me apresentado ao mundo da tecnologia e que é possível integrar disciplinas que, à primeira vista, não estão relacionadas. Também agradeço aos professores de ambos os laboratórios: Naylor Vilas Boas, Thiago Leitão, Maria Cristina Cabral, Andrés Passaro e também a todas as equipes de pesquisa que tive a oportunidade de dividir as reuniões no PROURB.

Aos meus amigos Vanessa Alves, André Tardelli, João Vitor, Pedro Moya, Wesley Picone, Kaori Laburu e Lucas Diniz por todo o amor, carinho e paciência que tiveram durante essa trajetória mais do que longa.

E um agradecimento especial à Andressa Souza, que infelizmente não está mais conosco para que pudéssemos dividir esse momento. Uma pessoa extremamente importante na minha vida, que me deu todas as oportunidades no início da graduação e me fez não desistir. Obrigada por tudo.

SUMÁRIO

1. INTRODUÇÃO	4
1.1 TFG DE ARQUITETURA	4
1.2 REPOSITÓRIO, ACERVO E OBJETO DIGITAL	7
1.3 VISUALIZAÇÃO DE UM ACERVO CULTURAL	8
2. OBJETIVOS	13
3. ESTUDOS DE CASO DE DIFUSÃO DE ACERVOS DE TFGs ONLINE	13
3.1 CONTEXTO HISTÓRICO NA FAU-UFRJ	13
3.2 PORTAL MEDIATECA	15
3.3 REFERÊNCIAS DE ACERVOS DE TFGs	16
4. OS DADOS DO PORTAL MEDIATECA	19
5. METODOLOGIA	21
5.1 REPERTÓRIO TÉCNICO	23
6. RESULTADOS	27
6.1 VIKUS VIEWER MEDIATECA	27
6.2 VISUALIZAÇÕES COM O IPVY	31
7. PERSPECTIVAS	35
8. REFERÊNCIAS BIBLIOGRÁFICAS	36
9. LISTA DE IMAGENS	38

1. INTRODUÇÃO

Este trabalho foi desenvolvido para o Trabalho Final de Graduação 2 do curso de Arquitetura e Urbanismo da Universidade Federal do Rio de Janeiro (FAU-UFRJ). Trata-se de uma experimentação de visualização de dados do acervo digital de Trabalho Final de Graduação (TFG) da FAU-UFRJ - o Portal MEDIATECA, com a finalidade de criar uma síntese para compreender a produção acadêmica a partir de uma análise do conjunto de dados sobre os trabalhos, parte enviados pelos autores dos trabalhos (imagem, título, resumo, localização, palavras-chave), parte da Comissão Coordenadora de Trabalho Final de Graduação (banca de avaliação, orientador, ano de publicação) e da catalogação da MEDIATECA (categorias, temas), fomentando a discussão sobre visualização da informação por arquitetos e urbanistas.

A formação do arquiteto é extremamente diversificada pela necessidade de compreender os diferentes fatores que compõem o processo de concepção de um projeto arquitetônico e/ou urbanístico. A MEDIATECA tem importância neste processo, uma vez que seu acervo é composto pelos trabalhos de conclusão e a sua difusão permite que os alunos, assim como pesquisadores, arquitetos formados e outros profissionais de áreas afins, possam consultar sua base para desenvolver trabalhos e pesquisas¹. Visto a relevância desta plataforma para diversas áreas, a visualização desse conjunto

¹ FREIRE, W.; PARAIZO, R. C. ; KAMIMURA, R. . **Curadoria para a exposição digital-interativa da MEDIATECA da FAU/UFRJ**. In: XXXVII ENSEA / XX CONABEA - Desafios do Ensino de Arquitetura e Urbanismo no Século XXI, 2019, Rio de Janeiro. Caderno 42 ABEA. Rio de Janeiro: ABEA, 2019. p. 504-514;

de dados permite investigar, para além da concepção de um projeto, a forma que a produção do conhecimento acontece na instituição ao longo do tempo. O tamanho do seu acervo, que está em constante expansão conforme os alunos se graduam, torna desafiador a análise manual das informações enviados de cada trabalho, o que requer a combinação da análise computacional com a análise humana, uma vez que só é possível trabalhar com grandes conjuntos de dados com o poder de processamento de um computador².

1.1 TFG DE ARQUITETURA

A regulamentação do Trabalho Final de Graduação dos cursos vigentes no país é feita pelo Ministério da Educação (MEC). Através do texto da resolução específica para Arquitetura e Urbanismo³, as instituições têm liberdade para definir como serão condicionados esses trabalhos em função das determinações. Há um certo consenso em determinar que os TFGs sejam da categoria **projeto**, uma vez que isso representaria a síntese da graduação, envolvendo e aplicando a maioria das disciplinas obrigatórias.

Atualmente na FAU-UFRJ existe uma comissão organizadora que elabora todo semestre as resoluções para o desenvolvimento do TFG, que é dividido em duas partes - no nono período e no décimo período. Esta comissão, intitulada Comissão Coordenadora do Trabalho Final de Graduação (CCTFG), é composta por um grupo

² MANOVICH, L. **What is Visualization?** Manovich, 2010. Disponível em: <<http://manovich.net/index.php/projects/what-is-visualization>>. Acesso em: 11 ago. 2021.

³ **RESOLUÇÃO Nº 2, DE 17 DE JUNHO DE 2010**. Ministério da Educação, 2010. Disponível em: <http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=5651-rces002-10&category_slug=junho-2010-pdf&Itemid=30192>. Acesso em: 12 jan. 2022.

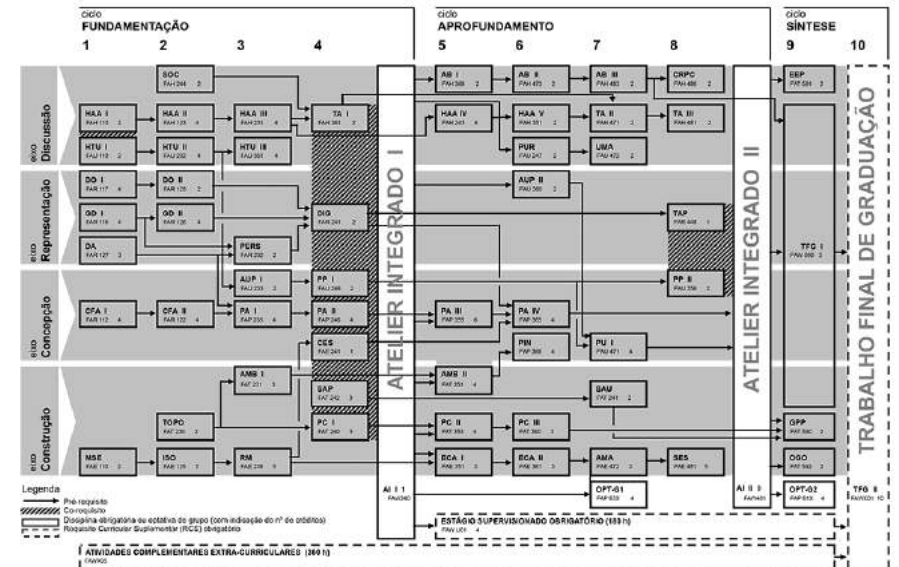
de docentes, diretor adjunto de graduação, coordenador de operações acadêmicas, grupo de monitores-bolsistas e representantes de turmas. Todo período é elaborado um documento e encaminhado aos alunos com as resoluções e orientações do TFG, pois estas informações podem sofrer alterações de acordo com novos membros da CCTFG, especificidades acadêmicas da UFRJ ou fatores externos, como novas resoluções vindos do MEC.

Utilizando como referência o regulamento que rege este trabalho⁴, entende-se que "o objetivo geral do trabalho é representar a prontidão do(a) estudante para o exercício profissional da Arquitetura e do Urbanismo, de modo alinhado com o Projeto Pedagógico do Curso". (CCTFG, 2020) Deste modo, a partir do Projeto Pedagógico criado em 2006, os TFGs são realizados de maneira individual, com tema e formatação livre, porém que estejam de acordo com o projeto pedagógico mencionado anteriormente.

Esta liberdade do tema e formato permite que os alunos explorem diversas representações que estejam mais alinhadas com o trabalho. Ao longo da graduação, somos ensinados a desenvolver um discurso em paralelo com o projeto. Este discurso pode ser construído primeiro, a partir de conceitos e teorias, ou a partir de experimentações formais de projeto, que justifiquem as tomadas de decisão projetual. Há dois momentos na grade curricular da FAU (ver imagem 01) em que há uma experimentação dessas metodologias, e de outras, para o desenvolvimento de projetos.

O conceito de Atelier Integrado vem da capacidade de integrar diversas disciplinas, resultando em apenas um trabalho. Com isso,

para determinar os ciclos pedagógicos, o atelier serve como uma síntese das disciplinas anteriores a ele, avaliando o aluno - ou grupo de alunos -, semelhante a uma banca de TFG.



(Imagem 01: Grade curricular do projeto pedagógico 2066 da FAU-UFRJ. Fonte: FAU-UFRJ)

Os produtos gerados no atelier costumam seguir um padrão dos concursos de arquitetura: uma prancha resumo/síntese, pranchas técnicas (plantas, cortes, fachadas, etc) e/ou imagens síntese (renders, fotografias, colagens, etc). A entrega de todos esses produtos - ou parte deles - depende da curadoria dos concursos, assim como do grupo de professores que cada turma do atelier possui. O mais comum é a prancha resumo ou síntese, pois nela está contida todos os outros produtos, em menor quantidade, que seja essencial para compreender o projeto todo. Esta prancha precisa ser capaz de, não só sintetizar todo o discurso, mas também se sustentar sem a presença de seu autor.

⁴ Regulamento do trabalho final de conclusão de curso de graduação em Arquitetura e Urbanismo da Universidade Federal do Rio de Janeiro - FAUFRJ. Comissão Coordenadora do Trabalho Final de Graduação; Faculdade de Arquitetura e Urbanismo - Rio de Janeiro, RJ: FAU-UFRJ, 2020. Disponível em: <http://www.fau.ufrj.br/wp-content/uploads/2020/04/TFG_Regulamento_Aprovado_11mar2020.pdf> Acesso em: 24 nov 2021;



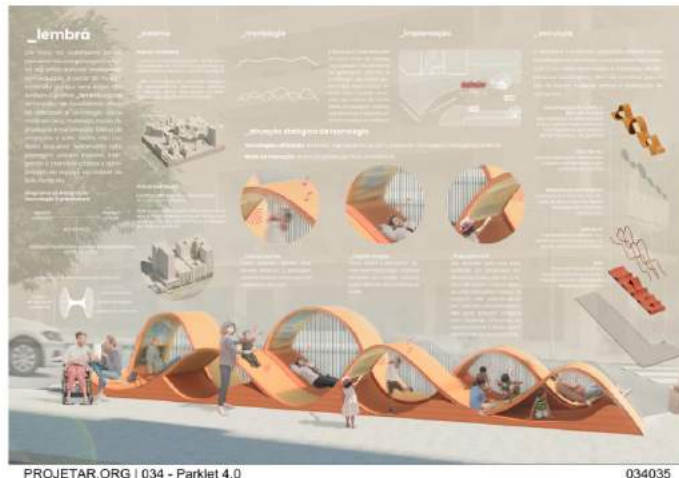
ENTRAR ARQUITETURA DESIGN TCC

PROJETAR.ORG

PORTUGUÊS

SOBRE PUBLICAÇÕES PARCEIROS JURI

1º COLOCADO



Paola Bonetto Ferrari
Universidade Federal de Minas Gerais - UFMG - Belo Horizonte/MG

(Imagem 02: Prancha resumo do projeto vencedor do concurso Parklet 4.0, organizado pelo antigo portal Projetar.org. Fonte: <<https://projetar.org/vencedores/59/parklet-4.0-034>>. Acesso em: 10 fev 2022)

Ainda que seja comum para se trabalhar os projetos de arquitetura, a linguagem adotada nas pranchas depende muito do arquiteto que a desenvolveu ou do projeto. Muitos autores utilizam a estética da prancha como parte do discurso ou de elementos marcantes do projeto na forma de grafismos, compondo imagens, norteando a cor predominante e assim por diante. A imagem passa a ser elemento fundamental para representar o projeto.

Durante muito tempo, a FAU apenas aceitou TFGs que fossem da categoria **projetal**, por entender que estes trabalhos estariam alinhados com os projetos pedagógicos e por fazer parte da tradição

da instituição. Ainda que de maneira livre, muitos trabalhos foram elaborados na ideia de pranchas síntese e pranchas técnicas, pois para avaliar um projeto arquitetônico, se faz necessário a análise técnica de acordo com o tema do trabalho: estrutura, saneamento, habitação, afins. Ao criar duas novas categorias, **teórico** e **experimental**, a FAU rompe com a tradição por trabalhos de projetos arquitetônicos/urbanos e abre espaço para trabalhos que discutem a teoria ou então pesquisem algum tema relacionado, por exemplo.

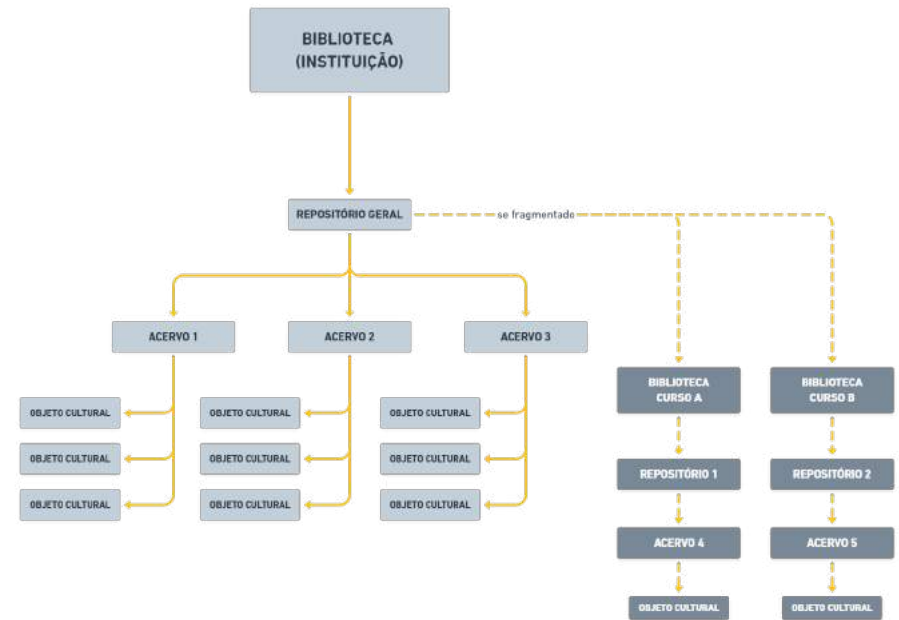
Esta nova demanda de trabalhos é completamente diferente da estrutura anterior, já que trabalhos teóricos precisam estar de acordo com a NBR14724, que fala sobre a formatação de trabalhos científicos, monografias e TCCs (trabalho de conclusão de curso), e os trabalhos experimentais vai depender da metodologia e resultados obtidos, assim como os trabalhos de uma Iniciação Científica.

A CCTFG a fim de estabelecer uma uniformização mínima dos trabalhos, fala em sua resolução na exigência dos produtos: **imagem-síntese** (formato JPG, PNG, GIF ou PDF), **material gráfico/textual** (formato PDF) e um **resumo do trabalho** de até 400 caracteres (com espaços). A apresentação, principalmente no modelo remoto, fica com a parte de “formatação livre”, podendo explorar diversas técnicas e ferramentas para transmitir o conceito do trabalho. Isso acontece pois o TFG de arquitetura por si só já é complexo para a documentação e armazenamento, uma vez que não há um formato padrão, mesmo que haja uma tendência de formatos. Ao introduzir outras duas categorias distintas, aumenta-se as dificuldades para o armazenamento destes trabalhos.

1.2 REPOSITÓRIO, ACERVO E OBJETO DIGITAL

Segundo o Instituto Brasileiro de Museus (IBRAM)⁵, pode-se entender o acervo digital a partir de duas categorias: o acervo digitalizado (aquele que mantém obras físicas como quadros, estátuas, fotografias e afins, em um espaço físico real e passam por um processo de digitalização) e o acervo nato digital (não possui obras físicas, apenas virtuais). O entendimento a partir dessas duas categorias é fundamental para compreender o processo histórico dos acervos digitais de trabalhos acadêmicos de arquitetura nos últimos anos.

Dentro das universidades, as bibliotecas são responsáveis por catalogar, armazenar e controlar o acesso das informações e produções da instituição através dos repositórios, que podem ser únicos e/ou fragmentados entre as diferentes instâncias dentro da universidade (ver imagem 01). No caso particular do objeto TFG de arquitetura e urbanismo, está inserido como um objeto do acervo de uma subdivisão do repositório geral da instituição universitária. Em outras palavras, trata-se de um repositório-filho do repositório geral, em que sua gestão e acesso fica a encargo da biblioteca do curso de Arquitetura e Urbanismo.



(Imagem 03: Diagrama de estrutura dos repositórios de uma instituição universitária. Fonte: Autora)

Como explicado anteriormente, a representação faz parte do projeto arquitetônico, comunicando uma ideia do arquiteto para e/ou sobre um espaço. As diferentes formas de representação desse espaço, como maquetes físicas, diagramas, pranchas e imagens sínteses e etc, geram desafios para a documentação desses trabalhos, pois o estabelecimento de uma formatação e padronização de temas poderia limitar a produção dos trabalhos e condicionar a um pensamento, enquanto a ausência do mesmo cria diversos tipos de catalogação da informação, tornando heterogêneo os dados sobre os trabalhos. A evolução das tecnologias e de ferramentas que auxiliam no desenvolvimento dessas representações também justifica os diferentes métodos de representação desses trabalhos. Um exemplo deste fato é a representação gráfica de plantas técnicas: até a difusão

⁵ **Acervos digitais nos museus: manual para realização de projetos.** Instituto Brasileiro de Museus; Universidade Federal de Goiás - Brasília, DF: Ibram, 2020. Disponível em: <<https://www.museus.gov.br/wp-content/uploads/2021/05/Acervos-Digitais-nos-Museus.pdf>> Acesso em: 28 set de 2021

de softwares de desenho CAD (Computer Assisted Design), esses desenhos eram feitos à mão com nanquim e papel manteiga, enquanto que nos dias de hoje já há outro processo de criação e pensamento desses desenhos através da metodologia BIM (Building Information Modeling).

A partir disso, ao falarmos de acervos digitais voltados para TFGs de arquitetura, estamos trabalhando sobre um possível acervo digital híbrido, entre o digitalizado e o nato digital. Dependendo de quando a catalogação dos trabalhos começa, é possível ter uma coleção exclusivamente digital - sem um original físico - o que não descarta a possibilidade de haver reproduções físicas de um conteúdo criado digitalmente. Para acompanhar as evoluções nas representações dos trabalhos, é preciso que a visualização desses repositórios também se modifique ao longo do tempo, gerando um trabalho de pesquisa contínuo aos gestores destes repositórios de repensar sobre a formatação da visualização e, conseqüentemente, na formatação dos dados do trabalho. O armazenamento desses trabalhos fica sob responsabilidade da MEDIATECA, que promove a digitalização, difusão e gestão deste acervo.

1.3 VISUALIZAÇÃO DE UM ACERVO CULTURAL

Contextualizado o que é um acervo de trabalho final de arquitetura e onde está inserido dentro da estrutura da instituição, é preciso entendê-lo também como um acervo cultural. Dado o valor cultural que esses trabalhos possuem (documentos que expressam a interpretação particular de síntese da graduação), as análises sobre este tipo de acervo instigam reflexões no campo da teoria e da prática do ensino, além da sua difusão ser fonte de pesquisa para

diversos públicos, tais como: alunos de graduação e pós, professores, pesquisadores, por se tratar de fonte primária de conhecimento produzido pela instituição⁶.

Utilizamos a visualização da informação para conseguir fazer essas interpretações. Para MANOVICH (2010), a visualização de informações (em abreviação, infovis) pode ser definida como um mapeamento entre dados discretos (valor específico ou categórico) e uma representação visual. Significa que para cada número absoluto e/ou categoria que pode ser extraído do repositório, terá uma representação gráfica proporcional ao montante. É também importante que esta representação gráfica seja a mais clara possível sobre a informação que a representa, pois o contrário poderá gerar entendimento equivocado sobre o acervo analisado. Um caso recente sobre uma má interpretação a partir do design da visualização tendencioso foi a divulgação do quadro de medalhas da Olimpíada de Tóquio 2020 pela mídia nos Estados Unidos da América⁷. Diversos jornais veicularam gráficos manipulados, onde alguns misturavam a organização por somatório do número total de medalhas conquistadas, colocando os EUA como o primeiro na classificação da organização por somatório de medalhas de ouro, o que colocaria a China em primeiro lugar.

