

ARQUIGRAFIA E ARCHITEXT MINING: O APOIO À PESQUISA EM ARQUITETURA A PARTIR DE IMAGENS E TEXTOS

Artur Rozestraten
Ana Esteban Maluenda

Resumo

Este trabalho revisa duas ferramentas digitais pensadas como apoio à pesquisa em arquitetura e urbanismo apoiadas em dois formatos de informação distintos: imagens (Arquigrafia) e textos (ArchiteXt Mining). O “Arquigrafia” é um projeto iconográfico temático Web 2.0, centrado na área de arquitetura e urbanismo, que se caracteriza por uma natureza híbrida – ao congrega usuários institucionais e usuários particulares cooperando na constituição de uma mesma constelação de imagens digitais em crescimento contínuo online desde 2010 – como plataforma pública, gratuita e aberta. O “ArchiteXt Mining” (que é um acrônimo de Architectural Text Mining) é um projeto de pesquisa subvencionado pelo Governo da Espanha que propõe o uso das mais avançadas técnicas de análise de dados para a criação de uma nova ferramenta. Sua função é facilitar o trabalho de pesquisadores que se valem de revistas de arquitetura como fonte de informação. As possibilidades atuais da ciência da computação nos permitem pretender algo impossível até então: realizar uma análise geral dos conteúdos de publicações periódicas especializadas. Esta comunicação irá apresentar aspectos da trajetória e dos propósitos motivadores destes dois projetos, suas interações complementares e seus desafios atuais e futuros.

Palavras-chave: plataforma digital; imagens; textos; fotografia; revistas.

Abstract

This paper presents two digital tools designed as a support for research in architecture and urbanism and supported in two different formats of information: the images (Arquigrafia) and the texts (ArchiteXt Mining). ‘Arquigrafia’ is a Web 2.0 thematic iconographic project, focused on the area of architecture and urbanism, which is characterized by its hybrid nature - it brings together institutional and private users who cooperate in the construction of a constellation of digital images in continuous online growth since 2010- as a public platform, free and open. ‘ArchiteXt Mining’ (which is the acronym of Architectural Text Mining) is a research project funded by the Government of Spain that proposes the use of the most advanced data analysis techniques for the creation of a new tool. Its aim is to facilitate the work of researchers who use architectural journals as a source of information. The current possibilities of computer science allow us to pretend something impossible to date: perform a general analysis of the contents of specialized periodicals. This communication presents different aspects of the trajectory and purposes of motivation of both projects, their complementary interactions and their current and future challenges.

Keywords: digital platform; images; texts; photography; journals.

Resumen

Este trabajo revisa dos herramientas digitales ideadas como un apoyo a la investigación en arquitectura y urbanismo y soportadas en dos formatos de información distintos: las imágenes (Arquigrafia) y los textos (ArchiteXt Mining). ‘Arquigrafia’ es un proyecto iconográfico temático Web 2.0, centrado en el área de arquitectura y urbanismo, que se caracteriza por poseer una naturaleza híbrida –congrega usuarios institucionales y particulares que cooperan en la construcción de una constelación de imágenes digitales en crecimiento continuo online desde 2010– como plataforma pública, gratuita y abierta. ‘ArchiteXt Mining’ (que es el acrónimo de Architectural Text Mining) es un proyecto de investigación subvencionado por el Gobierno de España que propone el uso de las más avanzadas técnicas de análisis de datos para la creación de una nueva herramienta. Su función es facilitar el trabajo de los investigadores que utilizan las revistas de arquitectura como fuente de información. Las posibilidades actuales de la ciencia de la computación nos permiten pretender algo imposible hasta la fecha: realizar un análisis general de los contenidos de las publicaciones periódicas especializadas. Esta comunicación presenta distintos aspectos de la trayectoria y los propósitos de motivación de ambos proyectos, sus interacciones complementarias y sus desafíos actuales y futuros.

Palabras-Clave: plataforma digital, imágenes, textos, fotografía, revistas.

INTRODUÇÃO

Este trabalho revisa duas ferramentas digitais pensadas como apoio à pesquisa em arquitetura e urbanismo apoiadas em dois formatos de informação distintos: imagens (Arquigrafia) e textos (Architext Mining).

O Arquigrafia é um projeto iconográfico temático Web 2.0, centrado na área de Arquitetura e Urbanismo, que se caracteriza por uma natureza híbrida – que congrega usuários institucionais e usuários particulares cooperando na constituição de uma mesma constelação de imagens digitais em crescimento contínuo online desde 2010 – como plataforma pública, gratuita e aberta, cujo percurso formativo construiu conhecimento em um amplo espectro de ações conceituais e técnicas, que envolvem desde a conservação de material fotográfico original, passando pela digitalização de 42.000 imagens, incluindo um processo de constituição de um “beta perpétuo” pautado pelo Design Centrado no Usuário (DCU) até o desenvolvimento de procedimentos metodológicos e soluções de software com código aberto (+GRAFIA) e aplicativo Android que podem ser empregadas em outras áreas de conhecimento. Três referenciais históricos são fundamentais na compreensão da natureza do Arquigrafia: o Museu de Crescimento Ilimitado de Le Corbusier (1939), o Museu Imaginário de André Malraux (1947) e o esforço colaborativo de alunos, docentes e profissionais na constituição do acervo fotográfico da

Biblioteca da FAUUSP, desde os anos 1960, com a produção coletiva e a doação de originais para fins educacionais.

O Arquigrafia pode ser compreendido como um projeto com várias camadas (*layers*) integradas.

A mais visível destas camadas é o Website <www.arquigrafia.org.br> assim como também seu desdobramento como aplicativo Android e sua versão com um redesenho responsivo finalizada entre fins de agosto e início de setembro de 2018. Nesta plataforma iconográfica colaborativa convivem, interagem e complementam-se usuários institucionais e usuários particulares.

Esta camada Web mais visível se apoia e depende de uma camada mais profunda que é um Laboratório experimental, multidisciplinar, científico-tecnológico que congrega docentes-pesquisadores e estudantes de graduação e pós-graduação de diferentes institutos da Universidade de São Paulo. Como laboratório, o ARQUIGRAFIA possibilita abordagens experimentais, testes e interações formativas em distintos níveis de formação, da graduação à pós-graduação.

Em uma camada ainda mais profunda, tais interações acadêmicas se integram e estimulam as dinâmicas do Grupo de Pesquisa CNPq “Representações: Imaginário e Tecnologia” (RITE) que articula o diálogo interno e externo ao projeto, por meio de comunicações em eventos científicos, publicações e organização de encontros que estimulam a produção e a difusão de conhecimento no âmbito das Humanidades Digitais.

Desde fins de 2016, com a primeira versão do aplicativo Android e a abertura de acesso ao código fonte do sistema como um template denominado +GRAFIA, o projeto constitui uma camada acima de todas as mencionadas onde se desenvolvem projetos piloto, como atualmente o Open-Air Museum (OAM), que pretende acrescentar ao sistema informações em áudio georreferenciadas.

Parte dos desafios atuais do projeto Arquigrafia diz respeito ao esforço para situá-lo comparativamente junto a outras iniciativas iconográficas na Web 2.0, o que demanda, de início um esforço para o delineamento de critérios elementares para caracterizar tais iniciativas.

Outra parte destes desafios contempla a identificação e a análise das possibilidades técnicas de integração de fontes de textos confiáveis que venham complementar a indexação das imagens, evitando assim a produção editorial própria de textos redundantes. Além disso, o Arquigrafia pretende hoje definir um Guia de Estilos capaz de balizar tanto o redesenho e a programação responsiva atuais quanto o desenvolvimento colaborativo de futuras versões.

O “ArchiteXt Mining. Análise da arquitetura moderna espanhola através de seus textos (1939-1975)” é um projeto de pesquisa subvencionado pelo Governo da Espanha através da convocatória 2015 de “Projetos de Excelência do Ministério da Economia e Competitividade”.

Apesar do êxito do desenvolvimento de análise de dados como ferramenta em distintas disciplinas, a pesquisa em teoria

da arquitetura não fez até o momento um uso eficiente de tais tecnologias. Este é o momento de propor novas investigações aprofundadas baseadas em dados objetivos. Para tanto, se propõe a aplicação de técnicas de Text Mining (Mineração de textos) para tirar partido da melhor base de dados à disposição da disciplina: as revistas de arquitetura. O objetivo principal é oferecer uma nova perspectiva da transformação da produção arquitetônica e de como foi divulgada através de textos publicados nestes meios.

ArchiteXt Mining (que é um acrônimo de Architectural Text Mining) propõe o uso das mais avançadas técnicas de análise de dados para a criação de uma nova ferramenta. Sua função é facilitar o trabalho de pesquisadores que se valem de revistas de arquitetura como fonte de informação. As possibilidades atuais da ciência da computação nos permitem pretender algo impossível até então: realizar uma análise geral dos conteúdos de publicações periódicas especializadas. Estamos criando uma poderosa base de dados que vamos hospedar em uma plataforma web pública, acessível a toda a comunidade científica. Esta base de dados conterà não apenas campos bibliográficos clássicos, mas também muitos outros dados relevantes para os pesquisadores, como o tipo de texto, o tema de que trata, a seção em que está incluído, dados sobre o edifício, as personalidades ou eventos a que se dedica, etc. O acesso a toda esta informação é um grande passo adiante e proporciona aos pesquisadores uma poderosa ferramenta

para realizar análises quantitativas dos conteúdos dos textos e buscas que os ajudem no início de seus estudos. Entretanto, ademais, outros tipos de valores serão somados. Mediante a aplicação de diferentes técnicas estatísticas próprias do *text mining* sobre os textos completos dos artigos armazenados na base de dados, como estudos de semelhanças entre textos, correlações entre palavras ou estudos da frequência global ou linear de determinadas palavras em um texto.

Em uma primeira etapa, este projeto se dedica ao caso espanhol como uma espécie de piloto de uma investigação muito maior em escala mundial. Em particular, esta fase contemplará as revistas de arquitetura espanholas publicadas durante a ditadura franquista (1939-1975), um momento de máxima dificuldade quanto às relações internacionais. Assim poderemos comparar os temas importantes para os arquitetos espanhóis com outras matérias de interesse em revistas europeias, americanas e asiáticas que proporcionarão uma nova interpretação da arquitetura espanhola no panorama mundial.

A seguir serão apresentados aspectos gerais dos dois projetos aqui em pauta, considerando seus propósitos, suas trajetórias, suas interações complementares e seus desafios atuais e futuros.

ASPECTOS DO ARQUIGRAFIA

Idealizado em 2008 e desenvolvido desde 2010 e das por uma equipe multidisciplinar de docentes e discentes USP (FAU,

IME, ECA, FD) – com recursos da FAPESP (2009-18342-0 e 2012/24409-2), das Pró-reitorias de Pesquisa e Cultura e Extensão da Universidade de São Paulo e do INCT InterSCity – o projeto Arquigrafia segue em seu propósito de constituir uma abordagem iconográfica da arquitetura brasileira que ampare a pesquisa, a construção contínua de conhecimento, a crítica e também o projeto de arquitetura. Desde sua idealização, há 10 anos, o Arquigrafia se propunha a ser um projeto digital eminentemente visual, de cunho educacional, complementar a outras iniciativas editoriais e educacionais contemporâneas, direcionado, em primeiro lugar, aos estudantes de arquitetura e, por extensão, aos professores, pesquisadores, profissionais da área, fotógrafos e demais interessados no tema.

Em 2008, o projeto já evidenciava o desejo de constituição de uma plataforma Web para o amparo ao estudo da arquitetura brasileira a partir de imagens. A opção de foco sobre o Brasil nunca teve a intenção de ser restritiva, mas sim, de definir um ponto de partida para um processo de conhecimento que se estende, de início à comunidade lusófona e é indissociável da arquitetura mundial. Tratando-se de um projeto essencialmente visual, sua base são fotografias (aéreas, externas, internas e de detalhes), predominantemente, mas também desenhos (croquis, plantas, cortes, elevações e perspectivas), além da possibilidade já testada de integrar vídeos. Esta desejada constelação de imagens vem complementar o material visual publicado a respeito da arquitetura

brasileira, e interage diretamente com a comunidade dedicada ao estudo, à pesquisa e a iniciativas várias relacionadas às questões urbanas, patrimoniais, ambientais, paisagísticas, em um amplo espectro de interesses e formações para além da Arquitetura e do Urbanismo.

Pretendia-se a construção de uma plataforma na qual fosse possível navegar sobre conjuntos temáticos de imagens de arquiteturas de diferentes períodos históricos, e diferentes contextos geográficos do Brasil. O acervo digital de imagens seria constituído por duas vertentes integradas e complementares: o upload de imagens enviadas pelos usuários e a inserção de imagens de arquivo, provenientes de instituições como a Biblioteca da FAU, o IPHAN, o CONDEPHAAT, etc.

A fundamentação teórica do Arquigrafia encontrava então apoio nos procedimentos metodológicos de Aby Warburg (2003) e na conceituação de Alois Riegl (1959) e Henri Focillon (1939). De Warburg, apreendia-se, em especial, sua proposta de uma história visual da arte – no extremo, uma história da arte sem palavras –, mas também seu questionamento da fragmentação da história em períodos, e a revisão da noção de permanência, transmissão e transformação de conteúdos artísticos ao longo do tempo. De Riegl, retinha-se o entendimento da arquitetura como a interação artística que conforma matéria e espaço, e a noção de que o estudo das concepções espaciais deve ser feito diretamente sobre os aspectos sensíveis das arquiteturas. Já de Focillon advinha a noção central

de que, assim como o conteúdo fundamental da obra de arte é essencialmente formal, o da arquitetura é essencialmente construtivo-plástico-espacial.