No livro “Visual Explanations” de 1997, o autor Edward Tufte narra o caso da explosão do ônibus espacial Challenger em 1986, em que os engenheiros envolvidos no desenvolvimento do projeto divulgaram para os superiores um conjunto de análises favoráveis para a realização do teste, porém estas informações foram feitas de

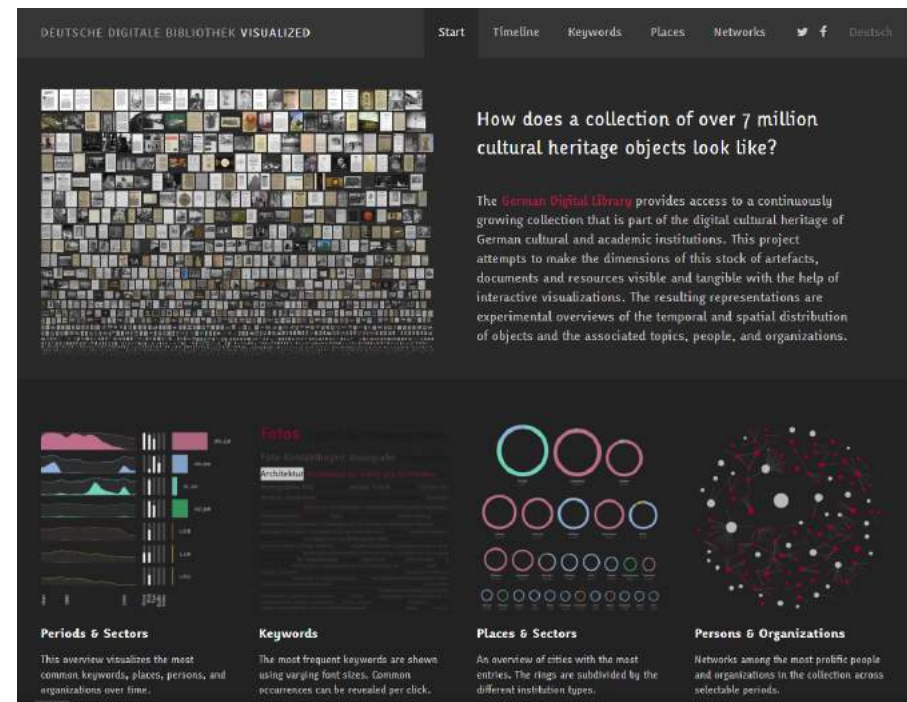
⁶ FREIRE, W.; PARAIZO, R. C. ; KAMIMURA, R. . Curadoria para a exposição digital-interativa da MEDIATECA da FAU/UFRJ. In: XXXVII ENSEA / XX CONABEA - Desafios do Ensino de Arquitetura e Urbanismo no Século XXI, 2019, Rio de Janeiro. Caderno 42 ABEA. Rio de Janeiro: ABEA, 2019. p. 504-514.

⁷ Olimpíada de Tóquio 2021: o 'truque' dos EUA para aparecer à frente da China no quadro de medalhas. Acesso em: 5 ago. 2021. Disponível em: <<https://www.bbc.com/portuguese/internacional-58095518>>

forma simples e com várias anotações e esboços não organizados. Após o acidente, Tufte recriou as visualizações propondo um novo design baseado em ilustrações e gráficos de suporte, com base na fonte primária da informação gerada pelos engenheiros. Ao apresentar os relatórios com os dados, utilizando as novas visualizações em comparação com o anterior, ficava claro que o acidente aconteceu devido às condições climáticas presentes no dia do teste e que não havia sido testada durante a fase de testes, evidenciando a importância do design da informação para compreender as informações que estão sendo visualizadas.

Para buscar a visualização ideal das informações sobre um acervo cultural, não basta ter o domínio pleno das ferramentas de visualização. A construção dessa visualização precisa partir do ponto de vista de um interpretador, alguém capacitado que entenda as particularidades que o objeto cultural possui. Um pesquisador que possui familiaridade com o acervo que será interpretado a partir de uma visualização, terá mais capacidade para determinar se os resultados obtidos estão ou não corretos, além de estar apto a identificar informações secundárias que não são claras no primeiro momento.

Uma boa referência de interpretação de um acervo cultural digital é a visualização do banco de dados da Biblioteca Digital Alemã (Deutsche Digitale Bibliothek) realizado através do laboratório Urban Complexity Lab da Universidade de Ciências Aplicadas de Potsdam (Potsdam University of Applied Sciences) em parceria com a biblioteca⁸. Trata-se de um repositório com acervo digital híbrido, que concentra objetos de várias instituições científicas e culturais alemãs (Imagem 04).



(Imagem 4: Deutsche Digitale Bibliothek Visualized. Fonte: <<https://uclab.fh-potsdam.de/ddb/>>. Acesso em: 05 out de 2021)

A visualização principal está na relação dos setores com o tipo de material em função da data de criação do objeto ou ao período que se refere. Outros dados estão ligados e podem ser interpretados paralelamente, como uma nuvem de palavras a partir dos metadados dos objetos, a localização da obra ou o lugar de referência e o nome da instituição/organização/pessoa que contribuiu para a obra. É possível selecionar um período na linha do tempo para filtrar todos os gráficos ao mesmo tempo, como na imagem 5.

⁸ **Deutsche Digitale Bibliothek Visualized.** Alemanha, 2015. Disponível em: <<https://uclab.fh-potsdam.de/ddb/>>. Acesso em: 05 out. de 2021



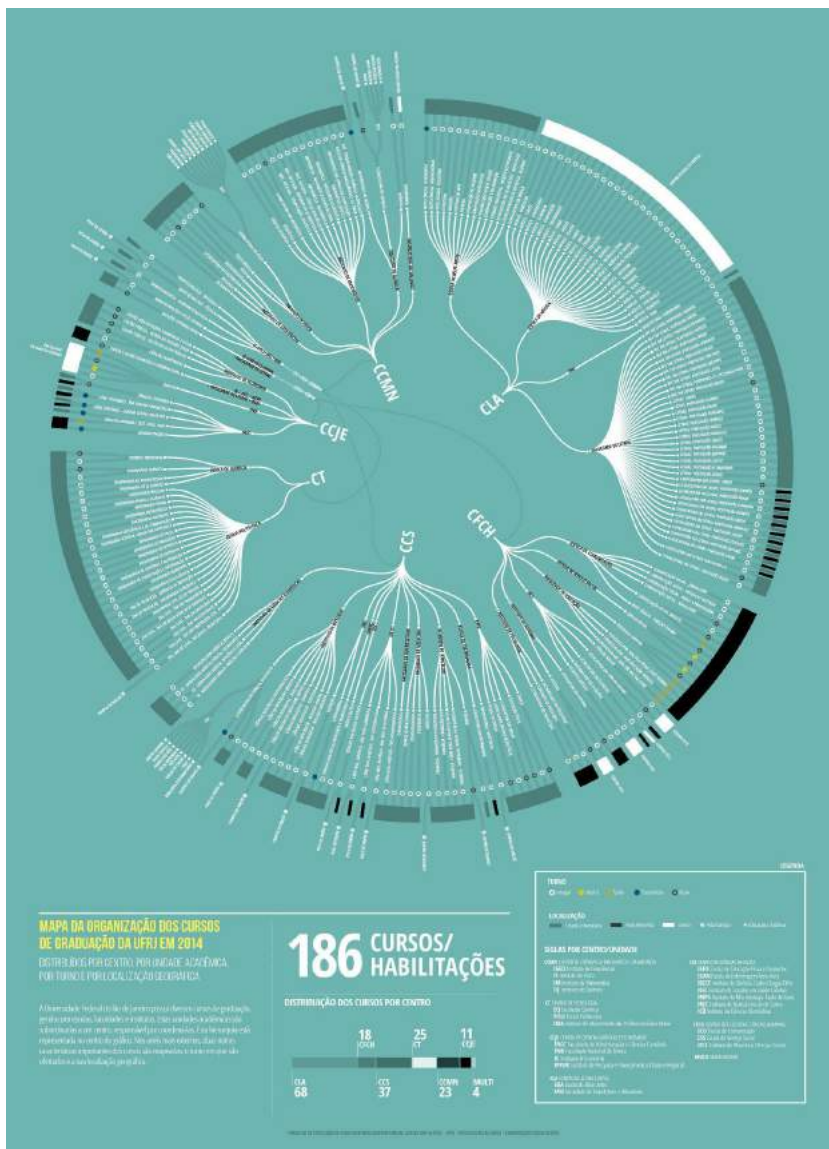
(Imagem 05: Visualização por linha do tempo do projeto Deutsche Digitale Bibliothek Visualized. Fonte: <<https://uclab.fh-potsdam.de/ddb/timeline/>>. Acesso em: 05 de out de 2021)

Outros levantamentos feitos são em formato de nuvem de palavras a partir dos metadados, gráficos em pizza da relação do tipo de material com a localização da instituição a que pertence a obra, e uma rede das instituições/organizações/pessoas que contribuíram para o repositório, mostrando o número de obras fornecidas.

Aproximando para o contexto da Universidade Federal do Rio de Janeiro, o trabalho final de graduação do aluno Gabriel Lira do Nascimento⁹ do curso de Comunicação Visual e Design da UFRJ, sob a orientação de Doris Kosminsky e Claudio Esperança, trata-se de um conjunto de pranchas com várias visualizações de dados sobre a UFRJ, lançando luz sobre informações e relações não tão óbvias, além de um panorama geral sobre a estrutura organizacional dos cursos de graduação. Seu levantamento foi feito a partir de consultas a publicações e documentos disponibilizados publicamente, além de entrevistas com funcionários. Mesmo que seu trabalho não se refira a um objeto físico - como uma pintura, por exemplo, ao analisar a instituição como um todo, podemos interpretá-la também como um objeto cultural, com grande relevância em estudar o perfil do aluno de graduação em relação aos cursos vigentes dentro do estudo realizado. Uma das pranchas de visualização gerada em seu trabalho pode ser vista na imagem 6.

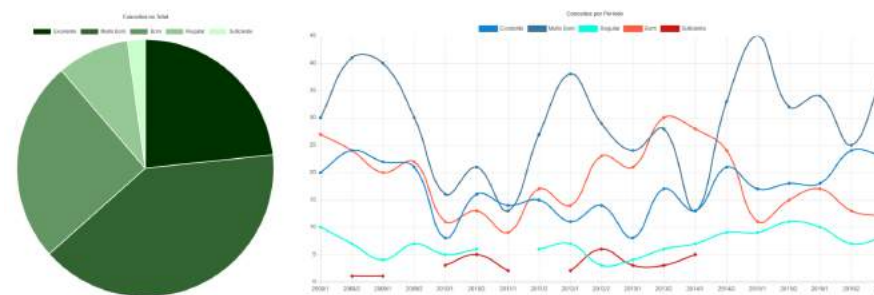
A análise de um repositório de TFG de arquitetura a partir da visualização da informação é a ampliação dessas análises para um caso particular, que busca expor informações da própria fonte e também de possíveis cruzamentos de dados. Mais do que a difusão do acervo para realização de consultas e pesquisas - papel do portal da Miateca, a interpretação da coleção de TFGs de arquitetura da FAU-UFRJ pode evidenciar uma estética, perfis de alunos, tendência de termos nos discursos e etc, que tornam únicos a experiência de ensino desta instituição.

⁹ NASCIMENTO, G. L. . **UFRJ Visualizada: infográficos sobre a Universidade do Brasil**. Trabalho de conclusão de curso (Projeto prático em Comunicação Visual e Design) - Faculdade de Comunicação Visual e Design, Escola de Belas Artes, Universidade Federal do Rio de Janeiro. Rio de Janeiro, 2015. Disponível em: <<https://labvis.eba.ufrj.br/projetos/infograficos-sobre-a-universidade-do-brasil/>>. Acesso em: 8 de out. de 2021.



(Imagem 06: Parte do trabalho final de graduação de Gabriel Lira do Nascimento. Fonte: <<https://labvis.eba.ufrj.br/projetos/infograficos-sobre-a-universidade-do-brasil/>>. Acesso em: 08 out. de 2021)

Em 2017 iniciou-se estudos sobre este acervo a partir de uma pesquisa de Iniciação Científica. O trabalho apresentado na Semana de Integração Acadêmica (SIAC) sob o título “Depuração e difusão da base de dados da MEDIATECA FAU-UFRJ”¹⁰ realizou experimentações iniciais sobre o conjunto de dados reunidos até aquele momento no banco de dados da MEDIATECA, resultando em amostragens por conceito (nota atribuída ao trabalho), relação de conceitos por período, quantitativo de TFGs que estão sendo recuperados (digitalização de material impresso, migração de mídias em CD/DVD para o banco de dados), quantitativos de TFGs por temas e relação tema por período (Imagem 07).



(Imagem 07: Gráfico quantitativo de TFGs por conceito e relação entre período e conceitos. Fonte: MEDIATECA)

A investigação dessas informações são promissoras, principalmente quando analisamos ao longo do tempo, porém a pesquisa não se aprofundou nos resultados. O design e as ferramentas utilizadas para a visualização também foram

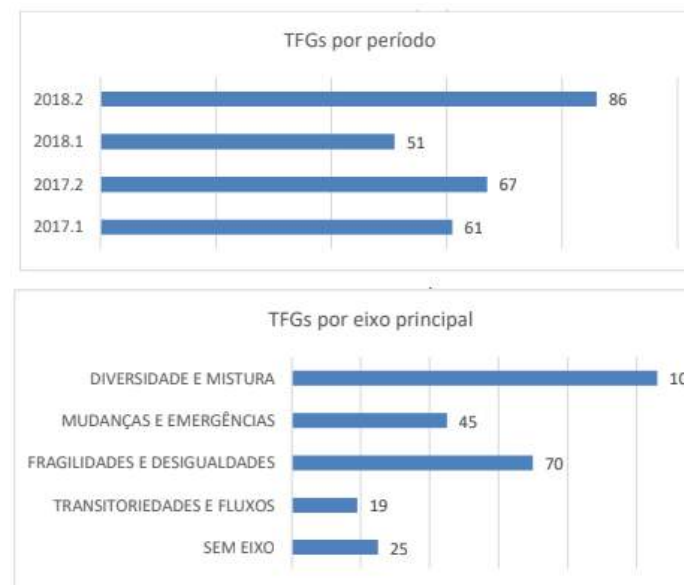
¹⁰ PEREIRA, G. S.; NETO, N. M. B. **Depuração e difusão da base de dados da MEDIATECA FAU-UFRJ**. Rio de Janeiro. 26 out. de 2017. Apresentação oral e slideshare. 15 slides, colorido. Apresentação para Semana de Integração Acadêmica da Universidade Federal do Rio de Janeiro

insuficientes para interpretar a complexidade que um TFG possui, reduzindo as informações em gráficos vetoriais simples com os quantitativos dos dados. A divulgação destes resultados se mantém restrita, pois foram apresentados apenas em formato de imagem em apresentação da SIAC de 2017, precisando de solicitação à equipe da MEDIATECA para acesso a apresentação.

*“By employing graphical primitives, (...)infovis is able to reveal patterns and structures in the data objects that these primitives represent. However, the price being paid for this power is extreme schematization. We throw away 99% of what is specific about each object to represent only 1% - in the hope of revealing patterns across this 1% of objects characteristics.” - Lev Manovich. **What is Visualization**. 2010.*

Uma investigação mais aprofundada sobre uma fatia deste acervo foi realizado pelos autores FREIRE, PARAIZO e KAMIMURA (2019) com os resultados obtidos divulgados no artigo sob o título “Curadoria para a exposição digital-interativa da MEDIATECA FAU/UFRJ” para o Encontro Nacional sobre o Ensino de Arquitetura e Urbanismo realizado em 2019 no Rio de Janeiro¹¹. No artigo, os autores demonstraram a metodologia utilizada para selecionar os TFGs com base em análises extraídas do banco de dados referente ao recorte dos períodos de 2017.1 à 2018.2 com os eixos temáticos do evento UIA 2020.

¹¹ FREIRE, W. V.; PARAIZO, R. C. ; KAMIMURA, R. . **Curadoria para a exposição digital-interativa da MEDIATECA da FAU/UFRJ**. In: XXXVII ENSEA / XX CONABEA - Desafios do Ensino de Arquitetura e Urbanismo no Século XXI, 2019, Rio de Janeiro. Caderno 42 ABEA. Rio de Janeiro: ABEA, 2019. p. 504-514



(Imagem 08: Gráfico 1 de quantitativo do número de TFGs por período e gráfico 2 de quantitativo do número de TFGs por eixo temático do evento UIA2020. Fonte: MEDIATECA)

Foi concluído com o conjunto de resultados apresentados no artigo que houve, de forma voluntária, adesão de um número significativo de alunos que desenvolveram projetos que estivessem de alguma forma alinhados com os temas levantados na UIA 2020 realizado no Rio de Janeiro em 2021. Contudo, os dados usados também foram extraídos a partir de uma consulta ao banco de dados e os seus resultados, tal como as análises, não estão presentes no site da MEDIATECA, apenas no artigo citado. Para realizar experimentações com esse acervo, só seria possível a partir da solicitação ao acesso desse banco de dados ou através da extração de alguns dados ocultos do site a partir da metodologia chamada de “raspagem de dados”

(Web Scraping)¹², porém requer conhecimento um pouco mais avançado de linguagem de programação voltado para ciência de dados (campo interdisciplinar que estuda dados estruturados e não-estruturados para a identificação de padrões, auxílio em tomadas de decisão e afins, a partir de análises utilizando a infovis).

Nos anos de 2020 e 2021, a FAU-UFRJ precisou atender às regras sanitárias para combater a disseminação da pandemia de COVID-19 na comunidade acadêmica¹³. Os períodos seguintes foram suspensos de aulas presenciais, tornando de forma provisória o ensino à distância. Este fato impactou diretamente na produção dos TFGs, assim como foi o evento da UIA2020 apontado anteriormente, pois as apresentações precisaram ser realizadas de forma exclusivamente virtual, usando ferramentas de videochamadas. Outros momentos durante a história do ensino na FAU-UFRJ também condicionaram mudanças na formatação e avaliação destes trabalhos, o que indica a importância do fator temporal na análise desse material.

2. OBJETIVOS

O objetivo deste trabalho é compreender o perfil da produção acadêmica a partir do TFG, através da visualização da informação de parte do banco de dados do Portal MEDIATECA. Para isto, pretende-se construir uma aplicação Web com ferramentas de visualização gratuitas e de código aberto para a difusão dos resultados obtidos,

¹² **Scraping Data**. In: DALE, K. Data Visualization with Python & JavaScript. 1ª ed. Califórnia, Estados Unidos da América: O'Reilly Media Inc., 2016

¹³ **UFRJ define ações para a comunidade universitária no combate à COVID-19**. Assessoria de imprensa da Reitoria, Universidade Federal do Rio de Janeiro. Rio de Janeiro, 2020. Disponível em: <<https://ufrj.br/2020/12/ufrj-define-acoes-para-comunidade-universitaria-no-combate-a-covid-19/>>. Acesso em: 8 out de 2021

possibilitando a incorporação no site da MEDIATECA para consulta. A documentação do processo, assim como o projeto da aplicação, serão disponibilizados na plataforma Github (plataforma de hospedagem de códigos, podendo ser públicas ou não) e Colab (ferramenta de desenvolvimento de códigos em linguagem Python), permitindo que pesquisas futuras possam prosseguir a partir da experimentação já realizada.

3. ESTUDOS DE CASO DE DIFUSÃO DE ACERVOS DE TFGs ONLINE

Para a realização de experimentações com o conjunto de TFGs da MEDIATECA, antes é necessário entender que objeto é este. Para MANOVICH (2009), o tempo impacta diretamente na compreensão do estudo de um objeto cultural e, portanto, se faz necessário o levantamento de informações à respeito da construção da base de dados da MEDIATECA, assim como o processo de desenvolvimento de um trabalho final de graduação, quais são essas etapas e quem está envolvido nesse processo. O entendimento de outras plataformas de difusão de acervos de TFGs pode levantar questões a serem abordadas nas análises posteriores e na etapa de construção da ferramenta de visualização de dados.

3.1 CONTEXTO HISTÓRICO NA FAU-UFRJ

Para atender as demandas da portaria nº 1770 de 21/12/94 do CFE e da nova Lei de Diretrizes e Bases de dezembro de 1996, a

Faculdade de Arquitetura e Urbanismo da UFRJ criou a reforma curricular em 1996 que ficou vigente até o ano de 2005¹⁴. Foi a partir desta reforma que foi implementado o trabalho de conclusão de curso de arquitetura chamado de Trabalho Final de Graduação, com as primeiras defesas neste formato realizadas dois anos depois do início do novo currículo. Foi estabelecido o desenvolvimento de projetos de arquitetura e/ou urbanismo de forma individual, podendo abordar qualquer tema pertencentes às disciplinas da graduação. Este currículo enfrentou diversas dificuldades na sua implementação e gerou várias modificações pontuais a partir das demandas do corpo docente e discente, como a reformulação do TFG no ano 2000, dividindo o desenvolvimento do trabalho em dois períodos, abreviados de TFG1 e TFG2.

9º Período						
Código	Nome	Créditos	C.H.G. Teórica/Prática/Extensão			Requisitos
FAW590	Fundamentos para o TFG	4.0	30	60	0	FAE471 (P), FAH486 (P), FAP485 (P), FAU484 (P)
Total de Créditos		4.0				

10º Período						
Código	Nome	Créditos	C.H.G. Teórica/Prática/Extensão			Requisitos
FAWX01	Trabalho Final de Graduação	0.0	300	0	0	FAW590 (P) FAW590 = FAP682
Total de Créditos		0.0				

(Imagem 09: Divisão da disciplina de desenvolvimento do Trabalho Final de Graduação no currículo acadêmico de 1996-2005 da FAU-UFRJ. Fonte: Sistema Integrado de Gestão Acadêmica - SIGA)

Por iniciativa da faculdade motivada pela implantação do novo currículo, em 2004 foi criado o portal “Mediateca do TFG” com o objetivo de promover o acesso a esses trabalhos na internet. Foi criada uma equipe de gestão que tomaria a frente do processo de

¹⁴ **Projeto Pedagógico: Reforma Curricular 2006-1.** Faculdade de Arquitetura e Urbanismo, Universidade Federal do Rio de Janeiro. Rio de Janeiro, 2006. Disponível em: <http://www2.fau.ufrj.br/wp-content/uploads/2017/05/FAU_Projeto_pedagogico_2006.pdf>. Acesso em: 08 out de 2021

armazenar, catalogar e divulgar os TFGs com a criação de uma base de dados, gerando o repositório de TFGs da FAU-UFRJ. Em pouco tempo, as ferramentas disponíveis na época e as estruturas de banco de dados não comportavam o volume de trabalhos que aumentava em função do período. O projeto foi descontinuado e os TFGs criados a partir deste período não passavam por processos adequados de armazenamento, tanto do material físico quanto do digital.

Com financiamento da Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro - FAPERJ, o projeto é retomado no ano de 2013 com o novo nome “Midioteca”. Também foi criada uma pesquisa paralela para a digitalização de material físico e transferência de dados salvos em dispositivos de armazenamento digital (como o CD/DVD), recuperando a memória dos trabalhos não publicados anteriores ao novo portal.