A contribuição complementar que o Arquigrafia poderia oferecer a outras iniciativas já existentes apresentava-se então como:

- A ênfase na imagem que assume e reforça a importância da apreensão visual no processo de conhecimento e na comunicação de conteúdos arquitetônicos. Esse enfoque iconográfico é especialmente importante para o público leigo – que encontra dificuldades para adentrar os textos sobre arquitetura e seus jargões – e para os estudantes de graduação que começam a sistematizar conhecimentos teóricos específicos sobre um fenômeno com o qual se relacionavam intuitivamente, a partir da vivência direta e do contato sensorial;
- O texto passa a ter um papel secundário, apenas como indicativo do tema visualmente exposto. Ao invés da função convencional do texto, que antecede as imagens e lança interpretações e explicações que “preparam” o olhar do leitor, a proposta é que a apresentação das imagens, sem preâmbulos, estimule associações de ideias e interpretações originais enraizadas na “pura” percepção visual;

- O entendimento da imagem da arquitetura – gráfica ou fotográfica – como metalinguagem (distinta e indissociável da arquitetura real), sugerindo que sua construção (assim como a arquitetura real) também expressa ideias de um determinado grupo de indivíduos sobre o fenômeno arquitetônico. Cassirer (2001), Panofsky (2002) e Francastel (1987, 1990) – direta e indiretamente relacionados às ideias de Warburg – há muito apontaram o comprometimento cultural e simbólico das representações do espaço arquitetônico;
- O questionamento da periodização tradicional da história das artes e da arquitetura – moderna, colonial, neocolonial, etc. – propondo, ao invés, a reunião de imagens de diferentes períodos históricos, e diferentes procedências geográficas, sob um mesmo tema sugerindo continuidades, rupturas e ressignificações;
- A edição temática – que cria um eixo trans-histórico e trans-geográfico – e trata com o mesmo padrão visual expressões arquitetônicas do período colonial, do séc. XX e do mundo contemporâneo, advindas das várias regiões do país, e abertas a relações com expressões arquitetônicas de outras épocas e lugares;
- O enfoque poligráfico ou multigráfico como alternativa ao formato monográfico que predomina nas iniciativas editoriais. A proposta é reunir exemplos diversos – anônimos, muitas vezes – produzidos em períodos e lugares distintos, para usos diferentes, com sistemas construtivos distintos, mas alinhavados por um mesmo tema ligado ao aspecto plástico da arquitetura;
- A apresentação simultânea de imagens como prioritária ao invés da apresentação isolada figura a figura. Essa condição visual *sui generis*, de uma imagem-síntese expõe, necessariamente, tensões entre as imagens – contrastes, semelhanças e diferenças – e sugere novas percepções e interpretações das qualidades plásticas (matéria e espaço) das arquiteturas apresentadas. Os recursos digitais Web também possibilitam uma aproximação conceitualmente mais coerente com a fundamentação teórica desse projeto, ou seja, com a apresentação simultânea e não-linear de imagens, diferente do que geralmente ocorre no formato de folhas em sequência em livros ou revistas convencionais;
- A intenção de se construir um enfoque sobre a arquitetura dentro da especificidade desta arte, isto é, não simplesmente como história dos edifícios, ou história dos arquitetos, mas principalmente como história das formas de con-

cepção e composição do diálogo entre matéria e vazio, ou como história das configurações plástico-espaciais de edifícios e espaços urbanos.

Tais objetivos iniciais convergiram para esforços de interpretação individual e coletiva de imagens na plataforma Arquigrafia com base nos “conceitos fundamentais” de Wölfflin (1945) e na técnica dos diferenciais semânticos desenvolvida pelo psicólogo norte-americano Charles E. Osgood (1990).

Em síntese, Osgood (1990, p.189) propõe que o processo de juízo ou interpretação (de um fenômeno plástico-espacial, como é o caso) pode ser representado por meio da atribuição de uma posição específica sobre uma escala experimental definida entre um par de termos opostos, como aqueles definidos por Wölfflin, por exemplo. Logo, frente a uma determinada imagem fotográfica que representa um certo aspecto de uma arquitetura ou espaço urbano é possível construir interpretações/juízos estéticos sobre a arquitetura representada na fotografia com base em diferenciais semânticos. A questão que a equipe do projeto Arquigrafia formulou então foi: Como esta possibilidade poderia ser explorada na Internet, valendo-se da noção de inteligência coletiva em um ambiente colaborativo de compartilhamento temático de imagens para o qual convergem, contribuem e interagem vários usuários?

Depois de vários testes realizados entre 2010 e 2014 definiu-se um conjunto

de diferenciais semânticos propostos aos usuários como um primeiro referencial experimental para a análise e organização de relações de semelhanças formais entre imagens. São 6 pares de qualidades, binômios plástico-espaciais, sendo 3 com relação à interação entre espaços internos e externos (aberta/fechada; interna/externa; translúcida/opaca) e 3 com relação à conformação plástica de superfícies e volumes (complexa/simples; simétrica/assimétrica; horizontal/vertical).

A conjugação de tais propósitos posicionou gradualmente o projeto no âmbito das Humanidades Digitais, mais precisamente relacionando-o ao escopo desejável de projetos Web 2.0 no âmbito cultural-educacional (O'REILLY, 2009; SHELLY & FRYDENBERG, 2011; ANTOUN, 2014).

Parte do problema que move as investigações atuais do Arquigrafia atualmente é que não há um consenso absoluto sobre quais seriam esses parâmetros Web 2.0, o que caracteriza justamente o campo contemporâneo de debates, aberto a proposições conceituais e experimentações no qual entende-se que se coloca a experiência de uma década do Arquigrafia.

Muito embora existam divergências sobre o tema, alguns critérios elementares podem ser tomados como ponto de partida para a caracterização da natureza de projetos Web 2.0. São eles:

- O entendimento do ambiente Web como uma plataforma onde é possível tanto “colher” quanto “semear” informações;

- A comunidade de usuários assume o controle sobre as informações inseridas, aferindo continuamente sua confiabilidade;
- Desenvolvimento participativo/colaborativo da Arquitetura do sistema;
- Software próprio com desenvolvimento contínuo de funcionalidades específicas como “beta perpétuo”;
- Acessibilidade responsiva em dispositivos variados;
- Escalabilidade sustentável;
- Fontes de informação remixáveis com possibilidades de alterações de informações a partir das licenças conferidas pelos autores;
- Dinâmicas de aproveitamento da Inteligência Coletiva.

Cabe, de início, considerar a afirmação de que, muito embora a Web 2.0 tenha apresentado condições crescentes de participação, colaboração e intercâmbio social, o impacto de tais possibilidades nas GLAMS está bem aquém do esperado (LÓPEZ et al., 2010; GARIBALDI, 2015; FOSTER & RAFFERTY, 2016; RIDGE, 2016), prevalecendo ainda parâmetros da Web 1.0 na constituição de bancos de imagens fechados, com várias restrições de acesso às informações e arquivos, e uma produção curatorial/editorial de conteúdos iconográficos e textuais.

Considerando que hoje existem várias fontes Web que já produzem informações

textuais confiáveis, a centralização da produção de conteúdo em texto em um núcleo editorial tende a ser redundante, abre margem para erros e imprecisões, e se torna financeiramente insustentável a médio, longo prazo. Nesse cenário, os parâmetros Web 2.0 de convergência de informações fundamentadas já existentes; de aferição contínua de sua confiabilidade por meio de uma comunidade crítica ativa que modera conteúdos; e de abertura a interações entre plataformas (APIs) explorando a “hackeabilidade” e a “remixabilidade” dentro dos limites de respeito aos direitos autorais; indicam claramente um caminho a ser investigado de maneira sistemática e criteriosa em prol da sustentabilidade de projetos iconográficos que dependem da interação com textos de qualidade.