Implantação do TFG

1998

Implantação do Trabalho Final de Graduação (TFG) de forma individual e tema livre

Reformulação TFG

2000

Divisão do desenvolvimento em dois períodos (TFG1 e TFG2)

Mediateca do TFG

2004

Primeira plataforma digital com acesso ao conteúdo completo de TFGs entregues no período (Padrão de armazenamento em CD/DVD)

Midioteca

2013

Lançamento da nova plataforma digital, com visualização parcial dos TFGs (título, resumo, imagem síntese e complementares, metadados)

Adição de outras mídias

FUTURO

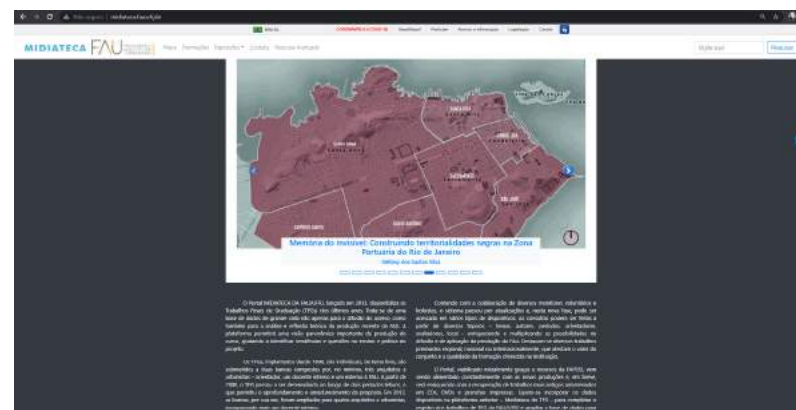
Vídeos e modelos 3D

(Imagem 10: Linha do tempo da Midioteca FAU-UFRJ. Fonte: Autora)

Ao construir o formato preliminar de TFG em 1998 (dois anos após a vigência da reforma curricular de 1996), foi criada uma comissão encarregada de determinar as diretrizes do desenvolvimento do trabalho final, a CCTFG. É função desta comissão orientar os alunos nos procedimentos burocráticos e acadêmicos, resultando em documentos que são distribuídos todo período para os alunos e orientadores. Estes documentos ficam disponíveis para consulta, assim como eventuais avisos e outros documentos auxiliares, no site da FAU-UFRJ. É de responsabilidade dos alunos e orientadores a leitura e a realização dos procedimentos nele indicados, para a validação do trabalho ao longo de seu desenvolvimento.

A comissão é o ponto de partida de um TFG e define alguns processos que irão formatar o resultado final. Após definir as etapas de desenvolvimento solicitando materiais determinados para a realização de cada banca de avaliação, o material resultante (gráfico e textual) da etapa do TFG2 é enviado para a MEDIATECA para a catalogação em paralelo com o preenchimento de um formulário, gerando os metadados do trabalho. A partir da alimentação do banco de dados após a finalização de cada período, os trabalhos são disponibilizados parcialmente no portal para consultas no período subsequente ao produzido.

3.2 PORTAL MEDIATECA



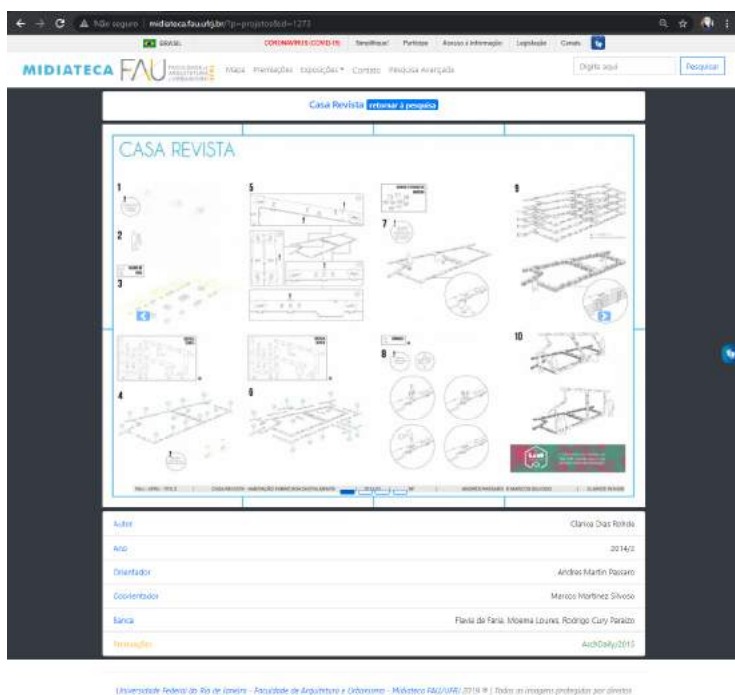
(Imagem 11: Portal MEDIATECA. Fonte: <<http://www.mEDIATECA.fau.ufrj.br/>>. Acesso em: 1 ago. de 2021)

O site da MEDIATECA¹⁵ é uma plataforma web, responsiva para dispositivos móveis, que permite a visualização de parte dos TFGs da Faculdade de Arquitetura e Urbanismo da FAU-UFRJ. Está estruturada com uma página inicial de introdução sobre o portal com uma visualização dinâmica de imagens-síntese de alguns trabalhos premiados, as páginas individuais dos trabalhos, um mapa, exposições de curadoria da própria equipe de gestão da plataforma e uma ferramenta de buscas. Essas buscas no acervo podem ser feitas de forma simples através de uma caixa de texto no canto superior direito, ou a partir da busca avançada utilizando filtros por nome do autor, título, orientadores, membros da banca, intervalo de período, termos presente no resumo, premiações, temas e palavras-chave. A maior parte destes filtros são gerados a partir do preenchimento do formulário enviado à equipe de gestão ao final de cada período,

¹⁵ **MEDIATECA da FAU-UFRJ.** Faculdade de Arquitetura e Urbanismo, Universidade Federal do Rio de Janeiro. Disponível em: <<http://www.mEDIATECA.fau.ufrj.br/>>. Acesso em: 1 ago. de 2021

como citado anteriormente, enquanto outras informações são acrescentadas de acordo com o tempo pela equipe do portal, como é o caso das premiações.

A partir das pesquisas realizadas internamente, a plataforma passa por constantes atualizações. Uma delas que deve ser destacada é o georreferenciamento dos trabalhos que possuem uma localização, resultando em um mapa interativo. Também é previsto a expansão do acervo com a adição dos trabalhos desenvolvidos na primeira etapa (TFG1) - dobrando o número total de trabalhos disponíveis -, e outros formatos complementares de trabalhos, como modelos 3D.



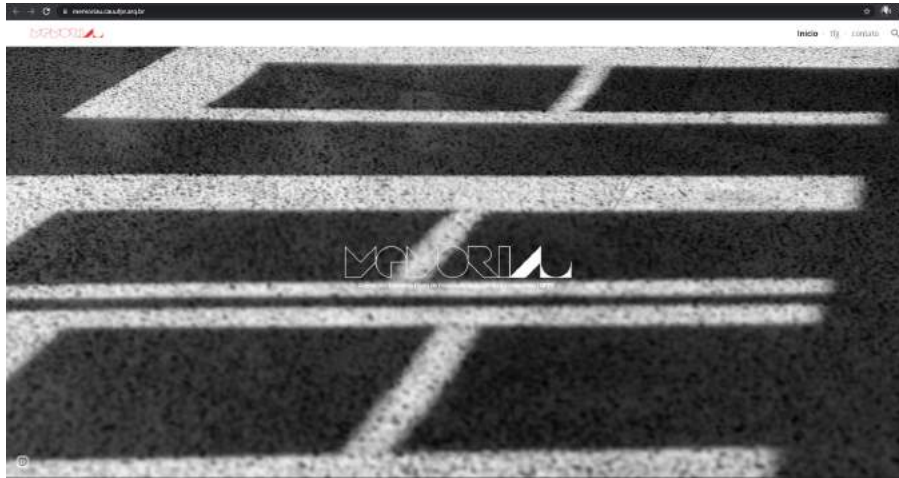
(Imagem 12: Página de visualização do trabalho “Casa Revista” da autora Clarice Dias Rohde. Fonte: <<http://www.midiateca.fau.ufrj.br/?p=projetos&id=1273>>. Acesso em: 1 ago. de 2021)

A apresentação do trabalho no portal acontece de forma parcial, ou seja, não é possível ter acesso ao material original criado pelo autor do trabalho, apenas as imagens selecionadas pelo mesmo, junto com os metadados do TFG. Estes metadados não são os mesmos para todos os trabalhos disponíveis do acervo, já que dependendo do período em que o trabalho foi produzido e defendido, o material pedido pela CCTFG vigente do período e o da equipe da Midiateca podem divergir, assim como a quantidade de imagens depende diretamente da curadoria do autor. Essas questões devem ser levadas em consideração ao analisar esse conjunto de informações disponíveis na plataforma. A escolha pela visualização parcial desses trabalhos não está exatamente clara, porém supõe-se ser uma questão relacionada à direito autoral e intelectual do TFG, já que a Midiateca tem a permissão de visualização, mas precisaria garantir a integridade do material.

3.3 REFERÊNCIAS DE ACERVOS DE TFGs

Segundo os autores FREIRE, PARAIZO e KAMIMURA (2019), o portal da Midiateca serviu como modelo na difusão do acervo de TFGs - de modo online - para outras instituições no país. Algumas semelhanças podem ser percebidas na plataforma “Memoriau”¹⁶ da Faculdade de Arquitetura e Urbanismo da Universidade Federal do Paraná - FAU-UFPR, criada em 2020.

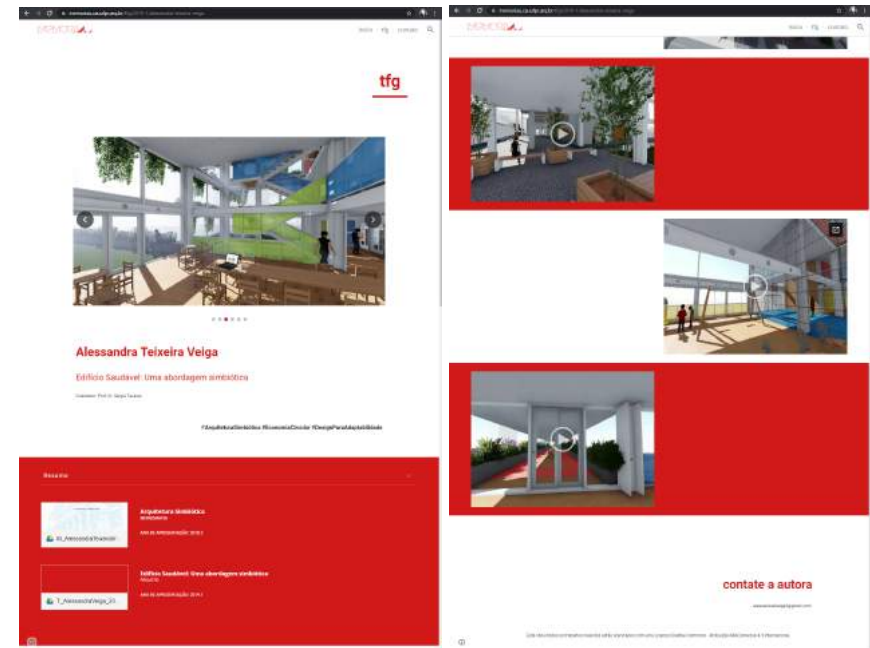
¹⁶ **Memoriau.** Faculdade de Arquitetura e Urbanismo, Universidade Federal do Paraná. Disponível em: <<https://memoriau.cauufpr.arq.br/>>. Acesso em: 09 out. de 2021;



(Imagem 13: MemoriaU. Fonte: <<https://memoriau.cauufpr.arq.br/>>. Acesso em: 09 out. de 2021)

A ferramenta também é responsiva para dispositivos móveis e de estrutura simples, tendo sua página principal com algumas informações a respeito do projeto e a equipe de gestão, instruções para submissão de trabalhos ao banco de dados e algumas imagens do edifício que pertence à Faculdade. A busca pode ser realizada por texto através de barra de pesquisa oculta no canto superior direito, ou por meio da página “tfg”, que divide os trabalhos por períodos. Diferente da MEDIATECA, esta ferramenta não permite busca avançada condicionada por outros tipos de filtros, como as palavras-chave que aparecem na página de um trabalho, mas se destaca ao apresentar o trabalho com os materiais desenvolvidos originais disponíveis para leitura, outros formatos de mídia (vídeos, animações de isométricas, etc) e informações para contato direto com o autor. A formatação da coleção de imagens em destaque na parte superior seguida dos metadados do trabalho também se aproxima do contexto apresentado pela FAU-UFRJ (ver imagens 14 e 15). Também não é possível obter

informações quantitativas sobre os trabalhos ou visualização de informações auxiliares para compreensão do acervo.



(Imagens 14 e 15: Página de visualização do trabalho “Edifício Saudável: Uma abordagem simbiótica” da autora Alessandra Teixeira Veiga. Fonte: <<https://memoriau.cauufpr.arq.br/tfg/2019-1/alessandra-teixeira-veiga>>. Acesso em: 09 out. de 2021)

Com o contexto histórico semelhante ao caso da MEDIATECA, a Faculdade de Arquitetura e Urbanismo da Universidade Federal de São Paulo - FAUSP retoma um projeto anterior de visualização de TFGs (TFG Online) com o lançamento de uma nova plataforma no ano de 2020 em parceria com a Biblioteca Digital de Trabalhos Acadêmicos, instituição responsável pela gestão do repositório

bibliográfico (Dédalus) da USP, sob o nome “Trabalho Final de Graduação: Publicação online”¹⁷.

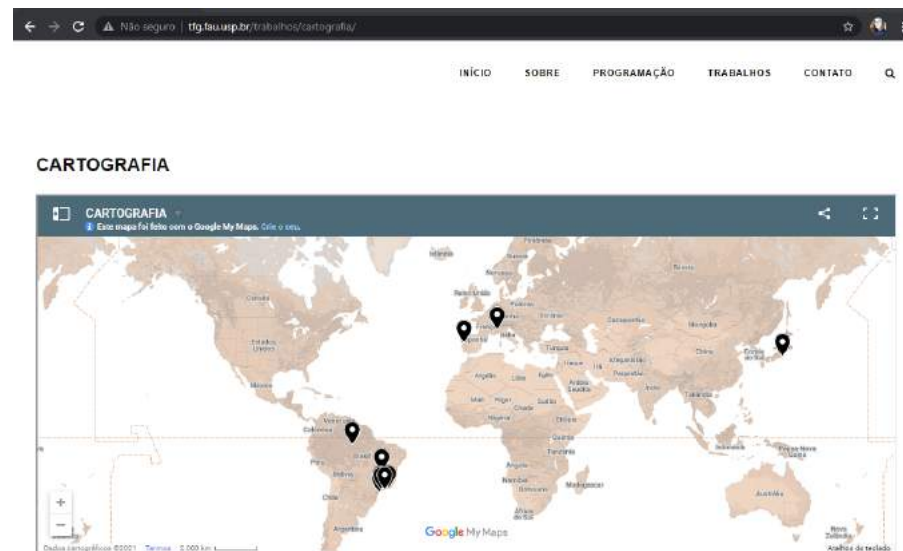


(Imagem 16: Trabalho Final de Graduação: Publicação online. Fonte: <<http://tfg.fau.usp.br/>>. Acesso em: 09 out de 2021)

Diferente das ferramentas anteriormente analisadas, o site também é fonte de divulgação das semanas de banca promovidas pela Comissão de Trabalho Final de Graduação da FAUUSP, que fazem parte do calendário acadêmico da faculdade como semana de integração como na FAU-UFRJ. A busca no acervo funciona por caixa de texto no canto superior direito e os filtros de ano/período, orientação e assunto. Não é possível sobrepor os filtros, uma vez que cada filtro redireciona para uma página com os dados já selecionados. Existe uma visualização por mapa interativo com os

¹⁷ Trabalho Final de Graduação: Publicação online. Fonte: <<http://tfg.fau.usp.br/>>. Acesso em: 09 out de 2021

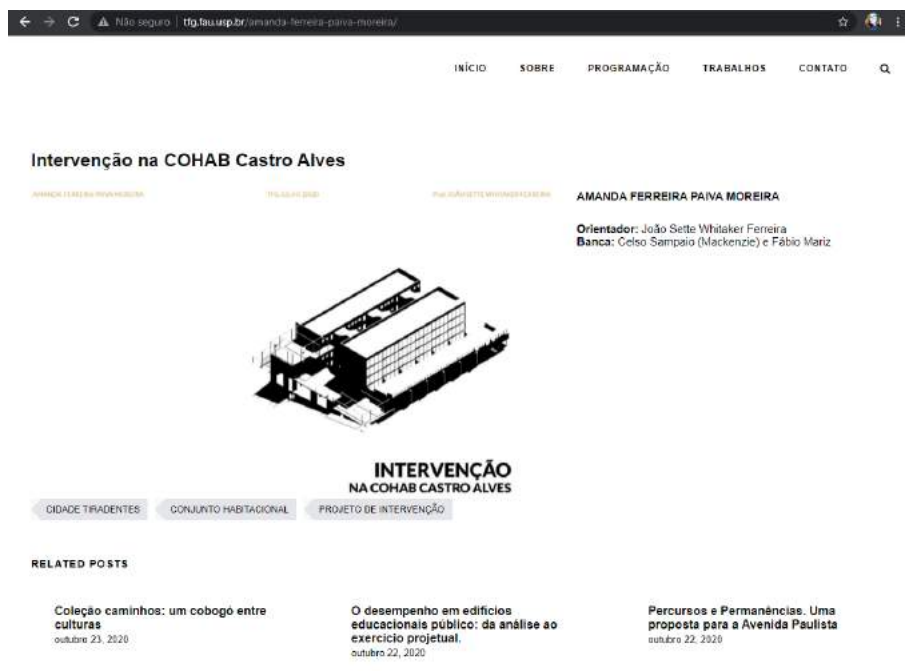
dados dos trabalhos georreferenciados, estrutura semelhante à produzida pela Midiateca (ver imagem 17).



(Imagem 17: Mapa interativo com TFGs georreferenciados. Fonte: <<http://tfg.fau.usp.br/trabalhos/cartografia/>>. Acesso em: 09 out. de 2021)

A página individual de visualização do trabalho mostra apenas uma imagem-síntese (ver imagem 18) com as informações que podem ser buscadas pelos filtros - autor, orientadores, membros da banca -, além de palavras-chave que estão associadas aos temas. Há um aviso em toda navegação de busca indicando que o acervo visualizado é apenas uma parte dos trabalhos publicados. O usuário é redirecionado a página do repositório Dédalus onde está hospedado o trabalho, tornando a plataforma apenas um intermediário entre o público-alvo deste tipo de material acadêmico e o seu local de armazenamento, indicando não ser a mesma equipe que faz a gestão, armazenamento e catalogação do acervo e a equipe de gestão da plataforma estudada. Com exceção da visualização dos trabalhos por

georreferenciamento, não há outras formas disponíveis que possam realizar análises em cima dos dados do acervo.



(Imagem 18: Página de visualização do trabalho “Intervenção no COHAB Castro Alves” da autora Amanda Ferreira Paiva Moreira. Fonte:

<<http://tfg.fau.usp.br/amanda-ferreira-paiva-moreira/>>. Acesso em: 09 out de 2021)

4. OS DADOS DO PORTAL MEDIATECA

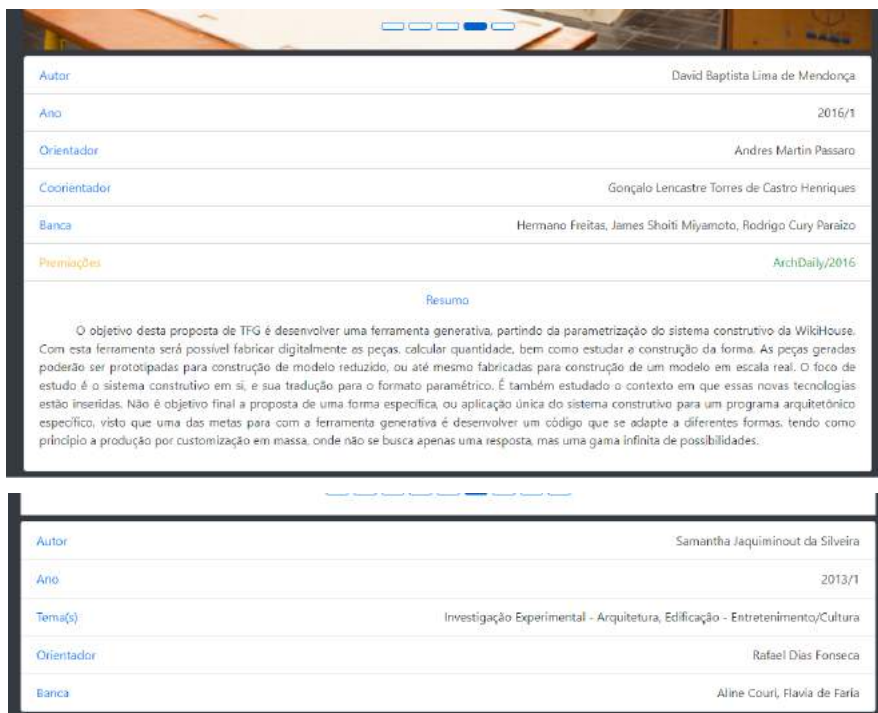
Como já levantado na seção anterior, cada trabalho publicado na MEDIATECA tem parte de seu conteúdo disponibilizado e conta com alguns dados de catalogação da curadoria da equipe gestora do

portal, além de outras informações que são referentes aos trabalhos, como os membros da banca examinadora, que são fornecidos pela CCTFG. Por conta das diferentes gestões, seja da CCTFG seja da MEDIATECA, cada um ao longo dos anos exigia um conjunto de informações dos alunos na hora de cadastrar os trabalhos, resultando na heterogeneidade dos dados. Alguns deles permanecem constantes, pois são tipos de dados que chamamos de **metadados**.