O principal problema que então se coloca é: como identificar na profusão de textos disponíveis na Web as fontes confiáveis de informação textuais?

As questões apresentadas até aqui evidenciam e reiteram a necessidade de uma atualização contínua dos ambientes colaborativos entendidos como “beta perpétuos” no contexto Web 2.0.

Ao longo desses 10 anos do projeto Arquigrafia foram produzidas 4 versões do sistema online: a primeira versão da plataforma foi concluída em 2010, a segunda em 2012, a terceira em 2015 e a quarta está prevista para meados de 2018. A experiência desse processo de produção, acrescido pelas exigências de manutenção e atualização contínua de funcionalidades, tarefas realizadas por

equipes distintas que se sucederam no tempo, evidenciou um problema metodológico-projetual: a necessidade de uma sistematização de parâmetros de integração entre WebDesign e programação conjugados em uma única base referencial estruturada como um Guia de Estilos (LYNCH & HORTON, 2016; ANDERSON, 2016).

Coloca-se então como meta atual do projeto a construção de um Guia de Estilos, conforme os parâmetros Web 2.0, que organize parâmetros capazes de orientar e coordenar as inevitáveis ações futuras de manutenção/produção características de um “beta perpétuo”, facultando ainda – na medida em que torna acessíveis e didáticas as interações entre programação e WebDesign –, a participação de programadores externos a equipe em tais atividades. Essa almejada participação de programadores voluntários na produção/manutenção do Arquigrafia só poderá ser feita, de fato, a partir da conclusão de um Guia de Estilos.

Depreende-se do que foi exposto até aqui a relevância científica e tecnológica do percurso metodológico proposto para o enfrentamento de problemas interdisciplinares e multifacetados a serem enfrentados hoje para o futuro do projeto Arquigrafia.

Tais problemas são sintetizados assim:

- A análise crítico-comparativa – com perspectiva histórica (desde 2010) e com visada sobre a condição atual – da posição relativa e da contribuição científica, tecnológica e cultural do Arquigrafia no cenário contemporâneo de plataformas iconográficas dedicadas à arquitetura e a espaços urbanos na Web 2.0 o que irá auxiliar no delineamento de diretrizes de desenvolvimento do projeto para os próximos anos;
- A reunião de parâmetros, diretrizes, linhas-guia e frameworks de WebDesign definidos como um Guia de Estilos que balize tanto o redesenho e a programação atual quanto o desenvolvimento de uma primeira versão responsiva (Website+App) do Arquigrafia (e do +GRAFIA) por ele pautada a ser concluída no final do primeiro semestre de 2019;
- Aperfeiçoamento do sistema de moderação que, valendo-se de dinâmicas de gamificação, promove um maior engajamento dos usuários na medida em que cria uma hierarquia progressiva de revisores, editores, moderadores e curadores, reconhecendo as contribuições indispensáveis das ações de tais usuários na qualificação das informações visuais e textuais disponíveis no sistema;
- Aperfeiçoamento de um vocabulário controlado específico para Arquitetura e Urbanismo, com base no Vocabulário Controlado SIBI/USP e em mapas conceituais experimentais que investiguem

estruturas de hierarquia, relação e vinculação entre termos/*tags* valendo-se de metáforas como árvores de termos e constelações de termos;

- A identificação e a análise das possibilidades técnicas de integração de fontes de dados confiáveis, tais como o IPHAN, CONDEPHAAT, DPH, Vitruvius, ArcoWeb, Getty Center, Wikipedia, World Digital Library e outras plataformas temáticas, que possam “fornecer” textos e *tags* que venham complementar a indexação das imagens, evitando assim a produção editorial própria de textos redundantes considerando o que já existe, com boa qualidade, na Internet;
- A experimentação do potencial que o piloto *Open-Air Museum* indica hoje considerando que a conjugação de dispositivos móveis e Internet sem fio torna possível acrescentar à vivência direta do espaço arquitetônico – preconizado por Zevi como protagonista da experiência da arquitetura –, a vivência simultânea de suas representações – desenhos, fotografia, filmes – *in loco*, promovendo assim uma ampliação imaginativa das interações com edifícios e cidades como realidades aumentadas por várias camadas de representações “*in loco*” ou “*in situ*”. Essa sobreposição de

camadas de informações visuais e textuais tem ainda o potencial de ir além das seleções e juízos consagrados pela historiografia, pois os recursos *wireless* possibilitam conjugar em plataformas colaborativas híbridas, juízos/opiniões/interpretações e acervos iconográficos oriundos de instituições (públicas e/ou privadas) a juízos/opiniões/interpretações e coleções iconográficas particulares construídas por leigos interessados no assunto ou por profissionais, especialistas e/ou pesquisadores especialistas. Tais sobreposições, interações, conjugações e convergências recentes, características das Humanidades Digitais, tem um potencial formativo, educacional, patrimonial, cultural ainda por explorar e é justamente nesse campo tecnológico a demandar pesquisa que se insere a investigação aqui proposta alinhada ao Programa eScience FAPESP.

Os resultados esperados dos esforços a serem realizados nos próximos meses, e que configuram a FASE III do projeto, contemplam aportes científicos e tecnológicos em duas frentes integradas e complementares entre si: contribuições metodológico-conceituais e contribuições propositivas-projetuais, ambas aplicadas diretamente no ambiente colaborativo Arquigrafia e +GRAFIA, com o piloto *Open-Air Museum* integrado.

O projeto piloto *Open-Air Museum* (Museu a Céu Aberto) promove um aporte experimental ao Arquigrafia do projeto *Smart Audio City Guide* < <https://vimeo.com/77111941> >, ambos desenvolvidos pela mesma equipe multidisciplinar de pesquisadores da Faculdade de Arquitetura e Urbanismo (FAU), do Instituto de Matemática e Estatística (IME), da Escola Politécnica (POLI), da Escola de Comunicações e Artes (ECA) e da Faculdade de Direito (FD) da Universidade de São Paulo (USP). O *Smart Audio City Guide* foi desenvolvido entre 2011 e 2013 como uma rede social e um ambiente colaborativo para *smartphones* com informações em áudio georreferenciadas, eminentemente não visuais, com o intuito de estimular interações de seus usuários com os ambientes urbanos e também entre si, considerando prioritariamente – mas não exclusivamente – o público com deficiências visuais, da baixa visão à cegueira.

O conceito de Museu a Céu Aberto, nesse piloto, considera que a interação com paisagens, espaços e objetos urbanos e a experiência sensorial direta de tais fenômenos pode ser intensificada e enriquecida com a presença *in loco* de representações – em imagens, textos e áudios – que estimulam ressignificações, reposicionamentos e revisões contínuas dos conhecimentos construídos a respeito de temas pertinentes às cidades, suas dinâmicas e transformações. Tal conceito contribui tanto à compreensão e ao reconhecimento da dimensão artística histórica que constitui a

materialidade presente nas cidades, quanto também colabora com o entendimento da transitoriedade das configurações urbanas e as decorrentes questões patrimoniais contemporâneas. Compreende-se a cidade como a grande obra de arte humana constituída por uma multiplicidade de obras de arte em diferentes escalas, dos objetos aos edifícios, dos grandes espaços públicos à sua região geográfica transformada.