Os metadados são dados que dizem respeito a um objeto e servem para identificá-lo. Podemos definir que o **nome do projeto, aluno, ano de publicação e orientador** fazem parte do conjunto de metadados de um TFG. Outros dados podem compor um conjunto de metadados, como por exemplo os **dados de catalogação** (código identificador, categoria, etc). Este tipo de dado surge quando armazenamos um objeto - físico ou virtual - e precisamos ter a relação de sua localização. Sem nos aprofundar neste assunto, que por si só é amplo e possui bastante divergências quanto a um formato padrão de catalogação entre as instituições de documentação, este trabalho se propõe a entender como o processo de armazenamento da MEDIATECA influencia nos dados, de forma a facilitar ou dificultar a pesquisa.

Outro ponto importante para levar em consideração é sabermos que tipo de objeto estamos armazenando. Como o TFG não tem uma estrutura padronizada e as suas diferentes categorias diferem nos resultados obtidos - projetual, teórico e experimental -, a equipe de gestão da MEDIATECA precisou estudar diversas maneiras de visualização desses trabalhos, o que condiciona em uma lista de dados e objetos que precisam exigir dos alunos para fazer a difusão. Como apresentado na seção anterior, a MEDIATECA armazena apenas as imagens referentes aos trabalhos e seus metadados. De acordo com publicações da própria equipe de pesquisa, a decisão é pautada no entendimento de que as imagens que compõem um TFG fazem

parte da dialética e, portanto, são produtos que podem sintetizar a ideia do projeto. Por este motivo, os trabalhos teóricos não estão armazenados e não podem ser visualizados no Portal, apesar de haver intenções de publicar esses trabalhos.



(Imagens 19 e 20 : Metadados do trabalho com ausência de tema e metadados de trabalho com ausência de resumo, respectivamente. Fonte: Portal Midiateca)

Ainda sobre as imagens, também não há uma uniformidade no tipo de imagem e nem a quantidade delas por trabalho. Esta decisão é totalmente do aluno e quase sempre está em função da representação escolhida para o TFG. Por se tratar de uma entrega de material pós-banca final, muitos alunos não enviam seus trabalhos

para a Midiateca, necessitando de intervenção da Comissão para ajudar na alimentação do repositório, adicionando a etapa do envio do material como obrigatório para poder dar entrada no diploma.

Alguns dados ausentes, como representado nas imagens 19 e 20, nem sempre é em razão de mudança nas resoluções da CCTFG ou da Midiateca. Parte deste apagão de dados é também pela dificuldade no acesso ao material gerado pelos alunos que não enviam seus trabalhos no período solicitado, mobilizando parte da equipe do Portal para a solicitação dos trabalhos.

Outros dados não são visíveis na página dos trabalhos e servem apenas para realizar pesquisas mais avançadas no acervo, como o caso das palavras-chave. Ao darmos entrada no cadastro de trabalho no TFG1 com a Comissão, o aluno precisa escolher uma lista de palavras-chave que definem seu trabalho ou que sejam conceitos abordados. Estas são diferentes dos temas, que são categorias pré-definidas pela equipe da CCTFG e o aluno, no ato do cadastro, precisa escolher dentro da lista fornecida, permitindo a Midiateca disponibilizar esses metadados junto aos trabalhos e realizar buscas com elas.

Existem metadados que não estão visíveis na visualização, porém estão armazenados no banco de dados, como as legendas das imagens. Para a finalidade da visualização, pode parecer não haver importância para essa informação, assim como as notas atribuídas aos trabalhos, porém para fins de pesquisa, podem contribuir bastante para compreender algumas tomadas de decisão do aluno em relação ao projeto.

5. METODOLOGIA

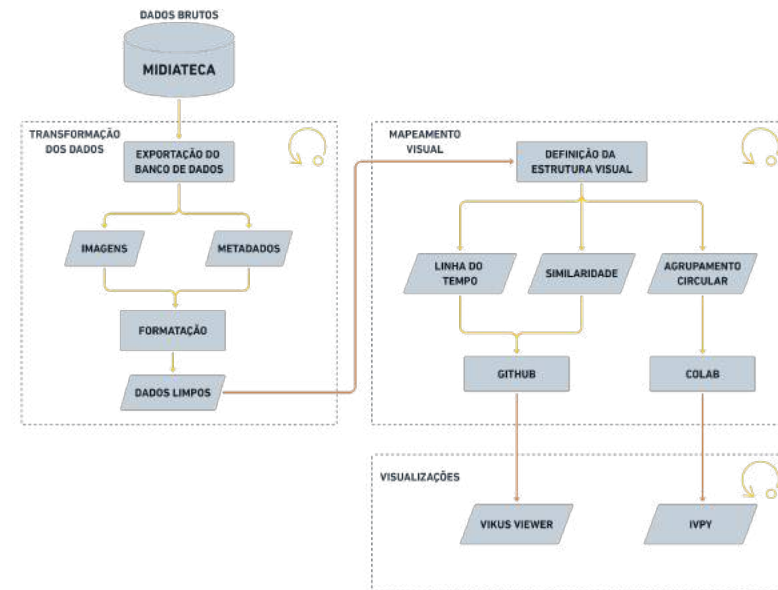
Durante o TFG1, foi necessário realizar um levantamento bibliográfico em paralelo do levantamento de repertório técnico para compreender a viabilidade da proposta do trabalho. A partir dos resultados obtidos primeiramente com a bibliografia, foi construído o pensamento crítico e conceitual acerca dos processos que envolvem a visualização de um acervo cultural, assim como os impactos do design da informação na compreensão das análises sobre os resultados. No segundo momento, fez-se a seleção de ferramentas gratuitas e disponíveis para modificações (open source), dentro dos conhecimentos prévios de programação nas linguagens Python e JavaScript. As ferramentas que não obedeceram esses critérios encontradas durante o levantamento, foram desconsideradas para implementação no TFG2.

A metodologia descrita é semelhante à que Mazza (2009) propõe sobre o processo de criar uma visualização, porém com um nível de complexidade por se tratar de tomada de decisão no uso de ferramentas para a visualização de um conjunto maior. Um exemplo excelente na aplicação desta metodologia é a dissertação de Gabriella Murta que foi apresentado parcialmente na 11ª Semana de Integração Acadêmica da UFRJ¹⁸.

Para o desenvolvimento do TFG2, primeiramente foi feita uma extração dos dados direto do banco de dados da MídiaTeCa através de login e senha, fornecidos pela equipe. Chamamos este tipo de material de **dados brutos**, pois trata-se de material primário, apenas com a formatação da estrutura do banco. Foram extraídos 11 tabelas

no total e uma pasta com todas as imagens disponíveis. A seguir tem a lista dos nomes das tabelas extraídas no dia 06/02/2022:

- alunos.csv
- convidado_banca_projeto.csv
- imagens.csv
- instituicoes.csv
- locais.csv
- palavras_chaves.csv
- professores.csv
- projetos.csv
- projetos_imagens.csv
- projetos_temas.csv
- temas.csv



(Imagem 21: Metodologia adaptada de Ricardo Mazza. Fonte: Autora)

¹⁸MURTA, Gabriella. *Estamparia customizada a partir da visualização de dados*. Formato em vídeo 8:51 min. Disponível em: <<https://www.youtube.com/watch?v=YLdawzJHdVg>>. Acesso em: 18 fev 2022.

Uma vez reunidos os dados brutos, a primeira etapa é a de **transformação dos dados**. Consiste em limpar, organizar e formatar de acordo com as necessidades, seja em função da visualização, seja em uniformização dos dados. Também chamamos este processo de **ETL** (*extract, transform and load*). Neste momento, a revisão das informações é essencial para que não haja equívocos na hora das análises realizadas pelo pesquisador, ressaltando a importância deste processo ser feito por alguém que seja familiarizado com as informações. É também nesta etapa que algumas tomadas de decisão são feitas em relação à formatação dos dados, como lidar com ausência de dados e qual tipo adotar.

Ao final do ETL, o resultado trata-se de um *dataset* ou **conjunto de dados** limpos estruturados - normalmente - em tabela de dados (o formato mais comum é o CSV, mas também há possibilidade de se trabalhar com XLSX, JSON, XML). Para a formatação dos dados da MídiaTeCa, foi utilizada a plataforma Colab da Google e está disponível para visualização e comentários no link: https://colab.research.google.com/drive/1xuz7mBkF_DBGfBUtrS21NUPqKS1zKcpO?usp=sharing.

A etapa seguinte consiste no **mapeamento visual** do que pretende-se estudar. Para MAZZA(2009), este momento determina como serão as visualizações de acordo com os dados que se tem, pois precisa-se espacializá-los em um ambiente, físico ou virtual, para que possa ser construído a visualização. Significa que a forma como vamos estruturar esses dados vão determinar como será a visualização, porém eles precisam ser interpretados para que seja decidido a melhor saída. Dados georreferenciados são tipos de dados que induz a visualização para o formato de mapa, já que são informações que possuem uma posição geográfica relativa (latitude/longitude), os dados temporais podem ser interpretados em

estrutura de linha do tempo e assim por diante, diferenciando apenas o design dessas visualizações.

Existem dados que podem ser difíceis de decidir a sua estrutura de visualização, uma vez que se tratam de valores que se isolados do contexto do *dataset*, podem gerar saídas que sejam de difícil interpretação ou cause confusão.

A partir da interpretação de que as imagens enviadas dos trabalhos são materiais importantes que compõe parte - ou total - do discurso do conceito do TFG, foi decidido que todas as visualizações precisavam tratar desse objeto, a fim de interpretar a representação por trás dos projetos dos alunos da FAU-UFRJ. Por causa disso, outra tomada de decisão, agora em função das imagens, foi a de analisar as cores e a semelhança entre si, resultando em um gráfico de dispersão por similaridade e agrupamento circular em função da cor. Para trabalhar com os dados quantitativos, decidi estruturar as imagens em linha do tempo, a fim de demonstrar a relação entre tempo e trabalhos.

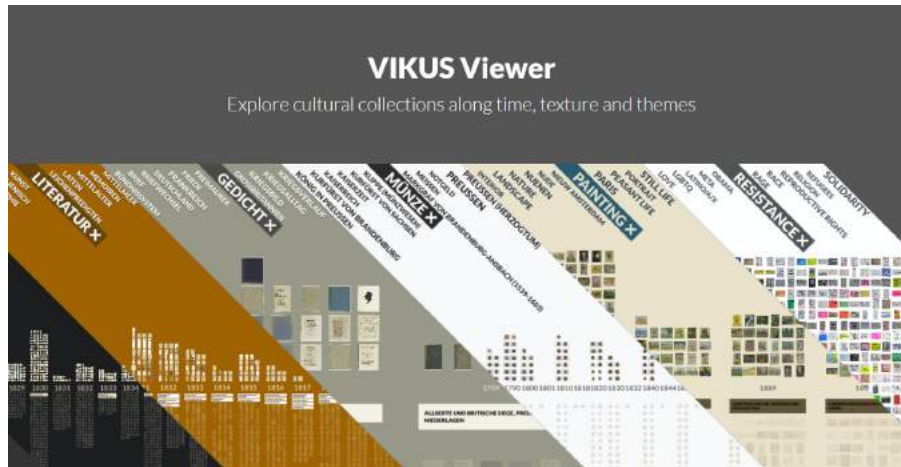
Após a decisão das estruturas visuais e seus ambientes de desenvolvimento (Github para a linha do tempo + similaridade e agrupamento circular no Colab), com os dados organizados para cada tipo de estrutura, realizamos as visualizações e as análises subsequentes.

Para cada etapa, Mazza também propõe que sejam cíclicas e que mesmo estruturado em uma sequência de vetor de direção determinada - a visualização final, elas podem ser repetidas quantas vezes forem necessárias até que sejam satisfeitas todas as perguntas do pesquisador. Também cita a importância da ação humana nas tomadas de decisão, pois é através da interferência e interpretação do pesquisador, que a máquina é capaz de realizar as visualizações.

Por fim, a visualização é feita em dois lugares distintos: o primeiro conjunto de visualizações através da ferramenta **VIKUS**

Viewer e a segunda pelo pacote de visualização de iconografias em linguagem Python **ivpy**. Todas as ferramentas comentadas nesta seção serão detalhadas melhor a seguir.

5.1 REPERTÓRIO TÉCNICO



(Imagem 22: Página de apresentação do projeto VIKUS Viewer. Fonte: <<https://vikusviewer.fh-potsdam.de/>> Acesso em: 01 set. de 2021.)

Desenvolvido por Christopher Pietsch a partir de experiências anteriores (pesquisas e projetos entre 2014 e 2017) no Urban Complexity Lab da University of Applied Sciences Potsdam na Alemanha, esta ferramenta permite criar um acervo digital iconográfico ordenado por tempo, similaridade, categorias e georreferenciamento, totalmente open source através de uma aplicação web voltado para coleções culturais. O projeto está disponível para download na página do GitHub do autor

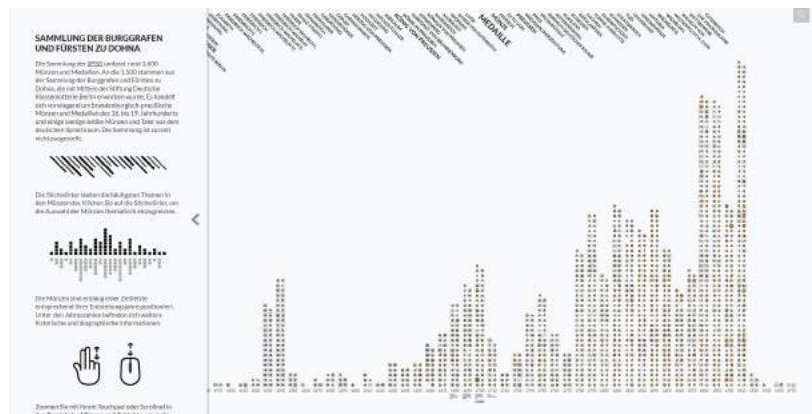
(<https://github.com/cpietsch/vikus-viewer>), acompanhado de tutorial e exemplos.

A ferramenta realiza navegação entre as imagens em alta resolução com suas informações (metadados) disponíveis ao selecionar uma imagem individualmente (ver imagem 23). A organização principal é em formato de linha do tempo horizontal, dispondo as imagens referentes a cada conjunto de tempo em sentido vertical. É possível adicionar textos para cada fatia vertical, de forma a trazer mais informações, ou contextualizar, aquele agrupamento. Isso permite analisar um conjunto de dados a partir das mudanças ao longo do tempo, construindo visualmente a memória por trás de cada objeto.

Para demonstrar o funcionamento da ferramenta, foi estudado visualizações de diferentes coleções de instituições que foram utilizadas durante o desenvolvimento do VIKUS Viewer, servindo de exemplos possíveis para se trabalhar. As moedas prussianas são parte da coleção “Sammlung der Burggrafen und Fürsten zu Dohna” da fundação alemã Preussische Schlösser und Gärten Berlin-Brandenburg, datadas entre o século XVI e XIX.

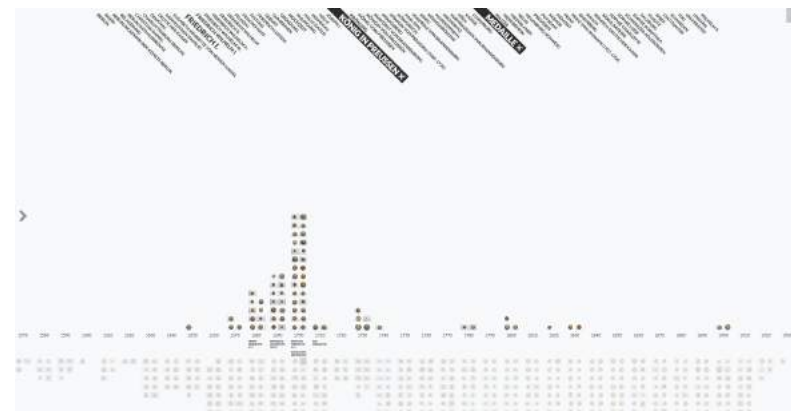


(Imagem 23: Visualização individual de uma moeda prussiana e seus metadados. Fonte: <<https://vikusviewer.fh-potsdam.de/muenzen/vis/>> Acesso em: 01 set de 2021.)



(Imagem 24: Visualização em linha do tempo das moedas prussianas. Fonte: <<https://vikusviewer.fh-potsdam.de/muenzen/vis/>> Acesso em: 01 set de 2021.)

É possível filtrar as moedas a partir das categorias disponíveis no canto superior e valem para qualquer tipo de visualização que esteja disponível, como linha do tempo, similaridade, mapa (no caso particular desta coleção, está disponível apenas a linha do tempo), podendo sobrepor os filtros e/ou realizar busca por texto. Ao filtrar o conjunto, o resultado da busca fica na parte superior da linha do tempo enquanto o restante tem a opacidade reduzida, sendo movida para abaixo da linha. Essas categorias foram definidas a partir dos temas presentes nas moedas, assim como os eventos que elas estão ligadas (como comemorações de aniversário, homenagem a alguma figura pública, etc.). Alguns fatos históricos e biografias relevantes são adicionados como quadro informativo abaixo do ano correspondente.



(Imagem 25: Conjunto de moedas prussianas filtradas por categoria. Fonte: <<https://vikusviewer.fh-potsdam.de/muenzen/vis/>> Acesso em: 01 set de 2021)

As outras visualizações funcionam de forma complementar à linha do tempo, aprofundando o entendimento sobre a coleção. A visualização por similaridade faz uma análise computacional das imagens em busca de padrões de cor, forma e textura e, uma vez identificado algum padrão, organiza as imagens aproximando as que são semelhantes entre si. Um excelente exemplo desta visualização é a que utiliza parte da coleção do pintor holandês Vincent Van Gogh do Van Gogh Museum. As categorias foram definidas por técnica utilizada pelo pintor e alguns tópicos presentes nos quadros.

A resposta da similaridade (ver imagens 26 e 27) conseguiu separar croquis e rascunhos das pinturas a partir do tom predominante e da luminosidade da imagem. Também identificou temas - como florestas e auto-retratos - a partir da forma e textura que esses quadros possuem. A disposição das imagens foi feita com um algoritmo baseado no t-Distributed Stochastic Neighbor Embedding (t-SNE), um método de visualização utilizado na estatística para um grande volume de dados, atribuindo uma coordenada para cada ponto de acordo com uma relação de agrupamento, podendo ser coordenadas bidimensionais ou

tridimensionais. Esse método gera agrupamentos diferentes se realizado com outros conjuntos de imagens, assim como a alteração da coleção atual utilizada (adição ou subtração de itens) pode responder um outro arranjo nos agrupamentos. Aplicando esse entendimento ao resultado obtido com as obras de Van Gogh, é possível afirmar que os padrões de organização de similaridade encontrados, como os croquis e rascunhos à esquerda e pinturas à direita, não serão empregados para outro conjunto de itens de outro pintor.

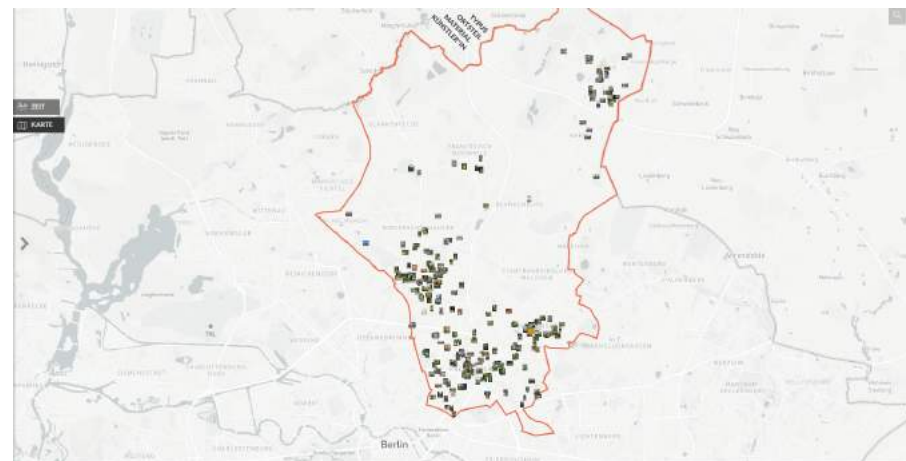


(Imagens 26 e 27: Coleção de desenhos e quadros de Van Gogh ordenados por similaridade e zoom in de agrupamento. Fonte: <<https://vikusviewer.fh-potsdam.de/vangogh/>> Acesso em: 01 set de 2021)

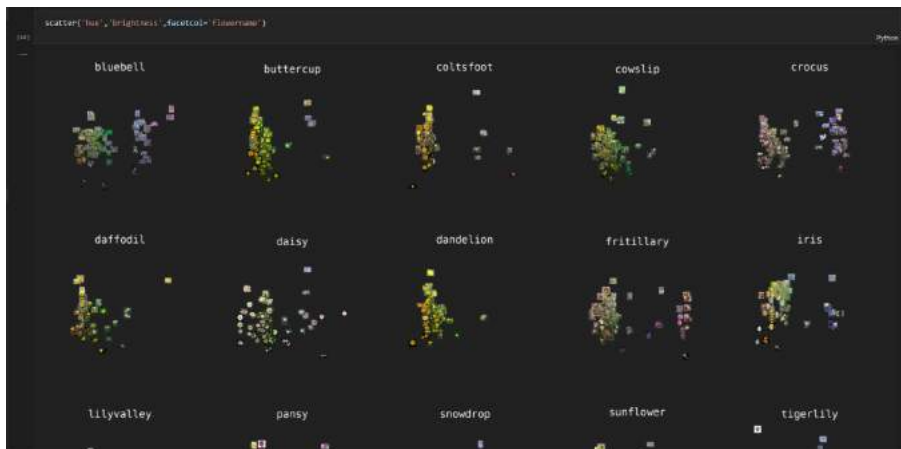
A visualização por mapa é a menos explorada da ferramenta, por exigir que os dados da coleção tenham informações geográficas

em seus metadados. A navegação passa a ser de um mapa em web padrão, permitindo aproximar e distanciar de um determinado ponto, enquanto as imagens são colocadas em suas coordenadas de ponto. É calculado o centro da imagem para que seja usado essa coordenada da imagem ligada à coordenada espacial do mapa, o que pode ocasionar em sobreposições de imagens quando suas coordenadas espaciais são próximas.