Outro sentido de Museu a Céu Aberto deriva do fato de que dispositivos móveis podem trazer à luz e tornar presentes e visíveis, audíveis, sensíveis aos usuários, diretamente nos espaços urbanos, reproduções digitais de documentos que habitualmente estão restritos às reservas técnicas de Galerias, Bibliotecas, Arquivos e Museus (GLAMs). Desse modo, a associação entre processos de digitalização, plataformas Web e Internet abundante podem levar para as ruas, para as mãos e olhos do público que transita pelas cidades, documentos que ressignificam as interações com tais espaços como obras de arte. Todos esses conceitos trazem inúmeros desafios em sua implementação. Por exemplo, os sistemas de GPS podem não possuir precisão suficiente para definir a localização de um usuário entre duas obras vizinhas, o que exigirá a implementação de algoritmos que considerem outros parâmetros para reduzir tais imprecisões. Adicionalmente, por se tratar de um ambiente colaborativo, o sistema precisa ser capaz de suportar um alto fluxo de dados sem prejudicar a experiência do usuário. Pois, ao mesmo tempo que a diversidade de conteúdo permite a absorção variada de informações, ela

pode tornar o uso confuso e burocrático, o que implica a necessidade de apresentar a informação adequada ao contexto em que o usuário se encontra. Uma boa experiência de acesso ao conteúdo pode resultar em maior interesse dos usuários em engajarem-se e compartilharem suas impressões. Em uma escala mais ampla, é por meio do engajamento de usuários que se torna possível um amplo acervo de um Museu a Céu Aberto de tantas e tão distintas arquiteturas e realidades urbanas atuais. O intercâmbio de impressões distintas sobre lugares que se transformam no tempo e são percebidos de maneiras singulares por usuários diferentes entre si favorece a compreensão do relativismo de juízos formulados sobre experiências sensíveis e estimula a construção de perspectivas comparativas entre as várias possíveis narrativas relacionadas a edifícios e espaços urbanos. Para alcançar esses objetivos, o projeto está sendo desenvolvido por meio de uma pesquisa-ação que se vale de experimentos em campo para a análise de dados em sua configuração real. (ROZESTRATEN, A. et al. 2017)

Com relação às contribuições metodológico-conceituais da conclusão da Fase III do projeto Arquigrafia, espera-se a produção do necessário aprofundamento quanto aos modos de realizar processos comparativos – tão característicos da antropologia cultural, por exemplo (LAPLANTINE, 2001; LÉVY, 2002) – entre projetos Web, partindo, evidentemente, de uma revisão crítica do “Estado da Arte” quanto ao que seriam critérios fundamentais para a Web 2.0 e seus des-

dobramentos futuros. Espera-se também que essa contribuição metodológica-conceitual abarque os procedimentos de moderação promovendo uma aferição contínua da qualidade das informações em plataformas colaborativas nas suas interações online.

Espera-se, para concluir esse delineamento de horizontes e desafios, que tais investigações venham corroborar a hipótese de que o projeto ARQUIGRAFIA é hoje, de fato, uma referência internacional de inovação, tanto do ponto de vista tecnológico, considerando sua plataforma iconográfica digital responsiva aberta e funcional, quanto do ponto de vista científico, na medida de sua contribuição na construção de conhecimento em todo o espectro de ações técnicas necessárias a projetos nas mais variadas áreas de conhecimento que lidem com desafios que envolvem desde a conservação de originais até a difusão Web/App de imagens digitais devidamente georreferenciadas e indexadas.

ASPECTOS DO ARCHITEXT MINING

Em pleno século XXI é impossível pensar em pesquisa científica sem considerar aspectos digitais. O mundo em que vivemos está cada vez mais interessado em dados. Algumas ferramentas como o *data warehousing* (armazenamento de dados) ou o *data mining* (mineração de dados) têm obtido resultados muito relevantes em termos de tratamento massivo da informação. O *text mining* (mineração de textos)

é uma técnica ainda mais específica que busca padrões e tendências nos textos e permite construir conhecimento, respondendo questões previamente formuladas ou apresentando padrões ocultos em conjuntos de textos. Esta classe de processos analíticos está sendo utilizada há alguns anos em várias disciplinas científicas e humanísticas com resultados positivos.

Por exemplo, em 1946 o sacerdote italiano Roberto Busa iniciou a construção do *Index Thomisticus*,¹ uma ferramenta para facilitar buscas de textos no conjunto da obra de São Tomás de Aquino. A princípio Busa utilizava técnicas digitais muito rudimentares, como cartões perfurados, para empreender suas análises. Entretanto, graças ao espetacular desenvolvimento da metodologia de trabalho nestas disciplinas, o projeto pode obter seus maiores resultados ao final do século XX.

Mais recentemente muitos outros projetos aplicaram ferramentas de *text mining*. São bastante conhecidos trabalhos de identificação de autores em peças literárias (SALLIS, SHANMUGATHAN, 2008). Concretamente, foram amplamente divulgadas as análises das obras de William Shakespeare que colocam em questão a autoria de alguns de seus textos.

Em arquitetura também existem alguns pioneiros na análise de dados, como Juan Pablo Bonta, autor de uma extensa pesquisa publicada no livro *American Architects and Texts* de 1996 (BONTA, 1996). Em seu trabalho, Bonta compilou dados citados em 380 textos sobre arquitetura americana

desde 1815. Hoje, o trabalho árduo de Bonta quantificando as aparições de certos arquitetos nos livros foi superado por ferramentas como o *Ngram Viewer* do *Google Books*,² que permite visualizar em poucos segundos a totalidade de citações de um determinado arquiteto existentes em todos os livros considerados no projeto. Por exemplo, se forem introduzidos os nomes de vários arquitetos bem conhecidos do século XX (Le Corbusier, Frank Lloyd Wright, Mies van der Rohe, Walter Gropius e Alvar Aalto), em questão de segundos é gerado um gráfico mostrando o número de citações que tais mestres obtiveram ao longo do tempo na literatura digitalizada pelo *Google Books*. Neste gráfico, pode-se observar que as citações de todos eles têm uma tendência crescente no tempo, entre outras coisas, devido ao fato de que a indústria editorial tem crescido continuamente ao longo dos últimos anos. Percebe-se então que essa ferramenta não gera diretamente dados sobre o interesse real a respeito de certas figuras em um determinado momento, ou seja, o desenvolvimento da indústria editorial contamina a qualificação de citações e invalida esse tipo de resultados. Considerando tal distorção, como poderíamos obter informação sobre a evolução da arquitetura moderna sem acumular esse tipo de erro? Dito de outra forma: é possível utilizar bases de dados que proporcionem informação sobre o desenvolvimento da arquitetura moderna em um momento específico? A resposta é sim.

¹ Ver <http://www.corpusthomisticum.org/it/index.age>.

² Ver <http://Books.google.com/ngrams>.