A coleção de fotografias de arte em espaços públicos do distrito de Pankow em Berlim, na Alemanha, demonstra as potencialidades do georreferenciamento dos dados. Na visualização por linha do tempo, as fotografias são categorizadas por artista, material, localização (distrito) e tipo de arte, com cada uma tendo suas subcategorias e são organizadas pelo ano de criação. Na visualização por mapa é possível navegar vendo a relação de concentração de artes nos espaços.



(Imagem 28: Mapa de Pankow com fotografias de arte em espaços públicos. Fonte: <<https://vikus.kunst-im-oeffentlichen-raum-pankow.de/>> Acesso em: 08 out. de 2021)

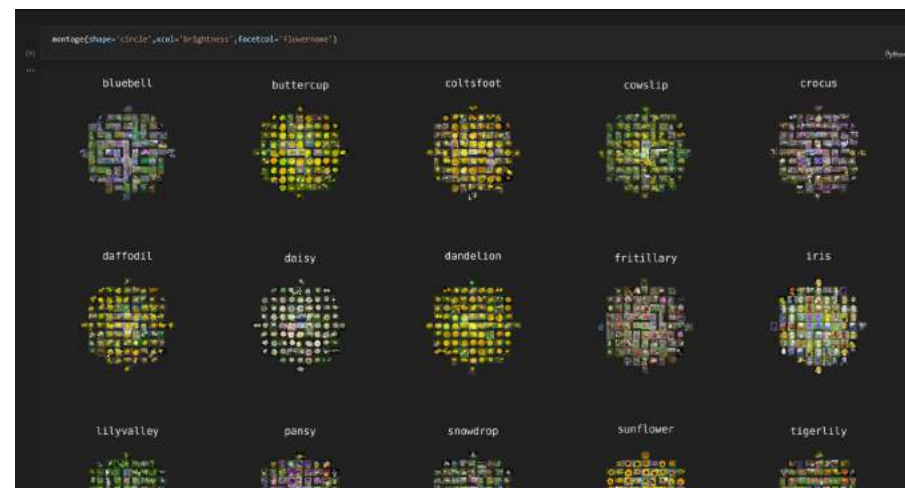


(Imagem 29: Visualização de dispersão estruturada por cor e categorias, utilizando a biblioteca ivpy. Fonte: Autora)

A biblioteca de software desenvolvida em linguagem de programação Python “Iconographic Visualization in Python” - ivpy, foi desenvolvida por Damon Crockett com o objetivo de criar visualizações de dados iconográficos a partir de informações extraídas das imagens e as organizando em agrupamentos (clusters) com coordenadas 2D, voltado para ambientes do tipo notebook¹⁹. O projeto é open source e está disponível para download na página do autor no Github (<https://github.com/damoncrockett/ivpy>), acompanhado de tutorial. O autor do projeto também publicou um artigo contando um pouco mais sobre o processo de criação da ferramenta e a aplicação em um estudo de caso.

¹⁹ Baseados no projeto *Jupyter Notebook*, uma aplicação web *open source* para o desenvolvimento de códigos de programação com hospedagem e processamento dos códigos na nuvem. Os usos mais comuns da aplicação são: manipulação de dados (extração, limpeza e formatação), visualização de dados, modelagem estatística, *machine learning* e afins. Disponível em: <<https://jupyter.org/>>. Acesso em: 02 ago de 2021

As potencialidades desta biblioteca se sustentam na variabilidade de experimentações que podem ser feitas a partir do mesmo conjunto de imagens, uso do processamento em nuvem dos notebooks e saídas no formato de imagem, podendo ser visualizadas no próprio notebook ou salvando os resultados. O processo das análises para as diferentes organizações em clusters também se baseia no t-SNE, como o VIKUS Viewer.



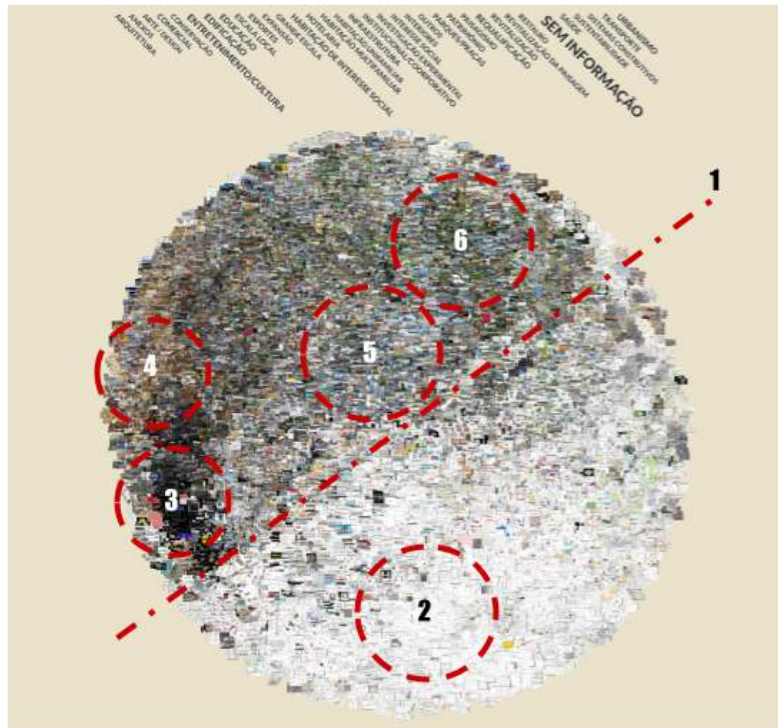
(Imagem 30: Visualização por organização circular ascendente do centro à extremidade, estruturada por luminosidade e categoria, utilizando a biblioteca ivpy. Fonte: Autora)

Permite fazer visualizações em:

- montage: organização com distribuição retangular ou circular, podendo ser crescente ou decrescente;
- histogram: histograma ou gráfico de barras;
- scatter: gráfico de dispersão;
- compose: combinação com mais de um tipo de organização.

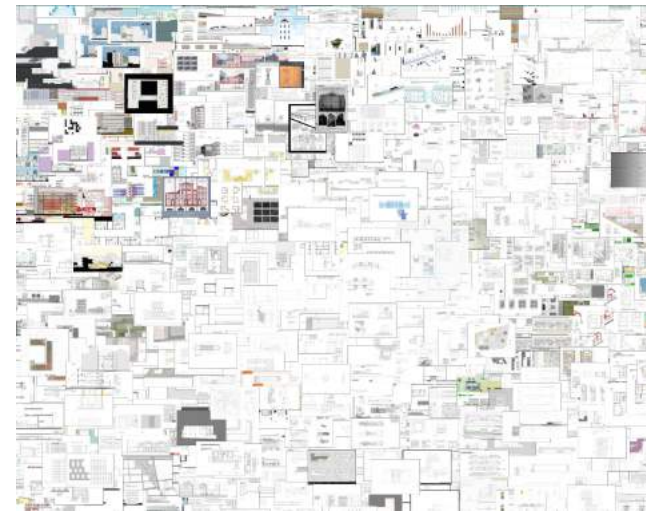
O resultado desta visualização está disponível através do link <https://mechlerc.github.io/vikus-viewer-midiateca/> e para acessar a similaridade, basta escolher a opção “similarity” a esquerda, após ocultar as instruções de navegação.

Diferente dos resultados dos testes realizados em outubro de 2021, com uma fatia dos dados da Midiateca (11 trabalhos), ao colocar todo o conjunto de imagens foi possível notar alguns padrões nas imagens. É através das bordas dos grupos que podemos perceber tendências, como o caso mais evidente entre uma área com cor predominante branca e uma área colorida. A imagem abaixo mostra algumas áreas interessantes para serem analisadas melhor.



(Imagem 32: Agrupamentos de similaridade. Fonte: Autora)

O primeiro agrupamento pode ser entendido como um eixo que parte quase que na metade da visualização. De um lado temos uma grande área com imagens com cor predominante branca, do outro temos as imagens coloridas. Ao aproximarmos da área em branco, percebe-se que em sua maioria são imagens com representações técnicas, como plantas, cortes, fachadas. Tem também croquis digitalizados e trabalhos em tons de cinza, mas o que mais chamou a atenção é perceber que, independente de quantas imagens um aluno produz, ao vermos o conjunto inteiro, as imagens parecem ter um equilíbrio entre a representação conceitual e a representação técnica de um projeto. A partir de extrações das informações das imagens (matiz, brilho e saturação) realizadas com o ivpy, que será explicado mais à frente, foi possível gerar o número de imagens que tem a cor predominante branca, ou uma escala de cinza próximo ao branco absoluto, e o total foi de 5394 imagens, quase metade do número total.



(Imagem 33: Agrupamento 2 - imagens com cor predominante branca. Fonte: Autora)

Assim como a área do branco chamou a atenção, a mancha do preto próxima ao eixo foi intrigante. O script não é 100% perfeito e por isso é possível encontrar imagens com cor predominante preto em outras áreas, inclusive no branco. O fato é que ao analisar por similaridade, outras coisas são levadas em consideração. O processo de análise requer converter a imagem em vetor e, a partir disso, fazer as comparações.

A composição da imagem - elementos geométricos, proporção de ocupação e afins - também parece ser critérios para os agrupamentos, aproximando imagens do mesmo trabalho por terem a mesma linguagem de representação. Ainda sobre a área escura, percebe-se que é de trabalhos com a estética que contrapõe a representação tradicional do papel em branco e elementos escuros destacando. São poucas as imagens que adotam essa estética e, por conta da aproximação das imagens do mesmo trabalho, tende a ser uma região populada pelos mesmos trabalhos, porém há uma tendência de dispersão da concentração, rumo a parte norte do círculo.



(Imagem 34: Agrupamento 3 - imagens com cor predominante preto. Fonte: Autora)

No agrupamento em tons mais amarelados, nota-se imagens de renders de ambientes internos, com a escolha de luminosidade quente. Por conter imagens mais escuras, a dispersão da área preta vem nesta direção. Também há presença da materialidade da madeira ou de móveis em tons pastéis.



(Imagem 35: Agrupamento 4 - imagens com cor predominante amarelo. Fonte: Autora)

As áreas 5 e 6 muitas vezes se confundem e compõe bastante do lado colorido. Mais do que a representação da cor, essas duas áreas falam bastante sobre a representação de ambientes externos. Parece haver uma tendência de usar o céu azul com algumas nuvens enquanto a vegetação, principalmente copa de árvores, um tom de verde, próximo ao verde-bandeira. Pode ser por limitação de bibliotecas de textura acessíveis ou então uma escolha estética dos alunos, porém há um grande volume de imagens com este padrão. Existem trabalhos que não seguem estas cores, mas por compor

proporcionalmente imagens semelhantes (relação chão e céu, vegetação), foram agrupadas próximo.



(Imagens 36 e 37: Agrupamento 5 e 6 - imagens com cor predominante azul e verde, respectivamente. Fonte: Autora)

A visualização por linha do tempo é a que menos resultou em análises significativas. Infelizmente parte deste problema se deve a um bug na escala da representação das imagens e não foi corrigido a tempo da apresentação deste trabalho. Contudo, a visualização estruturada em função do tempo não precisa necessariamente ser abandonada, mas repensar como as imagens organizadas pelos anos podem ajudar a entender os trabalhos.

Uma possibilidade de abordagem é não utilizar cada imagem como um célula da linha do tempo, mas sim agrupar por trabalhos. Reduziria a quantidade de objetos a serem renderizados na visualização por linha do tempo, assim como sintetizaria o volume de trabalhos entregues por ano - um dado mais interessante a ser explorado. Um ponto interessante desta estrutura de visualização é a capacidade de relacionar alguns eventos importantes, como a implantação do modelo remoto em 2020, e a produção dos trabalhos.

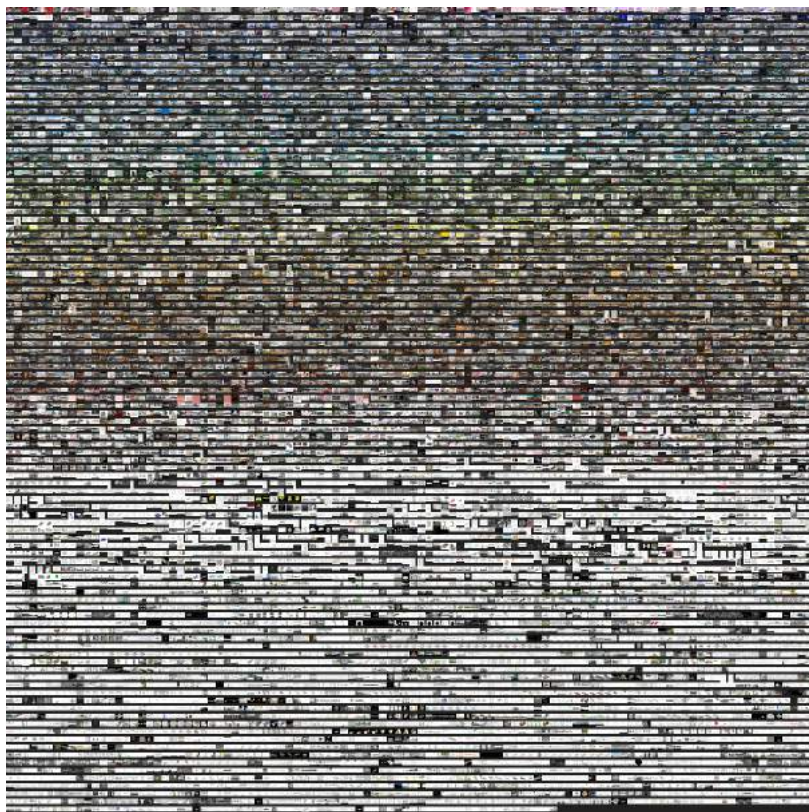


(Imagem 38: Visualização por linha do tempo. Fonte: Autora)

A vantagem de estar trabalhando com uma ferramenta interativa, é a possibilidade de realizar filtros usando a mesma

visualização. Foi utilizado os temas como categorias para filtrar as imagens, porém outras informações poderiam servir de categoria, como professores, palavras-chave, localização do projeto, etc. Basta que esta informação esteja no banco de dados da MEDIATECA e seja adicionada ao CSV de dados para o VIKUS Viewer processar a visualização.

6.2 VISUALIZAÇÕES COM O IVPY



(Imagem 39: Visualização total por matiz. Fonte: Autora)

Os resultados desta visualização, assim como alguns dados quantitativos sobre o *dataset* resultante do ETL, estão disponíveis para leitura e comentários através do link https://colab.research.google.com/drive/1XaBNHqyuKYEcfZ5l_mBefdN8NsSDaMg?usp=sharing. As imagens produzidas neste Colab foram exportadas em PNG e podem ser hospedadas em outro lugar. Por causa da maneira como a biblioteca funciona, não é possível interagir com os resultados de forma dinâmica. Para isto é preciso utilizar outras ferramentas, como a biblioteca de visualização Bokeh, a partir dos resultados extraídos das análises das imagens, salvo em um CSV.

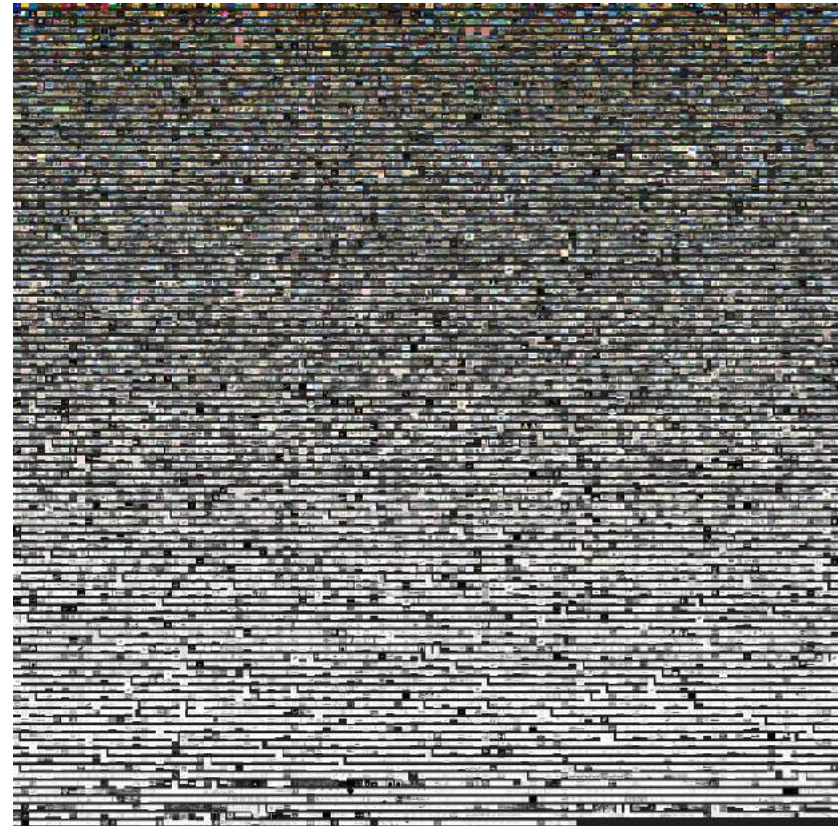
A vantagem de se trabalhar em um notebook é o processamento em nuvem, sem requisição de uma máquina adequada para o trabalho. Como já mencionado anteriormente, o processamento da análise de imagens é demorado e pesado, dedicando bastante dos recursos computacionais. Ao trabalhar na nuvem, o tempo de trabalho é reduzido e permite que o computador não esteja totalmente ocupado com as análises. A extração das informações utilizando o *ivpy* no Colab foi de aproximadamente 3 horas e 40 minutos.

Um ponto negativo para esta visualização é a saída em formato estático (imagem), sem a possibilidade de interação, uso de filtros secundários - como pesquisa por autor e nome de projeto - e de difícil visibilidade das unidades de imagem, necessitando buscar algumas respostas na pesquisa avançada do Portal MEDIATECA. Isto não prejudicou as análises aqui documentadas, apenas dificultou o processo e torna os resultados pouco aproveitáveis para interações com outros pesquisadores.

Foram extraídos apenas as informações a respeito das cores das imagens, trabalhando com a escala HSV (*hue, saturation and value*), porém é possível fazer outras análises, como descrito no artigo de

Damon Crockett²⁰, autor do projeto. Esta escala permite fazer agrupamentos por cada parte da escala, como a matiz, saturação e/ou brilho e juntos podemos definir alguns intervalos de cor, podendo categorizar as imagens em função de sua cor, como por exemplo: imagens que estão dentro do intervalo de cor predominante ciano. É preciso ter ciência que, ao trabalhar com análise de imagens por cor, qual a escala está sendo analisada e de que maneira essa informação é analisada do objeto. Para o ivpy, as informações a respeito da escala HSV estão se referindo à cor predominante na imagem. Existem outras maneiras para se abordar isso, como extrair uma paleta de cores, extrair a informação da cor predominante de fundo e da cor predominante do elemento, etc.

Foi decidido fazer 3 tipos de agrupamentos das imagens: total, por ano e total de um ano (parcial). A imagem 39 é o resultado do agrupamento total por matiz, ou seja, os maiores valores de matiz no topo até os valores iguais a 0 ao final da visualização. Podemos notar novamente um eixo de divisão na metade da composição, corroborando com a visualização por similaridade realizada com o VIKUS Viewer e explicada na seção anterior. Também fica mais claro o intervalo de cores mais comuns, como o azul e o vermelho. Não foram levantadas os quantitativos de imagens de cada intervalo.



(Imagem 40: Visualização total por saturação. Fonte: Autora)

A visualização total por saturação revela a proporção de trabalhos que utilizam cores saturadas nas suas representações, podendo ser parte do partido ou por influência da maneira como a imagem foi construída (renders noturnos, elementos gráficos com cores vibrantes, fundo de prancha colorido são alguns exemplos que podem influenciar na proporção deste resultado). Fato interessante é que o eixo de divisão de imagens parece não estar mais centralizado e sim mais abaixo, algo próximo a 60% da composição da

²⁰ CROCKETT, D. **ivpy: Iconographic Visualization Inside Computational Notebooks**. International Journal for Digital Art History, [S. l.], n. 4, p. 3.60–3.79, 2021. DOI: 10.11588/dah.2019.4.66401. Disponível em: <https://journals.uni-heidelberg.de/index.php/dah/article/view/66401>. Acesso em: 08 out. 2021

visualização. Isso porque as imagens em tons de cinza passam a formar a borda de divisão entre os trabalhos coloridos e brancos.