A arquitetura dos séculos XX e XXI tem uma base de dados extraordinária na qual os principais conceitos, eventos e edifícios foram registrados: os periódicos de arquitetura. A maioria dos pesquisadores especializados em arquitetura moderna os utilizam como fonte. No entanto, a pesquisa sobre periódicos continua a ser feita como há 50 anos, ou seja, diretamente nos acervos físicos das bibliotecas, conferindo exemplares, página por página. Evidentemente, nas últimas décadas, foram preparados índices para o seu conteúdo, mas em geral são incompletos e não incluem registros de textos menores, como seções de notícias. Nos últimos anos também proliferaram as digitalizações disponíveis online de revistas importantes. Mas tudo isso não é suficiente. A grande quantidade de informações contidas em periódicos arquitetônicos excede a capacidade dos pesquisadores. É necessário o emprego de recursos computacionais que sejam capazes de transformar esse enorme banco de dados em um formato legível que possa ser analisado por máquinas.

ArchiteXt Mining propõe o uso de técnicas avançadas de análise de dados para construir ferramentas para os investigadores que utilizam os periódicos de arquitetura em seu trabalho. De outro lado, ArchiteXt Mining pretende ser uma ferramenta colaborativa que proporcione informação à medida que a receba de pesquisadores e usuários. Outro aspecto a se destacar é que se trata de um projeto piloto que pretende explorar novas metodologias de pesquisa.

O projeto foi concebido na Espanha, com foco específico sobre a arquitetura moderna espanhola, mas aspira converter-se em algo muito maior. Contudo, antes de tudo, suas possibilidades estão sendo testadas sobre uma amostra bastante conhecida e controlada pela equipe de trabalho: as publicações periódicas de arquitetura espanhola entre 1939 y 1975, período da ditadura franquista. As mudanças políticas e culturais transcorridas no país nesses anos proporcionam um amplo leque de possibilidades a explorar.

O projeto inicial pretendia incluir cinco revistas de arquitetura espanholas da época (*Revista Nacional de Arquitectura y Arquitectura*, *Cuadernos de Arquitectura*, *Hogar y Arquitectura*, *Nueva Forma* e *Informes de la Construcción*) e cinco revistas europeias bastante conhecidas pertencentes aos principais núcleos de difusão de notícias da época: Inglaterra (*The Architectural Review* y *Architectural Design*), França (*L'Architecture d'Aujourd'hui*) e Itália (*Domus* y *Casabella*). Entretanto, devido à drástica redução de recursos concedidos, muito aquém do solicitado, a pesquisa se restringiu às duas revistas publicadas pelos “*colegios de arquitectos*” mais importantes da Espanha, o de Madri e o de Barcelona, deixando a inclusão de outros títulos e a comparação com periódicos estrangeiros para etapas futuras.

É importante salientar, de início, que este projeto não é uma digitalização de revistas. Os projetos de digitalização são muito importantes – sobretudo se incluem OCR (*Optical Character Recognition*) – o que permite localizar termos

específicos no conteúdo das revistas. Na Espanha, recentemente, foi concluída a digitalização da revista *Arquitectura* del Colegio de Arquitectos de Madrid (<http://www.coam.org/es/fundacion/biblioteca/revista-arquitectura-100-anios>) o que possibilitou realizar buscas por autor, arquiteto ou palavra-chave incluída no título e referências bibliográficas principais. No Brasil, a Universidade de São Paulo divulgou na Web a digitalização da revista *Acrópolis* (<http://www.acropole.fau.usp.br>) tornando possível buscar termos nos textos digitalizados. Comparativamente, a digitalização da *Acrópolis* é mais útil do que a da *Arquitectura* por permitir que o pesquisador defina o conceito/termo que deseja localizar sem se restringir a uma pré-definição de termos nos campos de busca.

ArchiteXt Mining deseja proporcionar algo mais. Não apenas pretende ajudar os pesquisadores a economizar tempo, como pretende, de certo modo, converter-se em uma fonte de inspiração para os mesmos. Para tanto, o que se propõe é a criação de uma base de dados sobre arquitetura publicada nos meios especializados aberta ao mundo acadêmico que ultrapasse a informação habitualmente contida nos índices disponíveis. A ferramenta proporcionará a possibilidade de realizar análises profundas de conteúdo dos artigos aplicando as metodologias de mineração de textos. A intenção é estabelecer padrões de semelhança e diferença não apenas entre revistas, mas também entre a arquitetura espanhola e a estrangeira. A análise quantitativa

das tendências será fundamental para a localização dos principais núcleos de recepção e emissão de notícias.

Outro objetivo do projeto é proporcionar uma lista objetiva dos textos que iniciaram tendências e daqueles que, ao contrário, foram meros reflexos ou continuação dos primeiros. De outro lado, pretendemos estabelecer *rankings* que indiquem a importância de arquitetos, edifícios, críticos, além de um número considerável de variáveis de interesse para os pesquisadores. Um dos aspectos mais consistentes do projeto é sua capacidade de crescimento. Pretendemos iniciar um projeto de alcance internacional que construa uma ferramenta analítica de grande potencial.

Para tanto, a primeira grande tarefa realizada foi o desenvolvimento de uma base de dados biblio-temática. Tal base é uma classificação inicial feita de modo tradicional pelos membros da equipe de trabalho. No entanto, já proporciona uma quantidade de informação para além dos índices tradicionais. A criação dessa base de dados, por si só, já justificaria o projeto. Contudo, o ArchiteXt Mining pretende proporcionar valores adicionais valendo-se de técnicas estatísticas da mineração de textos.

Enquanto se estava construindo a base de dados em questão, a equipe preparava as digitalizações dos textos com OCR. Com a tradução dos textos para linguagem de máquina entra em cena a equipe de informática que realiza a mineração de textos. Os textos se convertem em matrizes e são

eliminadas todas as palavras que não contribuem com significados específicos. Tais *stop-words* são artigos, preposições, conjunções e todas aquelas palavras que são irrelevantes para o propósito de nossa análise. Todas elas são importantes para conectar palavras, sentenças e parágrafos, mas têm um impacto irrelevante na análise de significado dos textos. Com isso são obtidas as matrizes (TDM, *Text Matrix Document*) que constituem o autêntico DNA dos textos e a unidade de análise dos mesmos. Com essa informação podemos estabelecer buscas e comparações em grande escala, o que é impossível para o cérebro humano. A tarefa mais básica é calcular a relação de palavras mais frequentes. Trata-se de um exercício interessante em si mesmo, já que proporciona perceber os termos mais frequentes e, assim, as diversas chaves em torno do significado e sentido do texto. No entanto a relação de frequências não serve apenas para que se obtenha uma compreensão geral do conteúdo do texto. Podemos armazená-la em nossa base de dados e utilizá-la no tratamento automatizado de informação para estabelecer distintos estudos comparativos.