(Imagem 41: Visualização total por brilho. Fonte: Autora)

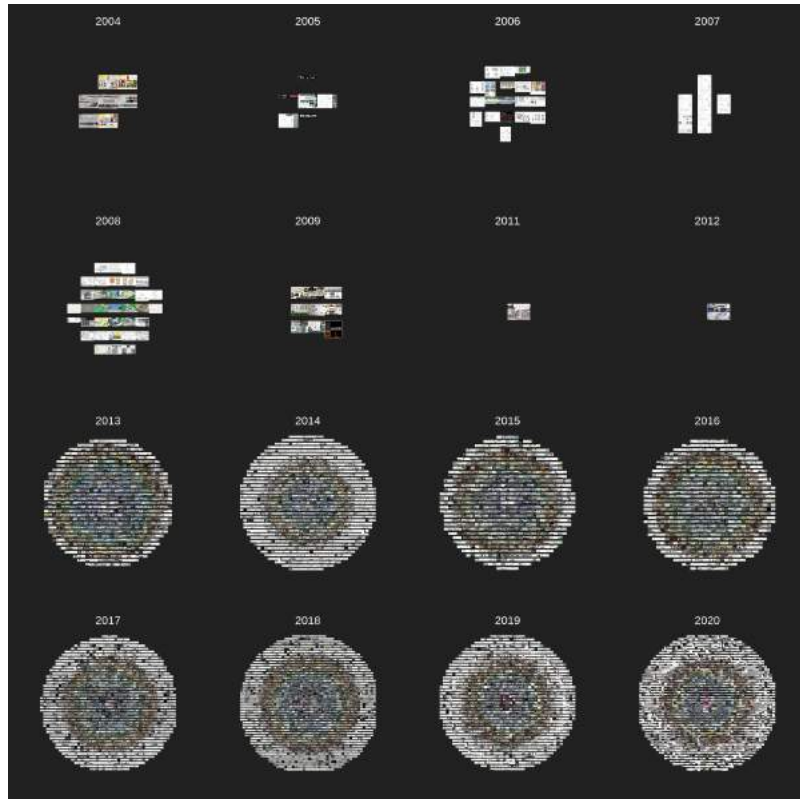
Apesar de ter separado as visualizações por saturação e brilho, essas informações podem ser trabalhadas juntas, resultando no intervalo entre o branco absoluto e o preto absoluto, para quando o valor de *hue* for igual a 0. Na visualização por brilho, destacam-se as imagens técnicas, pois a representação mais comum é trabalhar com o fundo branco e as linhas escuras, tradição de representação vindo das pranchas feitas à mão com nanquim e papel manteiga.

Ironicamente, o desenvolvimento destes produtos costumam ser realizados no software AutoCad, que trabalha em ambiente virtual de fundo preto com hierarquização de linhas através das cores. Os trabalhos que adotam a linguagem inversa na representação, seja técnica seja na prancha resumo, estão ao final da visualização. Por causa da maneira como a visualização por dispersão funciona, sobrepondo imagens semelhantes, pode não ser o suficiente para concluirmos se há um expressivo número de trabalhos que adotam essa linguagem, diferente deste tipo de organização, em que podemos ver lado-a-lado as imagens.

A fim de estabelecer comparações entre diferentes visualizações dos dados das imagens, propus a visualização em função do ano. Esta estrutura permite encontrar padrões ao longo do tempo ou até mesmo a indicação que não há uma tendência. Os dados entre os anos de 2004 e 2012 podem ser descartados, pois não há volume de dados suficiente para análises.

A imagem 40 representa o agrupamento por matiz em função do ano de publicação dos trabalhos. No ano de 2013, ano de publicação da nova plataforma Midiateca, houve um número significativo de imagens coloridas. A média do número de imagens entregues por projeto é algo em torno de 5 imagens por trabalho e o desvio-padrão reduz esse número, por volta de 3 imagens. Pode indicar que a tendência na entrega do material neste ano seja composto por uma prancha síntese e algumas imagens, sejam elas técnicas, sejam elas sínteses. No ano seguinte há uma ruptura completa neste padrão, pois a área de imagens com fundo branco passa a ocupar uma faixa maior na borda, quase metade do raio, além de ter proporcionalmente um diâmetro maior que a visualização do ano de 2013. Em números absolutos, foram enviadas 785 imagens em 2013 e 1205 imagens em 2014, porém o número total de projetos entregues nestes mesmos anos são próximos, 138 no primeiro e 134

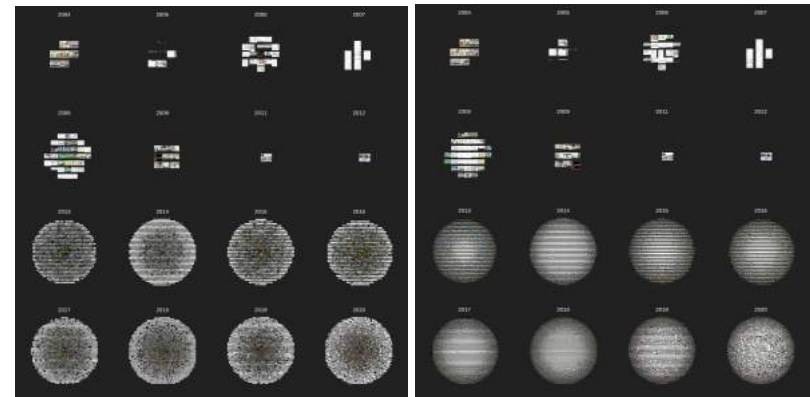
no segundo, indicando que a quantidade de trabalhos não interferiu no volume de imagens e nem na representação dos materiais enviados.



(Imagem 42: Visualização por matiz em função do ano. Fonte: Autora)

Uma possibilidade que explique estes resultados pode estar na diferença entre as resoluções da CCTFG e Miateca para a formatação e envio dos trabalhos. No ano de 2014 houve uma greve nacional das instituições federais e houve bastante adesão da parte da comunidade da FAU, o que pode ter influenciado no desenvolvimento dos trabalhos. Entre os anos de 2015 e 2016 é

retomado a proporção cor/branco encontrada em 2013 e novamente o desvio padrão no número de imagens enviadas é baixo, agora sendo em torno de 2 imagens por trabalho. Pesquisei alguns trabalhos destes três anos a fim de entender que tipo de imagem poderia ser e encontrei os seguintes tipos: **imagem síntese, prancha resumo, fragmentos da prancha resumo e/ou apresentação** - este último encontrei em diversos trabalhos em diferentes anos, aparentemente é um costume do aluno enviar fatias da prancha como material do TFG ao invés do material primário. De 2017 até 2020 parece haver uma retomada da proporção encontrada em 2014, também explicado no número do desvio padrão alto destes anos, com exceção de 2019. Vale ressaltar que o ano de 2020 possui 3 períodos aglutinados, diferente dos anos anteriores.

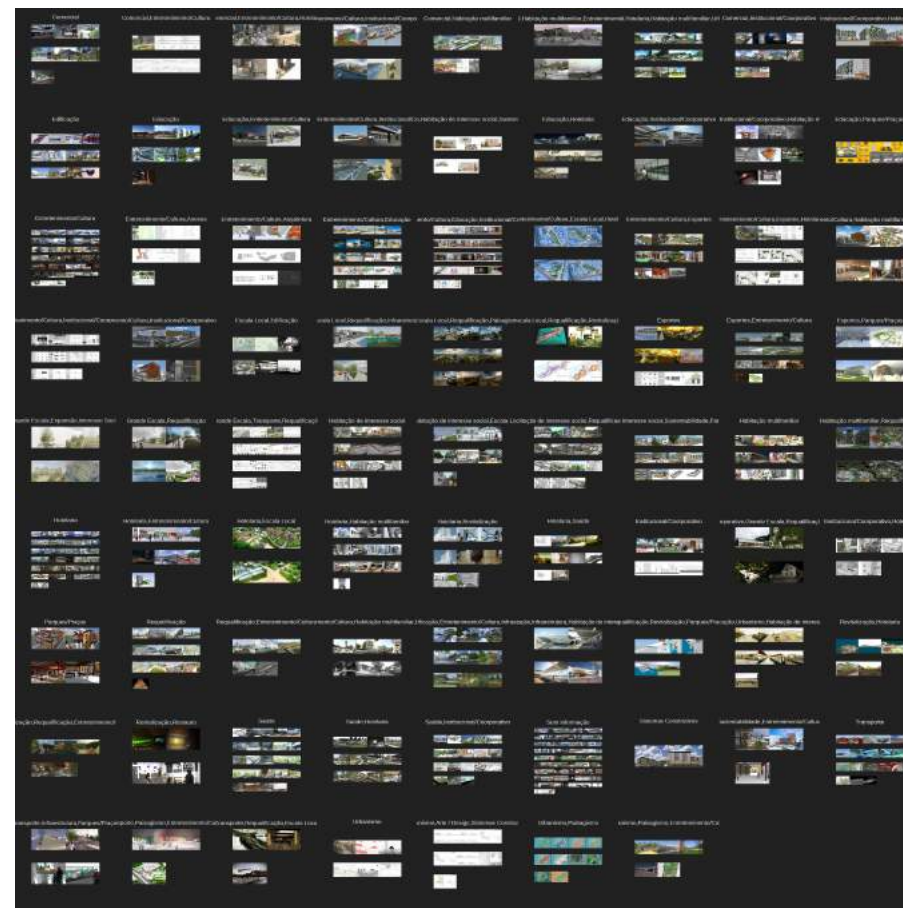


(Imagens 43 e 44: Visualização por saturação e brilho em função do ano, respectivamente. Fonte: Autora)

Foi experimentado também trabalhar com uma fatia dos dados, no caso o ano de 2013. Relacionei os valores de matiz em função dos temas de trabalhos, a fim de revelar se há algum indicativo entre a cor predominante e o tema do trabalho. Há alguns problemas nesta visualização e em grande parte causados pelo fatiamento dos dados.

Assim como não é possível realizar análises nos dados entre 2004 e 2012 por insuficiência de material, o mesmo ocorre nesta visualização. Para a estatística, há dois aspectos fundamentais: **categorias** e o **volume de dados**. As categorias servem para você realizar agrupamentos, como por ano e temas realizados nesta pesquisa, enquanto o volume de dados permite estabelecer padrões e formular hipóteses. Estas categorias podem variar de número e se subdividir em categorias menores, ou seja, o número de categorias é flexível e pode ser reduzido ou grande, dependendo do direcionamento das perguntas. Já o volume de dados pode comprometer as análises caso seja um número baixo. Não há um consenso de valor médio adequado, vai depender da pesquisa que está sendo realizada. No caso particular da visualização das imagens da MEDIATECA, o volume de dados do ano de 2013, apesar de ser um número expressivo, não é suficiente para concluirmos se agruparmos por temas (ver imagem 45).

Outro problema identificado é a forma como foi construído a coluna de temas na planilha de dados. Um mesmo trabalho pode ter mais de um tema ou um subtema, como por exemplo *habitação* é subtema de *edificação*. Isso faz com que tenha entradas do tipo lista, agregando vários temas e subtemas. Porém para o computador, significa uma entrada diferente dos demais, já que **edificação** é diferente de **edificação**, **habitação** e que também é diferente de **edificação**, **habitação**, **trabalho experimental**, resultando em mais agrupamentos. Mesmo que o volume de dados seja o total, por conta desses agrupamentos menores pode fazer com que a visualização não seja o suficiente, precisando repensar ou a forma de categorização dos trabalhos (algo na ordem da CCTFG e MEDIATECA), seja no agrupamento por outras categorias, como por **professores**. Vai depender de qual direcionamento a visualização terá.



(Imagem 45: Visualização da relação matiz e tema no ano de 2013. Fonte: Autora)

7. PERSPECTIVAS

O desafio de se trabalhar com os dados do Portal MEDIATECA estão relacionados a dois fatores: poder computacional e heterogeneidade dos dados. Para vencer o problema computacional, é preciso abrir mão de algumas decisões da pesquisa, como a

interatividade, ou então a ferramenta utilizada, retornando o processo de busca de ferramentas. Parte do processo também se sustenta na divulgação das ferramentas e processos envolvidos para que possa ser replicado em outras pesquisas. Um conjunto interessante de trabalhos a ser estudado são as pranchas resumo desenvolvidas no Atelier Integrado 1. Trata-se de um conjunto homogêneo de material e com consistentes metadados.

Sobre os dados do Portal, algumas abordagens ficaram de fora desta pesquisa relacionadas ao tempo de desenvolvimento e também na dificuldade de manipulação dos dados. Existem muitas entradas ausentes de temas e não está totalmente relacionado ao período inicial de documentação deste metadado (marco temporal), mas também por questões ligadas ao não fornecimento destas informações por parte do aluno, dificultando a análise. Outros metadados com problemas foram entradas duplicadas de alunos e professores, imagens duplicadas e outras sem relação com projetos (mais de 300 imagens descartadas), palavras-chaves sem relação com o conteúdo do TFG (entrada realizada pelo aluno) e entradas de alunos sem projeto correspondente. Estes problemas condicionaram as tomadas de decisão a respeito das visualizações.

Como o campo da representação é amplo, é possível levantar outras abordagens que não seja por cor para a análise dos TFGs. Uma possibilidade interessante é o desenvolvimento de uma inteligência artificial (IA) que será treinada para a identificação de tipos de imagem - render, colagem, planta-baixa, prancha-resumo são alguns tipos possíveis de serem treinados, basta que o pesquisador crie identifique os elementos presentes que caracterizam cada tipo, tenha um conjunto de dados de controle (são dados categorizados manualmente pelo pesquisador) a ser enviado para a IA e a partir da categorização dessas imagens, formular questões

pertinentes à representação de elementos arquitetônicos. Algumas perguntas possíveis de serem estudadas através deste método são:

- Qual é o tipo de representação mais comum da vegetação em plantas técnicas?
- A proporção de céu e chão nas imagens com colagem e/ou render são de 50% para cada?
- Quantas imagens possuem representação de figura humana?
- Quantos trabalhos com estética de croqui foram realizados por cada professor? E por ano?

Para esta abordagem, enfrenta-se questões como poder computacional e maturidade no desenvolvimento de uma inteligência artificial, adicionando complexidade à pesquisa. Este trabalho também se propôs a encontrar o meio termo e de viabilizar ferramentas para pessoas não familiarizadas com o ambiente de desenvolvimento, resultando em instruções e disponibilização dos códigos utilizados.

8. REFERÊNCIAS BIBLIOGRÁFICAS

1. **Acervos digitais nos museus: manual para realização de projetos.** Instituto Brasileiro de Museus; Universidade Federal de Goiás - Brasília, DF: Ibram, 2020. Disponível em: <<https://www.museus.gov.br/wp-content/uploads/2021/05/Acervos-Digitais-nos-Museus.pdf>> Acesso em: 28 set de 2021;
2. CROCKETT, D. **ivpy: Iconographic Visualization Inside Computational Notebooks.** International Journal for Digital Art History, [S. l.], n. 4, p. 3.60–3.79, 2021. DOI: 10.11588/dah.2019.4.66401.

- Disponível em:
<<https://journals.ub.uni-heidelberg.de/index.php/dah/article/view/66401>>.
Acesso em: 27 set. 2021;
3. DALE, K. **Data Visualization with Python & JavaScript**. 1ª ed. Califórnia, Estados Unidos da América: O'Reilly Media Inc., 2016. 556p.;
 4. **Deutsche Digitale Bibliothek Visualized**. Disponível em:
<<https://uclab.fh-potsdam.de/ddb/>>. Acesso em: 05 out. de 2021;
 5. FREIRE, W. V.; PARAIZO, R. C. ; KAMIMURA, R. . **Curadoria para a exposição digital-interativa da Mideoteca da FAU/UFRJ**. In: XXXVII ENSEA / XX CONABEA - Desafios do Ensino de Arquitetura e Urbanismo no Século XXI, 2019, Rio de Janeiro. Caderno 42 ABEA. Rio de Janeiro: ABEA, 2019. p. 504-514;
 6. KOSMINSKY, D. 2015. **Visualização de informação para divulgação científica: uma metodologia**. In: C. G. Spinillo; L. M. Fadel; V. T. Souto; T. B. P. Silva & R. J. Camara (Eds). Anais do 7º Congresso Internacional de Design da Informação/Proceedings of the 7th Information Design International Conference - CIDI 2015. Blucher Design Proceedings, n2, vol.2. São Paulo: Blucher, 2015. ISSN 2318-6968, DOI 10.5151/designpro-CIDI2015-62;
 7. KOSMINSKY, D.; CASTRO, B.; LUDWIG, L. **Existência Numérica**. 1ª ed. Rio de Janeiro, Brasil: Rio Book's, 2018. 288p.;
 8. MANOVICH, L. **What is Visualization?** Manovich, 2010. Disponível em:
<<http://manovich.net/index.php/projects/what-is-visualization>>. Acesso em: 11 ago. 2021;
 9. MANOVICH, L. **Cultural Analytics: Visualizing Cultural Patterns in the Era of "More Media"**. Manovich, 2009. Disponível em:
<<http://manovich.net/index.php/projects/cultural-analytics-visualizing-cultural-patterns>>. Acesso em: 29 ago. 2021;
 10. MAZZA, R. **Introduction to information visualization**. Londres, Inglaterra: Springer Science & Business Media, 2009;
 11. **Memoriau**. Faculdade de Arquitetura e Urbanismo, Universidade Federal do Paraná. Disponível em: <<https://memoriau.cauufpr.arq.br/>>. Acesso em: 09 out. de 2021;
 12. **Mideoteca da FAU-UFRJ**. Faculdade de Arquitetura e Urbanismo, Universidade Federal do Rio de Janeiro. Disponível em:
<<http://www.mideoteca.fau.ufrj.br/>>. Acesso em: 1 ago. de 2021;
 13. MURTA, Gabriella. **Estamparia customizada a partir da visualização de dados**. Formato em vídeo 8:51 min. Disponível em:
<<https://www.youtube.com/watch?v=YLdawzJHdVg>>. Acesso em: 18 fev 2022;
 14. NASCIMENTO, G. L. . **UFRJ Visualizada: infográficos sobre a Universidade do Brasil**. Trabalho de conclusão de curso (Projeto prático em Comunicação Visual e Design) - Faculdade de Comunicação Visual e Design, Escola de Belas Artes, Universidade Federal do Rio de Janeiro. Rio de Janeiro, 2015. Disponível em:
<<https://labvis.eba.ufrj.br/projetos/infograficos-sobre-a-universidade-do-brasil/>>. Acesso em: 8 de out. de 2021;
 15. **Olimpíada de Tóquio 2021: o 'truque' dos EUA para aparecer à frente da China no quadro de medalhas**. Acesso em: 5 ago. 2021. Disponível em: <<https://www.bbc.com/portuguese/internacional-58095518>>
 16. PEREIRA, G. S.; NETO, N. M. B. .**Depuração e difusão da base de dados da Mideoteca FAU/UFRJ**. Rio de Janeiro. 26 out. de 2017. Apresentação oral e slideshare. 15 slides, colorido. Apresentação para Semana de Integração Acadêmica da Universidade Federal do Rio de Janeiro;
 17. **Projeto Pedagógico: Reforma Curricular 2006-1**. Faculdade de Arquitetura e Urbanismo, Universidade Federal do Rio de Janeiro. Rio de Janeiro, 2006. Disponível em:

<http://www2.fau.ufrj.br/wp-content/uploads/2017/05/FAU_Projeto_pedagogico_2006.pdf>. Acesso em: 08 out de 2021;

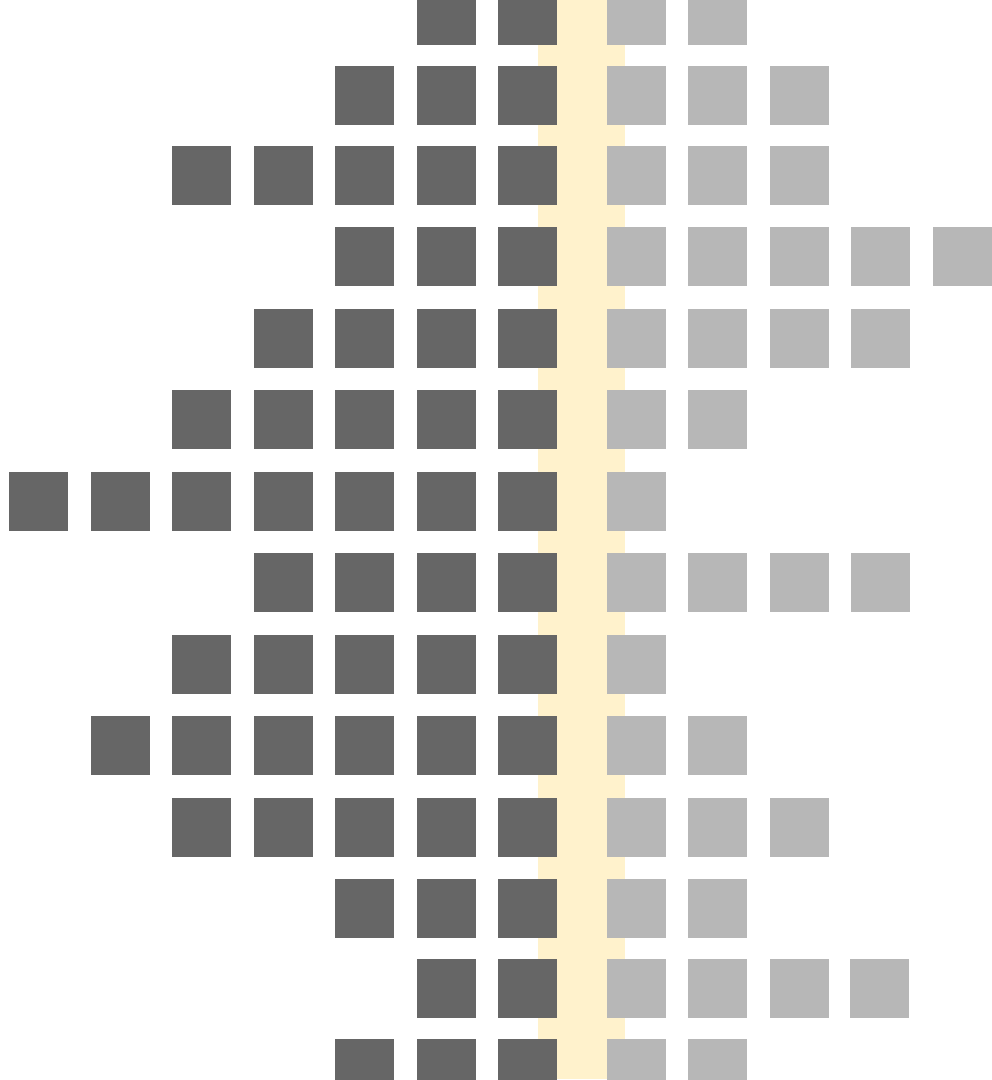
18. **Regulamento do trabalho final de conclusão de curso de graduação em Arquitetura e Urbanismo da Universidade Federal do Rio de Janeiro - FAUFRJ.** Comissão Coordenadora do Trabalho Final de Graduação; Faculdade de Arquitetura e Urbanismo - Rio de Janeiro, RJ: FAU-UFRJ, 2020. Disponível em: <http://www.fau.ufrj.br/wp-content/uploads/2020/04/TFG_Regulamento_Aprovado_11mar2020.pdf> Acesso em: 24 nov 2021;
19. **RESOLUÇÃO Nº 2, DE 17 DE JUNHO DE 2010.** Ministério da Educação, 2010. Disponível em: <http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=5651-rces002-10&category_slug=junho-2010-pdf&Itemid=30192>. Acesso em: 12 jan. 2022;
20. **Retórica gráfica na apresentação dos projetos de AIs.** In: CURY, R.; LEITÃO, T. Reflexões sobre o ensino integrado do projeto de arquitetura. 1ª ed. Rio de Janeiro: Rio Books, 2018;
21. **Trabalho Final de Graduação: Publicação online.** Faculdade de Arquitetura e Urbanismo, Universidade Federal de São Paulo. Disponível em: <<http://tfg.fau.usp.br/>>. Acesso em: 09 out de 2021;
22. TUFTE, E. R. **Visual Explanations.** 7ª ed. rev. Connecticut, Estados Unidos da América: Graphics Press LLC, 1997. 157p;
23. **UFRJ define ações para a comunidade universitária no combate à COVID-19.** Assessoria de imprensa da Reitoria, Universidade Federal do Rio de Janeiro. Rio de Janeiro, 2020. Disponível em: <<https://ufrj.br/2020/12/ufrj-define-acoes-para-comunidade-universitaria-no-combate-a-covid-19/>>. Acesso em: 8 out de 2021.