Estudos de frequência de palavras

Os estudos de frequência de palavras são comuns hoje para expor a informação contida em determinados conjuntos de dados aos usuários. São frequentemente utilizados para mostrar a frequência de determinados dados em uma mostra selecionada. Nos referimos às conhecidas

nuvens de palavras ou *word-clouds*. Como é bastante conhecido, quanto maior o tamanho de uma palavra em uma nuvem maior sua frequência em uma certa base de dados. É um conceito muito simples, mas que não invalida a eficácia da ferramenta na interpretação do conteúdo dos dados. Se, à frequência global de uma determinada palavra no texto adicionarmos o fator tempo, podemos desenhar um gráfico de frequência de palavras que mostra sua presença em um determinado momento do discurso. Se fizermos isso com várias palavras e estabelecermos paralelos, é fácil detectar conceitos que se apresentem nas três partes habituais de um texto: a introdução, o desenvolvimento e a conclusão. Entende-se que este passo é ainda mais eficaz que uma simples nuvem de palavras.

Estudos de correlações entre palavras

Com uma série de análises de correlação entre palavras é possível saber que palavras costumam apresentar-se associadas a outras em um texto. Tais tendências de associação dão pistas sobre dados ou circunstâncias ocultas que não tem motivo para figurar de forma explícita.

Estudos de semelhanças entre textos

Por último, propomos criar índices de semelhança entre textos. Una vez completado o processo de elaboração do DNA de todos os textos, podemos compará-los dois a dois e calcular o percentual de semelhança entre eles (SIM).

Tal percentual nos proporciona critérios objetivos para localizar séries de textos relacionados entre si, antes mesmo de iniciarmos uma leitura. Cabe dizer que não se trata de uma ciência exata, pois erros podem ser produzidos e acumulados. De todo modo, é uma técnica que auxilia a reduzir drasticamente o panorama de buscas.

Experimentamos diferentes métodos estatísticos para selecionar o que melhor se adapta às nossas necessidades. Tais métodos estão perfeitamente definidos na literatura estatística³. O primeiro é a somatória *de produtos de intervalo de frequências ponderadas*, que são termos baseados no produto escalar da frequência de palavras comuns entre dois textos ($d1$ e $d2$). Tal somatória deve ser ponderada com o número total de palavras do texto.

$$SIM(d1, d2) = \sum x_i y_i = x_1 y_1 + x_2 y_2 + x_3 y_3 + \dots + x_n y_n$$

Na fórmula anterior, são as frequências ponderadas da *i-word* () nos documentos, respectivamente. A frequência ponderada de certa palavra () em um texto é obtida mediante um quociente entre a frequência dessa palavra ($c(w_i, d1)$) e o número total de palavras do documento.

$$x_i = c(w_i, d1) / |d1|$$

$$y_i = c(w_i, d2) / |d2|$$

O segundo indicador é o *produto de somatórias de intervalos de frequências ponderadas*. Esses dois indicadores são bastante similares em termos de cálculo e utilizam as mesmas frequências

ponderadas. É muito importante ponderar os resultados. Nesse caso, as palavras frequentes em ambos os textos têm um peso considerável no cálculo.

$$SIM(d1, d2) = \sum x_i \sum y_i = (x_1 + x_2 + x_3 + \dots + x_n) \times (y_1 + y_2 + y_3 + \dots + y_m)$$

Ao contrário, o Índice *Jaccard* trabalha com a simples aparição de palavras, em atribuir ênfase a quantas vezes comparece no texto. Esse índice mostra a cardinalidade da interseção entre dois textos () dividida pela cardinalidade de sua união. No *Jaccard* não importa a frequência de certos termos, apenas sua presença.

$$SIM(d1, d2) = \frac{|d1 \cap d2|}{|d1 \cup d2|}$$

Após calcular esses três índices para um número significativo de pares de textos, nossa conclusão é de que o índice *Jaccard* é o que melhor se adapta às nossas necessidades. Em primeiro lugar porque normalmente seus valores permanecem na metade dos três índices. Em segundo lugar porque utiliza apenas os conceitos de união e interseção, o que é de fácil compreensão inclusive para principiantes nas técnicas estatísticas, como é o caso da maior parte da equipe de pesquisa. Finalmente, preferimos esse índice porque não favorece as altas frequências de palavras frente às baixas.

Os valores obtidos mediante tal método foram comparados por analistas de dados da equipe que estabeleceram um critério que conclui que quando dois textos mostram um índice de similitude em

³ Entre outros: (FELDMAN, SANGER, 2007) e (COLAS, BRAZDIL, 2006).

torno de 20% a 25% podemos considerar que tratam de temas parecidos.

Isto posto, colocam-se uma série de questões que temos tentado resolver em 2018. A primeira diz respeito às possíveis dificuldades que poderíamos encontrar ao tentar ir além das fronteiras espanholas e compara textos escritos em diferentes idiomas. Ainda que este não seja o objetivo primordial desta fase do projeto, fizemos testes de similitude entre textos escritos originalmente em espanhol e sua tradução em inglês. Mediante o uso de tradutores automáticos (Google tradutor ou similares), os resultados de similitude que obtivemos reproduzem o mesmo padrão: o índice SIM aumenta nos textos traduzidos para o inglês com relação aquele entre esses dois textos escritos em seu idioma original. O mesmo ocorre quando são utilizados os três métodos descritos anteriormente.

Nossa teoria para explicar este aumento do índice de similitude entre dois textos ao traduzi-los para outro idioma considera que os tradutores automáticos tendem a simplificar a coleção de palavras utilizadas nos textos e se valem de linguagem mais ou menos standard para expressar a mesma ideia. Provavelmente nos textos originais a presença de sinônimos, palavras polissêmicas e outros aspectos da linguagem proporcionam uma qualidade maior aos mesmos termos, enquanto que a tradução simplifica e utiliza as mesmas palavras para termos similares, o que explicaria o aumento do índice de similitude entre eles. Há que se considerar, portanto, que quando se comparam textos

traduzidos a outro idioma o índice de similitude pode sofrer um acréscimo que hoje estimamos em 10 a 15%. Este é um trabalho em andamento que ainda tem um longo caminho a percorrer.

Outra questão que nos preocupa bastante atualmente diz respeito a quais parâmetros e de que maneira queremos mostrar aos usuários da página web onde estará a ferramenta, de maneira que, em primeiro lugar, seja possível realizar buscas booleanas na base de dados bibliotemática. A ferramenta proporcionará os dados de registros bibliográficos que contêm os critérios introduzidos pelos usuários, que poderão buscar por palavra, por grupo de palavras ou por sintagmas, e que poderão ainda combinar tais parâmetros com os operadores clássicos para produzir resultados mais relevantes.

De outro lado, pretende-se que a página gere uma série de gráficos que ajudem o usuário a compreender os resultados de suas buscas, desde gráficos de barras que mostrem a palavra (ou grupo de palavras) y sua distribuição no tempo, a quantidade de artigos por ano e a aparição de certos termos em um dado momento. Se os usuários acessam um artigo em particular, além da informação bibliográfica poderá obter una nuvem de palavras de seu conteúdo ou um gráfico de pontos linear com as palavras mais frequentes.

Em nossa opinião, o resultado mais relevante que os usuários poderão obter será com o índice de similitude entre textos já que será possível obter um número determinado de artigos mais “similares” a um dado e, portanto, os

pesquisadores poderão partir de um número determinado de entradas muito mais reduzido que aquele obtido sem o uso dessa ferramenta.