9. LISTA DE IMAGENS

As imagens podem ser encontradas em seus tamanhos e qualidades originais através do link: <https://photos.app.goo.gl/wob9zohYpL1NEkw4A>

- Imagem 01: Grade curricular do projeto pedagógico 2066 da FAU-UFRJ;
- Imagem 02: Prancha resumo do projeto vencedor do concurso Parklet 4.0, organizado pelo antigo portal Projetar.org;
- Imagem 03: Diagrama de estrutura dos repositórios de uma instituição universitária;
- Imagem 4: Deutsche Digitale Bibliothek Visualized;
- Imagem 05: Visualização por linha do tempo do projeto Deutsche Digitale Bibliothek Visualized.;
- Imagem 06: Parte do trabalho final de graduação de Gabriel Lira do Nascimento;
- Imagem 07: Gráfico quantitativo de TFGs por conceito e relação entre período e conceitos;
- Imagem 08: Gráfico 1 de quantitativo do número de TFGs por período e gráfico 2 de quantitativo do número de TFGs por eixo temático do evento UIA 2020;
- Imagem 09: Divisão da disciplina de desenvolvimento do Trabalho Final de Graduação no currículo acadêmico de 1996-2005 da FAU-UFRJ;
- Imagem 10: Linha do tempo da MEDIATECA FAU-UFRJ;
- Imagem 11: Portal MEDIATECA;
- Imagem 12: Página de visualização do trabalho “Casa Revista” da autora Clarice Dias Rohde;
- Imagem 13: Memoriau;
- Imagem 14: Página de visualização do trabalho “Edifício Saudável: Uma abordagem simbiótica” da autora Alessandra Teixeira Veiga;
- Imagem 15: Página de visualização do trabalho “Edifício Saudável: Uma abordagem simbiótica” da autora Alessandra Teixeira Veiga;
- Imagem 16: Trabalho Final de Graduação: Publicação online;
- Imagem 17: Mapa interativo com TFGs georreferenciados;
- Imagem 18: Página de visualização do trabalho “Intervenção no COHAB Castro Alves” da autora Amanda Ferreira Paiva Moreira;
- Imagem 19: Metadados do trabalho com ausência de tema;
- Imagem 20: Metadados de trabalho com ausência de resumo;
- Imagem 21: Metodologia adaptada de Ricardo Mazza;
- Imagem 22: Página de apresentação do projeto VIKUS Viewer;
- Imagem 23: Visualização individual de uma moeda prussiana e seus metadados;

- Imagem 24: Visualização em linha do tempo das moedas prussianas;
- Imagem 25: Conjunto de moedas prussianas filtradas por categoria;
- Imagem 26: Coleção de desenhos e quadros de Van Gogh ordenados por similaridade;
- Imagem 27: Zoom in do agrupamento;
- Imagem 28: Mapa de Pankow com fotografias de arte em espaços públicos;
- Imagem 29: Visualização de dispersão estruturada por cor e categorias, utilizando a biblioteca ivpy;
- Imagem 30: Visualização por organização circular ascendente do centro à extremidade, estruturada por luminosidade e categoria, utilizando a biblioteca ivpy;
- Imagem 31: Visualização por similaridade de todas as imagens do Portal MEDIATECA;
- Imagem 32: Agrupamentos de similaridade;
- Imagem 33: Agrupamento 2 - imagens com cor predominante branco;
- Imagem 34: Agrupamento 3 - imagens com cor predominante preto;
- Imagem 35: Agrupamento 4 - imagens com cor predominante amarelo;
- Imagem 36: Agrupamento 5 - imagens com cor predominante azul;
- Imagem 37: Agrupamento 6 - imagens com cor predominante verde;
- Imagem 38: Visualização por linha do tempo;
- Imagem 39: Visualização total por matiz;
- Imagem 40: Visualização total por saturação;
- Imagem 41: Visualização total por brilho;
- Imagem 42: Visualização por matiz em função do ano;
- Imagem 43: Visualização por saturação em função do ano;
- Imagem 44: Visualização por brilho em função do ano;
- Imagem 45: Visualização da relação matiz e tema no ano de 2013.

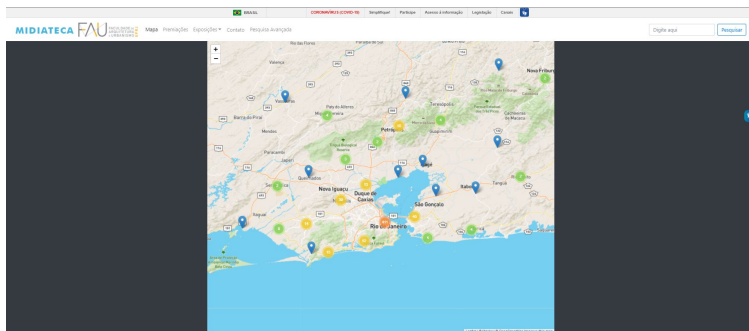
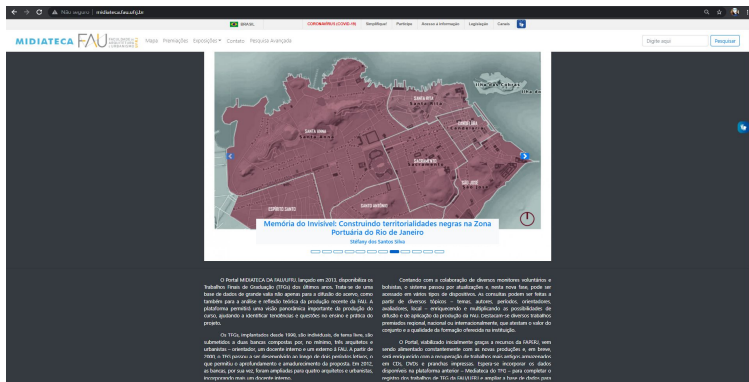


VISUALIZAÇÃO DE DADOS DA MIDIATECA FAU-UFRJ

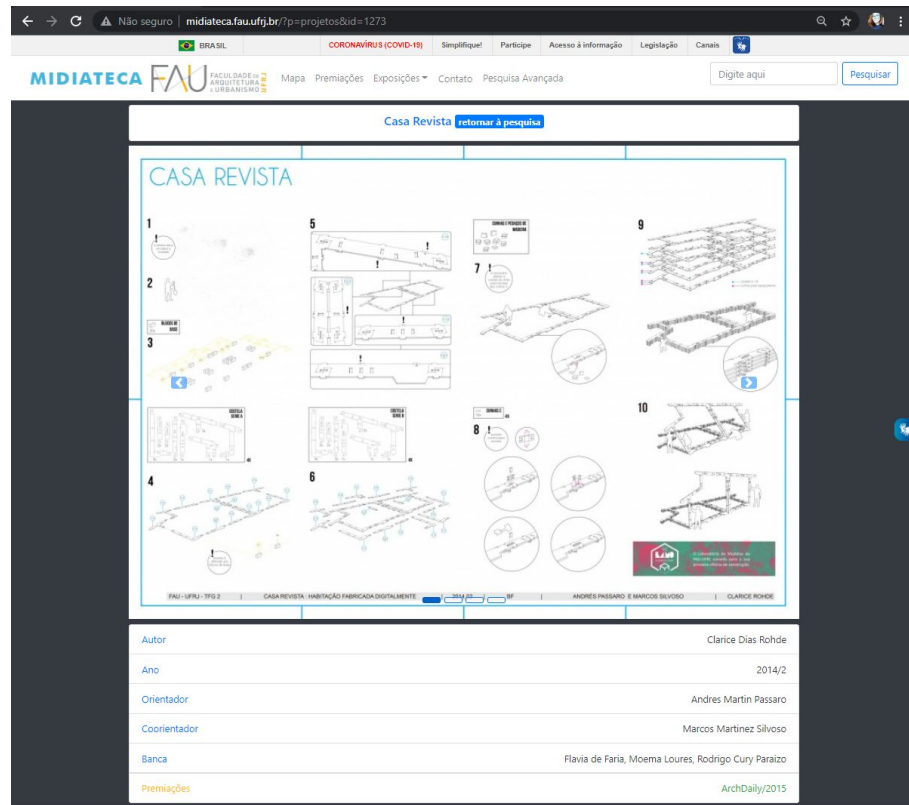
**ALUNA: CÍNTIA MECHLER DE CARVALHO
ORIENTADORES: RODRIGO CURY PARAIZO E
GONÇALO CASTRO HENRIQUES**

ESTUDO FINAL | TFG2 | 2022

PORTAL MEDIATECA



Universidade Federal do Rio de Janeiro - Faculdade de Arquitetura e Urbanismo - MEDIATECA FAU/UFRJ 2015 - 1 - Todos os direitos reservados por direitos autorais.



A Midiateca possui um **acervo híbrido**

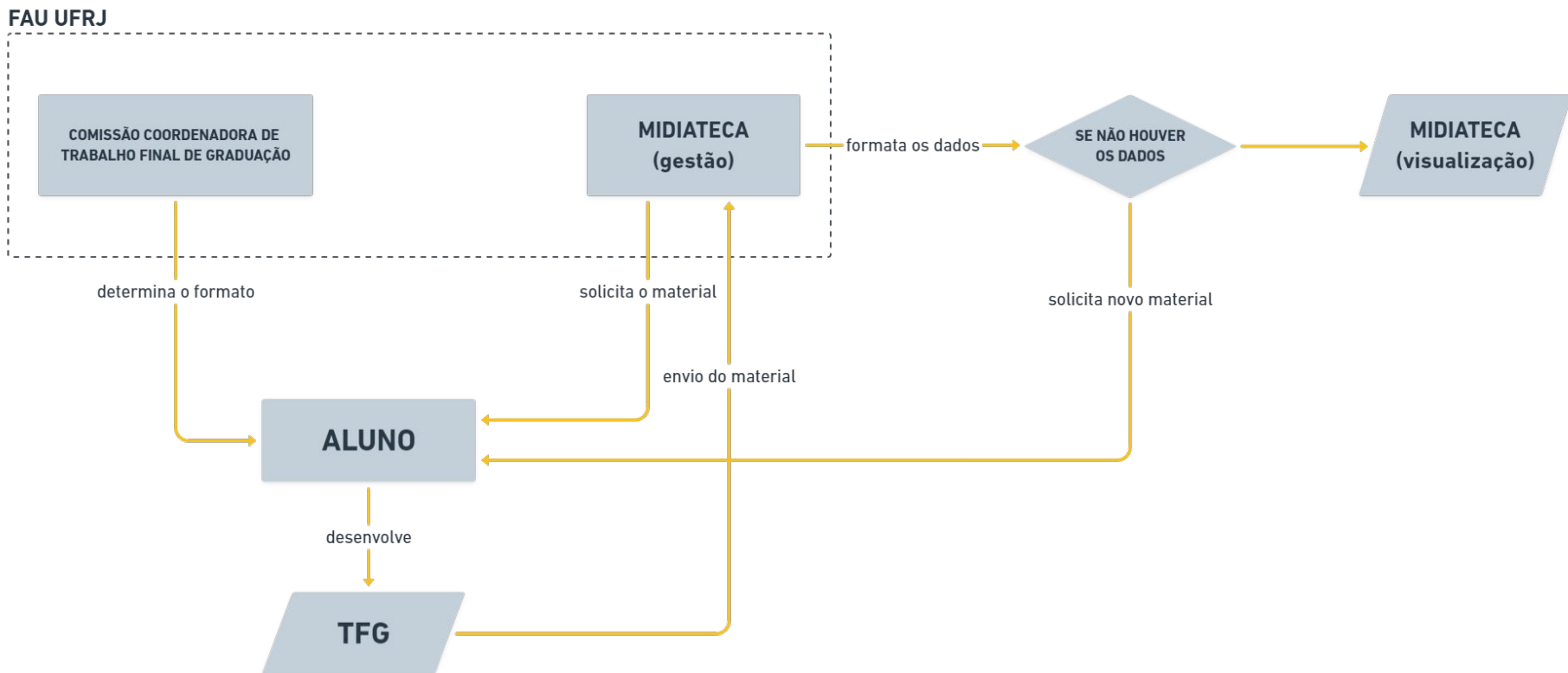
É um **repositório fragmentado** da UFRJ

Seus objetos (TFGs) podem ser entendidos como **objeto cultural**

O estudo de grandes acervos de objetos culturais (material visual) a partir de métodos computacionais chama-se **analítica cultural** (MANOVICH 2007).

POR QUE ESTUDAR A MIDIA TECA?

Fluxo de criação de um TFG



DESAFIOS

Dados heterogêneos e não padronizados

Problemas para adição de novos dados

Falta de comunicação entre CCTFG e MEDIATECA

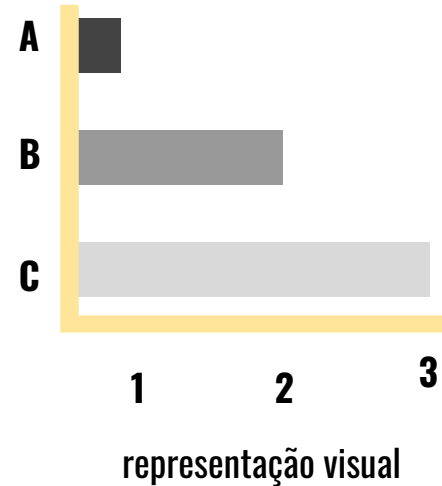
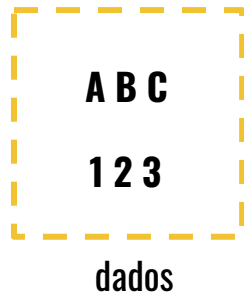
Dificuldade em contactar os alunos

Pouco material disponível sobre a MEDIATECA

Migração da base de dados (Pantheon)

A **visualização de informações** (*infovis*, abreviado) pode ser definida como um mapeamento entre dados discretos (valor específico ou categórico) e uma representação visual. (MANOVICH 2010)

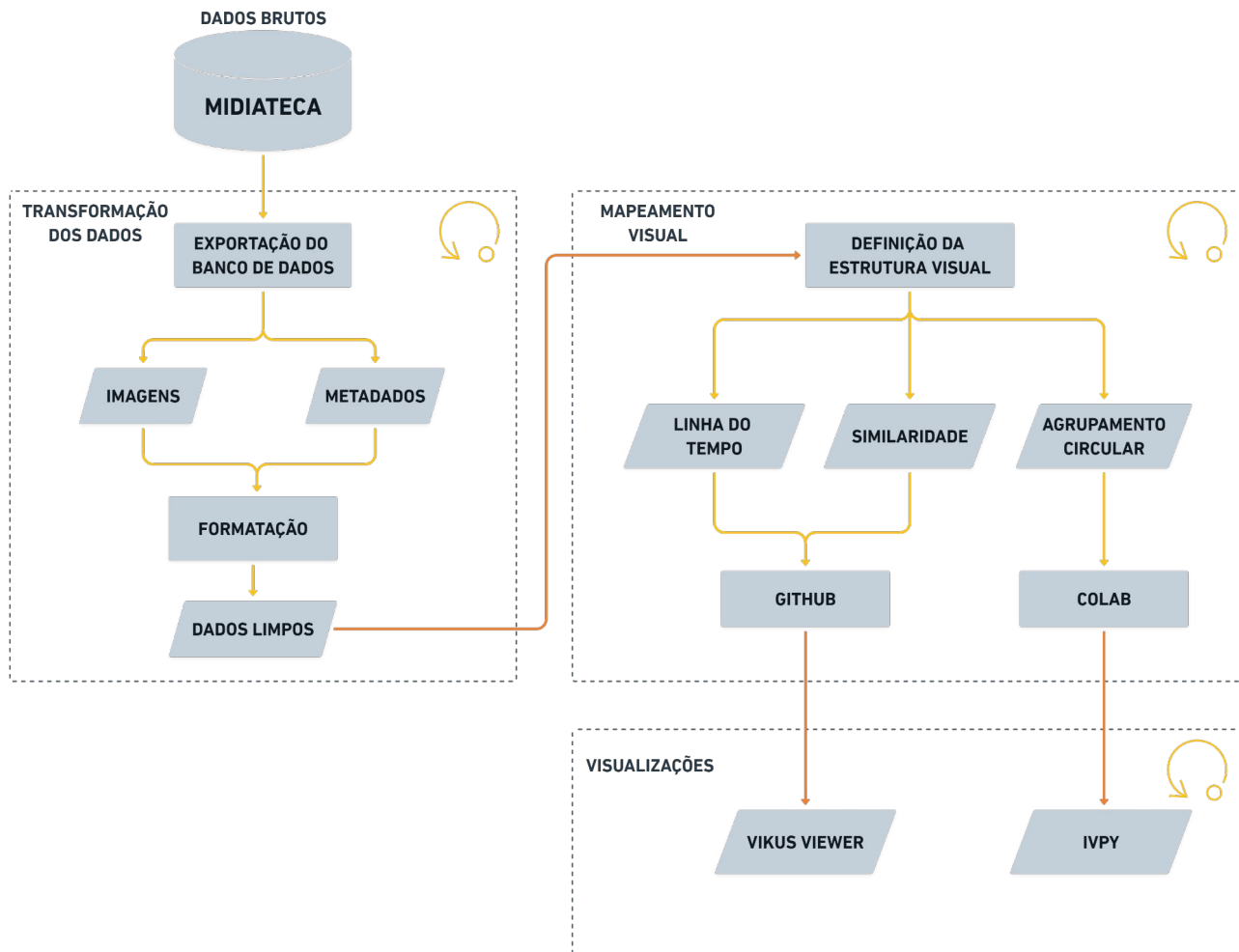
Através desse mapeamento, podemos fazer análises e interpretações das informações utilizadas.



O processo para criar uma infovis pode ser resumida em três grandes eventos:
transformação dos dados, mapeamento visual e visualização (MAZZA 2009)

Estes eventos estão estruturados em sequência linear. Cada unidade de evento pode ser repetido até que todas as questões sejam satisfeitas, resultando na infovis.

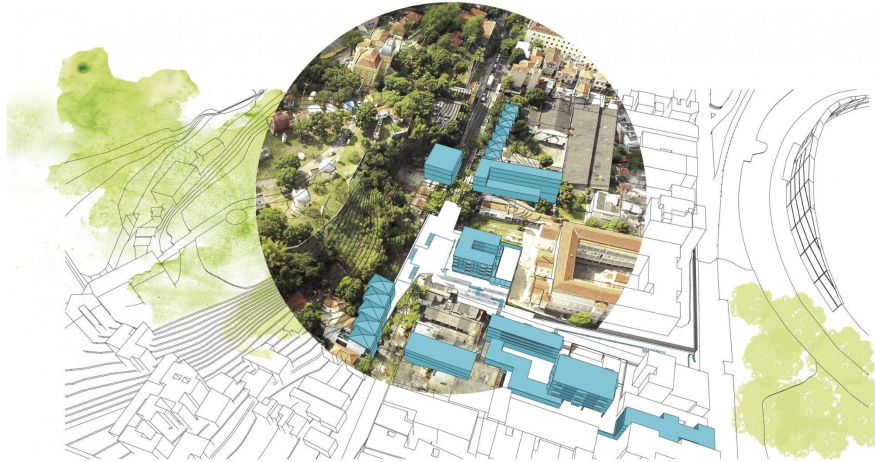
A presença humana no processo é importante, pois é o pesquisador que assume o papel na tomada de decisões.



POR QUE USAR AS IMAGENS?