CONSIDERAÇÕES FINAIS

Com base em imagens e textos, o Arquigrafia e o ArchiteXt Mining abordam a arquitetura a partir de seus documentos, de suas representações com o intuito de apresentarem contribuições complementares ao esforço dos pesquisadores da área de Arquitetura e Urbanismo. Ambos consideram que os recursos computacionais são auxiliares, mas não substituem a

imaginação, o trabalho, o critério e a sensibilidade dos pesquisadores. A condição complementar de tais plataformas pretende auxiliar, inspirar, servir de apoio ao trabalho, análise e interpretação de grandes quantidades de informação para além das possibilidades individuais. Os aspectos da trajetória, dos propósitos motivadores, das interações complementares e dos desafios atuais e futuros do Arquigrafia e do ArchiteXt Mining apresentados aqui pretendem embasar as comunicações e os diálogos a serem conduzidos no Simpósio Temático proposto e intitulado: Ferramentas e Plataformas Digitais para a documentação e a construção de conhecimento em Arte e Arquitetura.

ARTUR ROZESTRATEN é doutor em Estruturas Ambientais Urbanas e professor associado da Universidade de São Paulo | artur.rozestraten@usp.br
ANA ESTEBAN MALUENDA é doutora em Teoria e História da Arquitetura e professora da Universidad Politécnica de Madrid | ana.esteban.maluenda@upm.es

REFERÊNCIAS

- ANDERSON, Paul. **Web 2.0 and Beyond: Principles and Technologies**. Boca Raton: CRC Press, 2016.
- ANTOUN, Henrique (Org.). **Web 2.0 – participação e vigilância na era da comunicação distribuída**. Rio de Janeiro: Mauad Editora Ltda., 2014.
- BONTA, Juan Pablo. **American Architects and Texts: A Computer-Aided Analysis of the Literature**. Cambridge (Mass.): The MIT Press, 1996.
- CASSIRER, E. **A Filosofia das formas simbólicas**. São Paulo: Martins Fontes, 2001.
- CHARTIER, Roger. **El mundo como representación**. Historia cultural: entre práctica y representación. Barcelona: Gedisa, 1992.
- COLAS, Fabrice; BRAZDIL, Pavel. On the Behavior of SVM and Some Older Algorithms in Binary Text Classification Tasks. In: SOJKA, Petr; KOPEČEK, Ivan; PALA, Karel. (eds.) **Text, Speech and Dialogue** [9th International Conference, TSD 2006, Brno, Czech Republic, September 11-15, 2006. Proceedings]. Part of the Lecture Notes in Computer Science Book Series, vol 4188. Berlín, Heidelberg: Springer, 2006, 45-52.
- ESTEBAN MALUENDA, Ana; SAN PABLO MORENO, Luis. “ArchiteXt Minig: Taking advantage of Periodicals as an Architectural Data Base”. **CIRAS Discussion Paper**. Kyoto: Center for Information Resources or Area Studies, Kyoto University, N. 81, 2018, pp. 24-30.
- ESTEBAN MALUENDA, Ana; SAN PABLO MORENO, Luis; FERNÁNDEZ RODRIGUEZ, Francisco. “ArchiteXt Minig Project: Developments and Adjustments since the 2017 Symposium in Kyoto”. **CIRAS Discussion Paper**. Kyoto: Center for Information Resources or Area Studies, Kyoto University, N. 81, 2018, pp. 80-85.
- FELDMAN, Ronen; SANGER, James. **The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data**. Cambridge: Cambridge University Press, 2007.
- FOCILLON, Henri. **La vie des formes**. Paris: Alcan, 1939.
- FOSTER, Allen; RAFFERTY, Pauline. **Managing Digital Cultural Objects: Analysis, discovery and retrieval**. London: Facet Publishing, 2016.
- FRANCASTEL, P. **Imagem, visão e imaginação**. São Paulo: Martins Fontes, 1987.
- _____. **Pintura e sociedade**. São Paulo: Martins Fontes, 1990.
- GARIBALDI, Roberta. **The use of Web 2.0 tools by Italian contemporary art museums**. Museum Management and Curatorship, volume 30, Issue 3, 2015.
- LAPLANTINE, François. **L'Anthropologie**. Paris: Payot, 2001.
- LÉVY, J.J. **Entretiens avec François Laplantine – Anthropologies latérales**. Montreal: Liber, 2002.
- LÓPEZ et al. **The presence of Web 2.0 tools on museum websites: a comparative study between England, France, Spain, Italy, and the USA**. Museum Management and Curatorship, volume 25, Issue 2, 2010.
- LYNCH, Patrick J.; HORTON, Sarah. **Web Style Guide – Foundations of User Experience Design**. New Haven: Yale University Press, 2016.

- MALRAUX, André. **Le Musée Imaginaire: psychologie de l'art**. Paris: Skira, 1947.
- O'REILLY, Tim. **What is the Web 2.0 – Design Patterns and Business Models for the Next Generation of Software**. Sebastopol: O'Reilly Media Inc., 2009.
- OSGOOD, C.E. The Nature and Measurement of Meaning. In: **Language, Meaning and Culture: the selected papers of C.E. Osgood** / edited by Charles E. Osgood and Oliver C.S. Tzeng. New York: Praeger Publishers, 1990.
- PANOFSKY, E. **Significado nas artes visuais**. São Paulo: Perspectiva, 2002.
- RIDGE, Mia (Ed.). **Crowdsourcing our Cultural Heritage**. NY: Routledge, 2016.
- RIEGL, A. **Arte tardoromana**. Torino: Giulio Einaudi editore, 1959.
- ROZESTRATEN, Artur; PEREIRA, Diogo A. M. (Orgs.) **Arquigrafia entre 2009 e 2014**. São Paulo: FAUUSP, 2015.
- ROZESTRATEN, Artur; DE ANDRADE, Beatriz M.; GASTAL, Fernanda. **Manual de Procedimentos Técnicos do Projeto Arquigrafia**. 2ª Edição. São Paulo: FAUUSP, 2018.
- ROZESTRATEN, Artur S.; BERTHOLDO, Ana Paula O.; KERSUL FARIA, João Henrique; CORRÊA DA SILVA, Flávio Soares. Projeto Piloto Arquigrafia – Open-Air Museum. **Computação Brasil** - Revista da Sociedade Brasileira de Computação (SBC). Porto Alegre: SBC, n.35, ed. 3, 2017.
- SALLIS, Philip; SHANMUGANATHAN, Subana. A Blended Text Mining Method for Authorship Authentication. In: Analysis, Modelling & Simulation, 2008. AICMS 08. Second Asia International Conference. On: IEEE Explore Digital Library (<https://ieeexplore.ieee.org/document/4530518/>, search data: 05/08/2018)
- SHELLY, Garry B.; FRYDENBERG, Mark. **Web 2.0: Concepts and applications**. Shelly Cashman Series. Boston: Bentley University, 2011.
- WARBURG, Aby. **Der Bilderatlas Mnemosyne**. Berlin: Akademie Verlag, 2003.
- WÖLFFLIN, E. **Conceptos fundamentales de la Historia del Arte**. Madrid: Espasa-Calpe S.A., 1945.