A linguagem na representação também faz parte do projeto para o arquiteto - **retórica gráfica**

A imagem (síntese, técnicas, simulações de ambientes, etc) é uma das ferramentas utilizadas para expressar e materializar as ideias



*Abrindo olhos para São Cristóvão: Soluções metodológicas por uma cidade sem muros -
Nathália Domingues da Silva Meyohas, 2018*

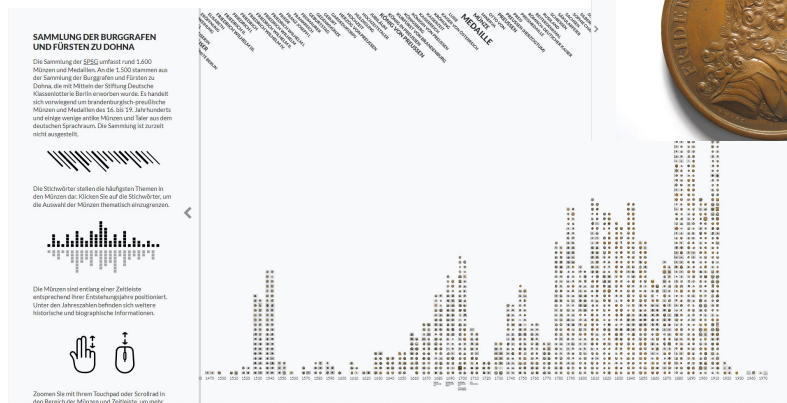
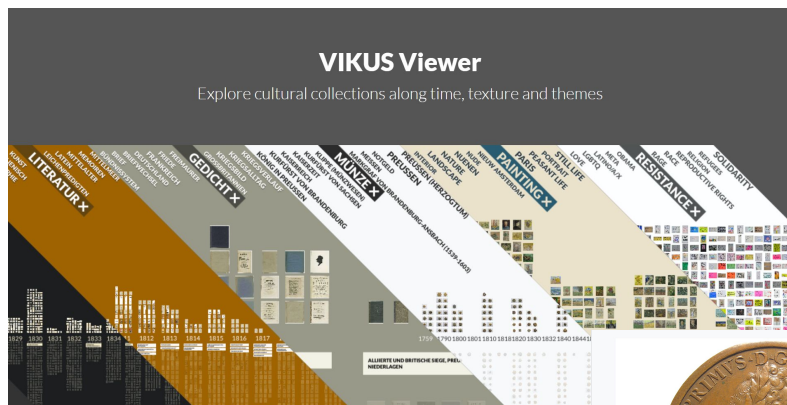


Uma nova expressão da COBAL do Humaitá - Caroline Nogueira Martins, 2013

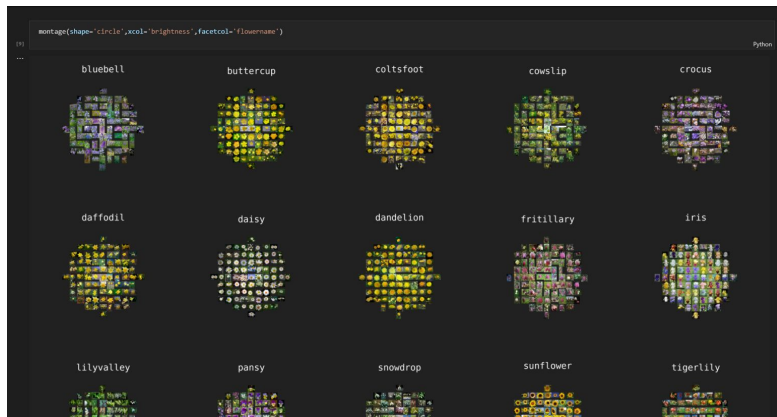
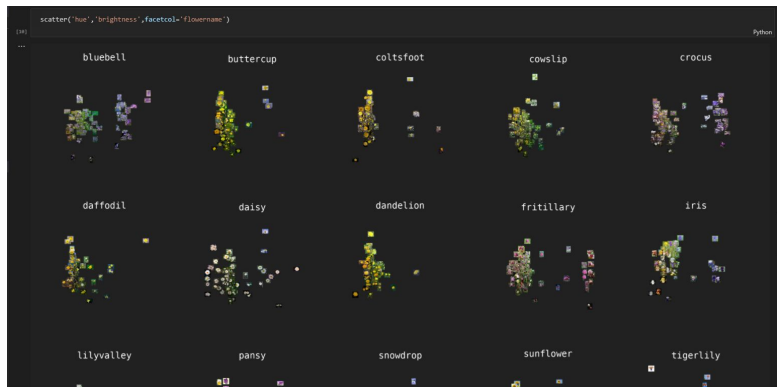


*Escola de Permacultura: Arquitetura como um organismo vivo e de memória -
Sâmara Viana Antero França, 2018*

VIKUS Viewer



- VIKUS Viewer é uma ferramenta de visualização de acervos iconográficos *open source*, organizado por tempo, textura, categorias e mapa;
- Dados estruturados em tabela .CSV;
- Filtros por categorias pré-determinadas e busca;
- Utiliza mapeamento por t-SNE (*t-Distributed Stochastic Neighbor Embedding*);
- Permite modificações de estilos de interface.



Visualizações com a biblioteca ivpy. Fonte: Autora

- ivpy é uma biblioteca de software desenvolvida em linguagem Python;
- Desenvolvimento para ambiente de notebook;
- Cria visualizações de dados a partir da análise de imagens;
- Dados estruturados em tabela .CSV;
- Permite combinar tipos de organização e informações diferentes das imagens para gerar a saída;
- Resultado apenas em imagem.

RESULTADOS

Notebook com documentação das etapas e código em python das transformações de dados (ETL) da Midiateca

Notebook com documentação das etapas e código em python das análises do acervo por matiz, brilho e saturação, em diferentes agrupamentos

Site interativo com linha do tempo e gráfico de dispersão, com filtros por temas e pesquisa por metadados

TOTAL DE IMAGENS POR ANO

	_ano	total
0	2004	7
1	2005	6
2	2006	19
3	2007	6
4	2008	35
5	2009	9
6	2011	1
7	2012	1
8	2013	785
9	2014	1205
10	2015	726
11	2016	770
12	2017	1533
13	2018	2233
14	2019	1476
15	2020	1971

TOTAL DE TRABALHOS POR PERÍODO

	_ano	_periodo	total
16	2015	1	77
17	2015	2	66
18	2016	1	77
19	2016	2	80
20	2017	1	81
21	2017	2	86
22	2018	1	68
23	2018	2	104
24	2019	1	76
25	2019	2	75
26	2020	1	36
27	2020	2	53
28	2020	4	65

MÉDIA DE IMAGENS POR ANO

	_ano	media
0	2004	3.500000
1	2005	3.000000
2	2006	2.714286
3	2007	6.000000
4	2008	4.375000
5	2009	1.500000
6	2011	1.000000
7	2012	1.000000
8	2013	5.688406
9	2014	8.992537
10	2015	5.041667
11	2016	4.873418
12	2017	9.179641
13	2018	12.982558
14	2019	9.774834
15	2020	12.798701

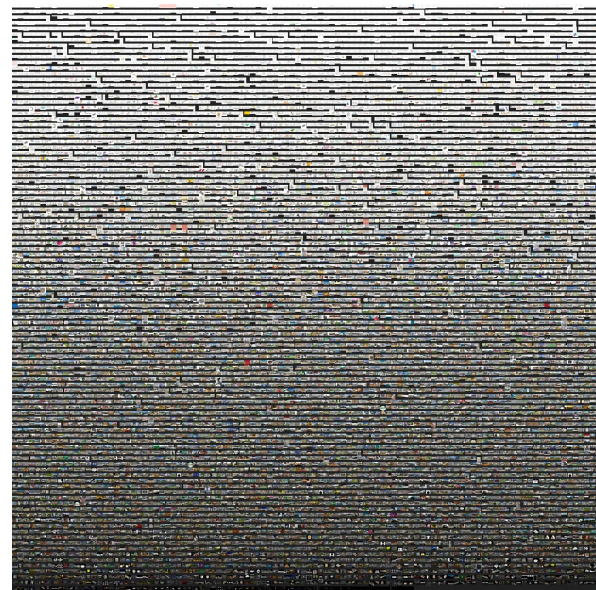
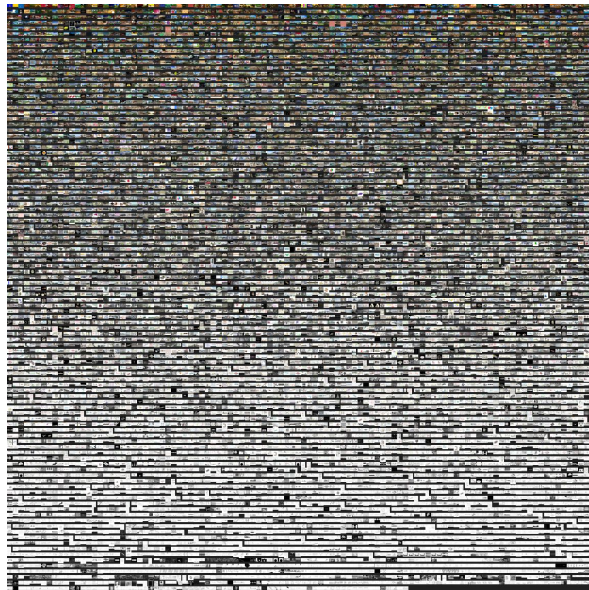
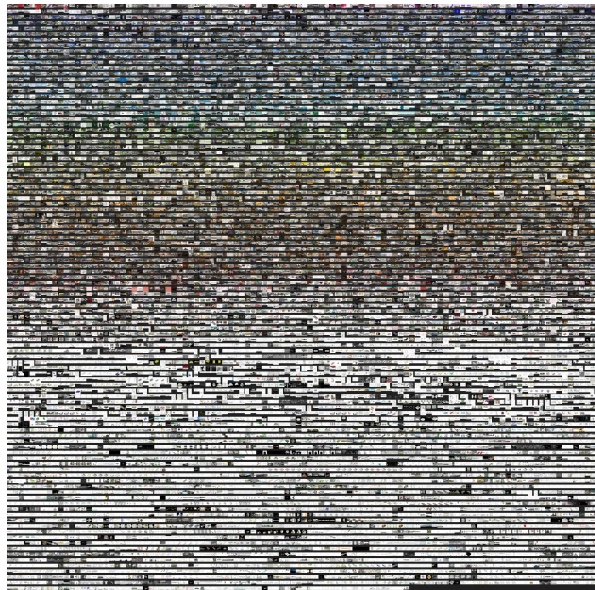
DESVIO-PADRÃO DE IMAGENS POR ANO

	_ano	desvio
0	2004	0.707107
1	2005	0.000000
2	2006	1.496026
3	2007	NaN
4	2008	2.669270
5	2009	0.836660
6	2011	NaN
7	2012	NaN
8	2013	3.207657
9	2014	9.334628
10	2015	2.101864
11	2016	2.143674
12	2017	7.053415
13	2018	7.041625
14	2019	3.770362
15	2020	7.051982

TOTAL DE IMAGENS COM COR PREDOMINANTE BRANCO OU CINZA CLARO POR ANO

	_ano	total
0	2004	2
1	2005	3
2	2006	17
3	2007	6
4	2008	22
5	2009	3
6	2011	1
7	2013	194
8	2014	746
9	2015	279
10	2016	295
11	2017	738
12	2018	1055
13	2019	818
14	2020	1215

TODAS AS IMAGENS ORGANIZADAS POR MATIZ, SATURAÇÃO E BRILHO



Visualizações com a biblioteca ivpy. Fonte: Autora

TODAS AS IMAGENS ORGANIZADAS POR MATIZ E AGRUPADAS POR ANO

2004



2005



2006



2007



2008



2009



2011



2012



2013



2014



2015



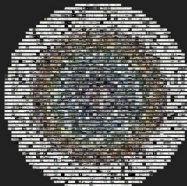
2016



2017



2018



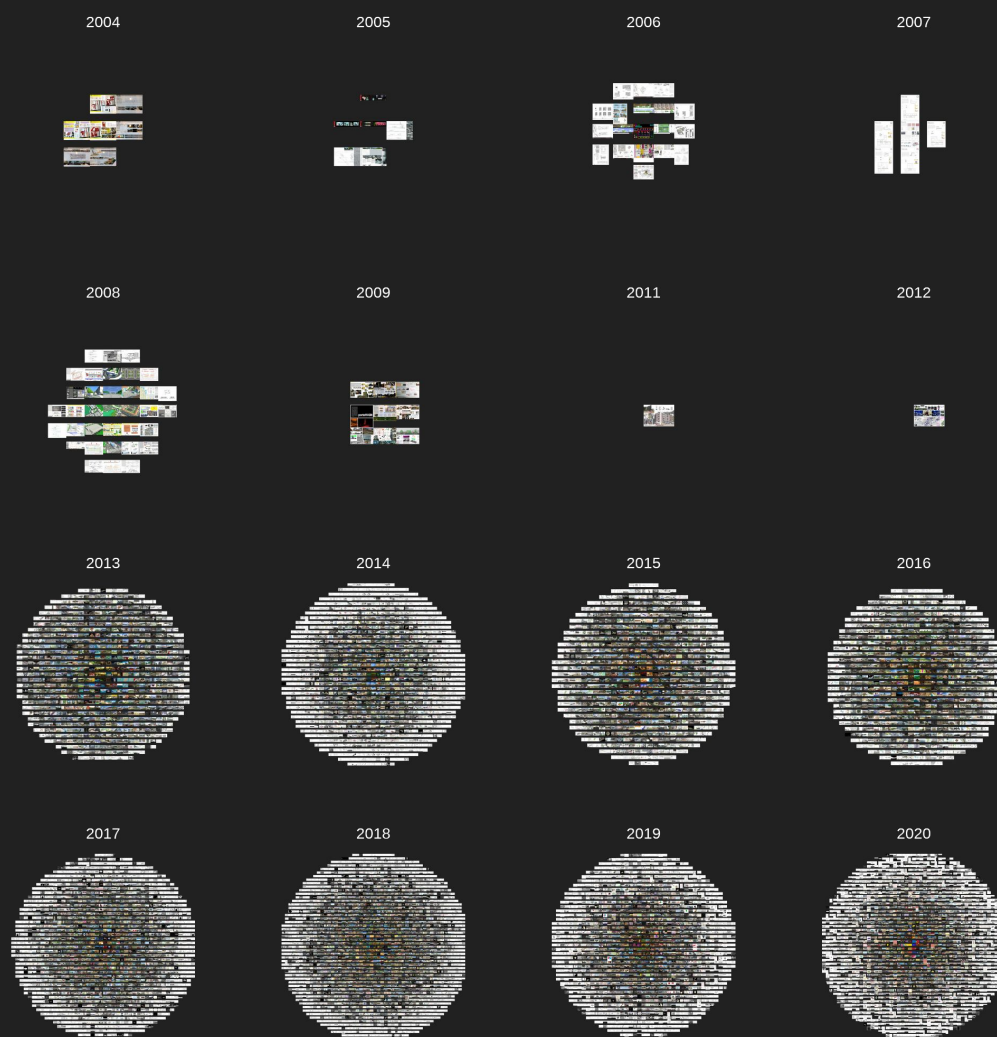
2019



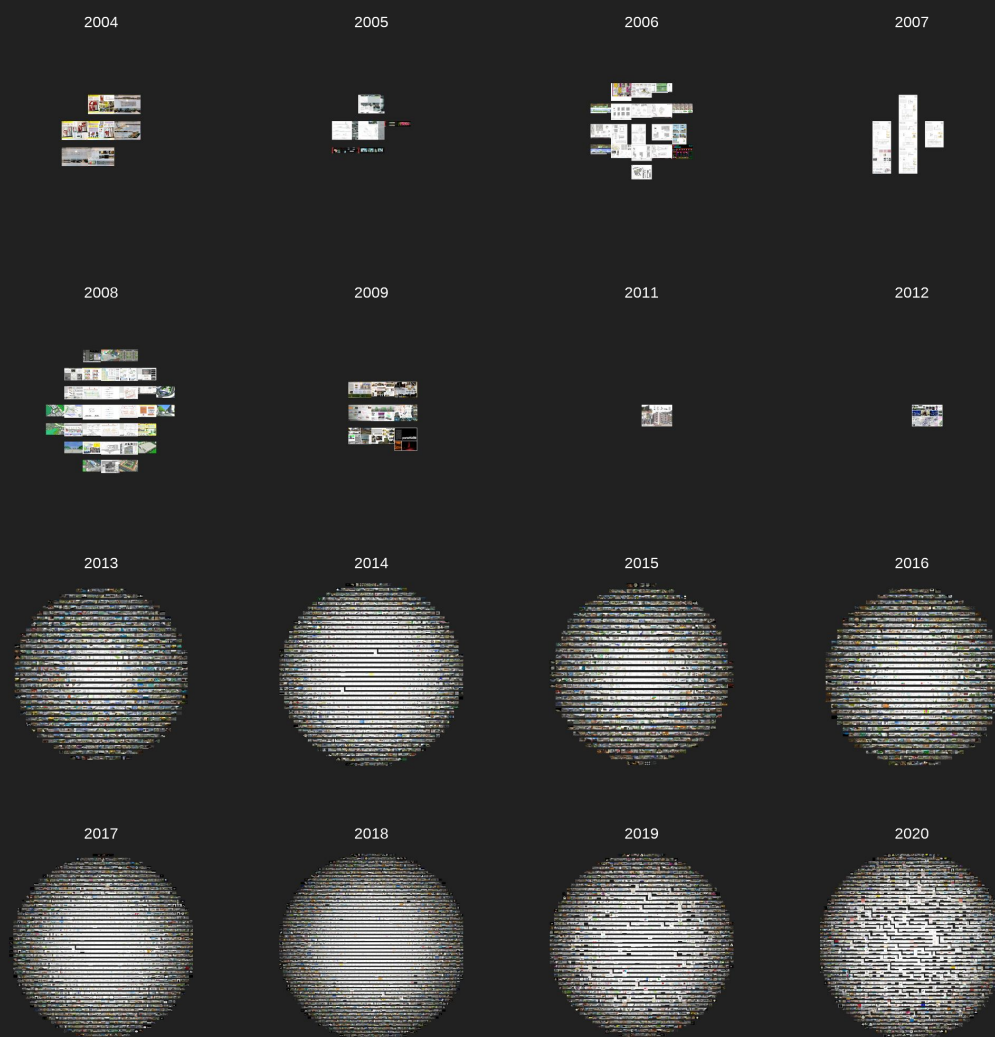
2020



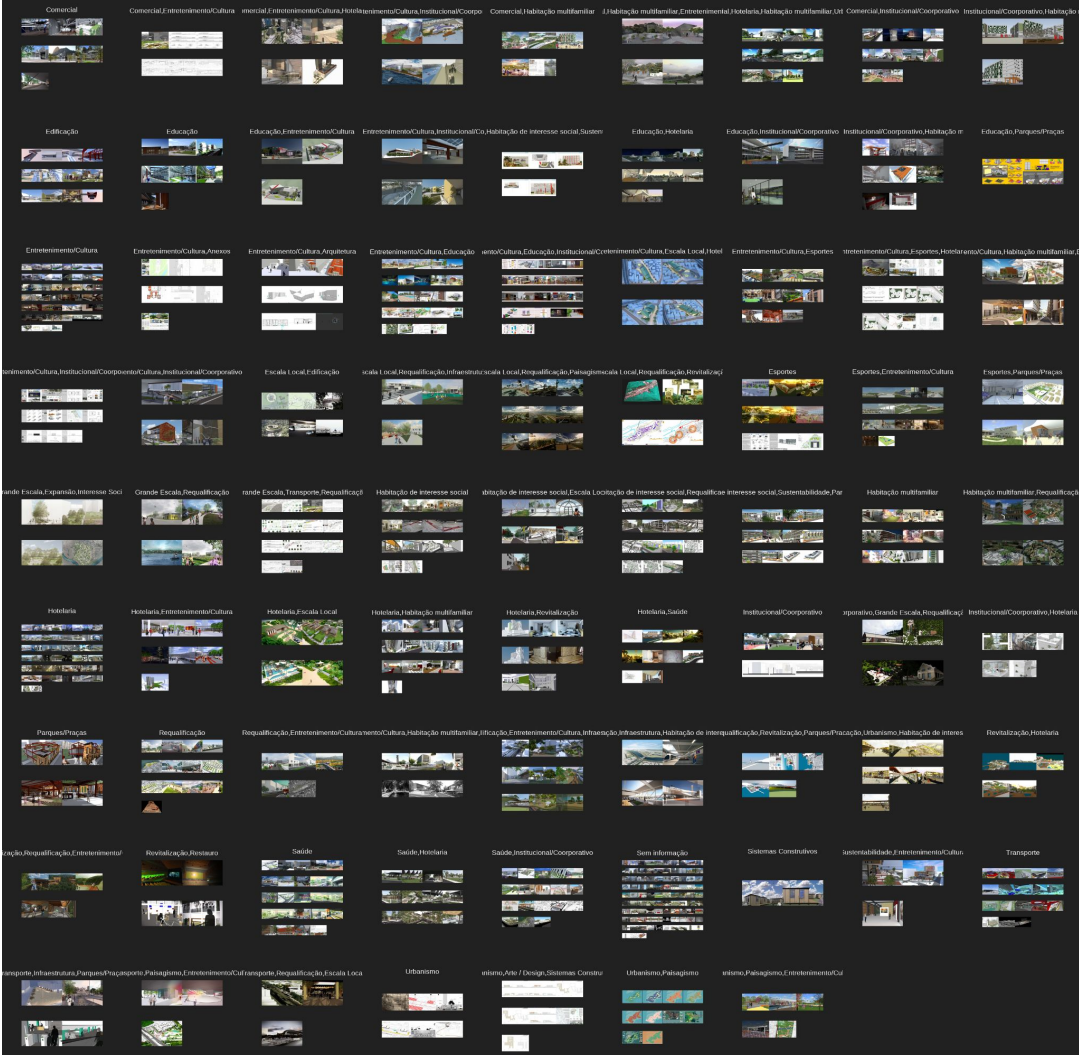
**TODAS AS IMAGENS ORGANIZADAS POR
SATURAÇÃO E AGRUPADAS POR ANO**



**TODAS AS IMAGENS ORGANIZADAS POR
BRILHO E AGRUPADAS POR ANO**



TODAS AS IMAGENS DO ANO DE 2013 ORGANIZADAS POR MATIZ E AGRUPADAS POR TEMA



A apresentação dos resultados com o VIKUS será realizada com navegação no site

É possível reproduzir a metodologia utilizada na MEDIATECA para outros conjuntos de dados, como por exemplo os projetos do Atelier Integrado 1

Algumas outras abordagens a partir de imagens:

- **extração de paleta de cores por trabalho;**
- **inteligência artificial + aprendizado de máquina para agrupar tipos de imagens (planta-baixa, corte, vista interna, etc) e/ou tipos de representação (vegetação, figura humana, rosa-dos-ventos, etc);**
- **proporção de ocupação em prancha-resumo - relação texto e imagem**

