

Ranking Buildings and Mining the Web for Popular Architectural Patterns

Ujwal Gadiraju
L3S Research Center
Leibniz Universität Hannover,
Germany
gadiraju@L3S.de

Stefan Dietze
L3S Research Center
Leibniz Universität Hannover,
Germany
dietze@L3S.de

Ernesto Diaz-Aviles^{*}
IBM Research
Dublin Research Lab,
Ireland
e.diaz-aviles@ie.ibm.com

ABSTRACT

Knowledge about the reception of architectural structures is crucial for architects and urban planners. Yet obtaining such information has been a challenging and costly activity. However, with the advent of the Web, a vast amount of structured and unstructured data describing architectural structures has become available publicly. This includes information about the perception and use of buildings (for instance, through social media), and structured information about the building's features and characteristics (for instance, through public Linked Data). Hence, first mining (i) the popularity of buildings from the social Web and (ii) then correlating such rankings with certain features of buildings, can provide an efficient method to identify successful architectural patterns. In this paper we propose an approach to rank buildings through the automated mining of Flickr metadata. By further correlating such rankings with building properties described in Linked Data we are able to identify popular patterns for particular building types (airports, bridges, churches, halls, and skyscrapers). Our approach combines crowdsourcing with Web mining techniques to establish influential factors, as well as ground truth to evaluate our rankings. Our extensive experimental results depict that methods tailored to specific structure types allow an accurate measurement of their public perception.

Categories and Subject Descriptors

H.3.3 [Information Search and Retrieval];
K.4 [Computer and Society]

General Terms

Human Factors, Experimentation, Measurement.

Keywords

Web Mining, Crowdsourcing, Architectural Structures, Influential Factors, Perception.

^{*}Work partially done while at L3S Research Center.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

WebSci '15 June 28 - July 01, 2015, Oxford, United Kingdom

Copyright 2015 ACM 978-1-4503-3672-7/15/06

<http://dx.doi.org/10.1145/2786451.2786467> ...\$15.00.

1. INTRODUCTION

Urban planning and architecture encompass the requirement to assess the popularity or perception of built structures (and their evolution) over time. This helps to better understand the impact of a structure, identify needs for restructuring or to draw useful conclusions about successful architectural patterns and features. Thus, information about how people think about a building that they use or see, or how they feel about it, could prove to be invaluable for architects, urban planners, designers, building operators, and policy makers alike. For example, keeping track of the evolving feelings of people towards a building and its surroundings can help to ensure adequate maintenance and trigger retrofit scenarios. On the other hand, equipped with prior knowledge of specific features that are well-perceived by the public, builders and designers can make better-informed design choices and predict the impact of building projects.

Until now, there has been limited research in the problem of ranking architectural structures based on their associated perception. So far, obtaining feedback about the perception of buildings has been a challenging and costly, yet important activity for stakeholders. Gathering such data historically required a significant amount of manual labour. With the advent of the Web, a substantial amount of data has become available publicly. This data provides information about the perception and use of buildings, for instance through social media. The social Web provides a multitude of channels, such as Twitter, Flickr, Foursquare, etc. for users to voice their opinions about situations and contexts in which they are in, often involving particular buildings. This provides a rich source for deriving information about the popularity and perception of certain structures of different types, such as airports, churches, bridges, and so forth. The Web also contains structured data about particular building features, for example, size, architectural style, and built date of certain buildings through public Linked Data. Here in particular, reference datasets such as Freebase¹ or DBpedia² offer useful data describing a wide range of architectural structures, concepts or geographically relevant regions.

The perception of an architectural structure itself has historically been studied to be a combination of the aesthetic as well as functionality aspects of the structure [28, 29]. The impact of such buildings of varying types on the built environment, as well as how these buildings are perceived, thus varies. For example, intuitively we can say that in case of

¹<http://www.freebase.com/>

²<http://www.dbpedia.org/>

churches, the appearance plays a vital role in the emotions induced amongst people. However, in case of airports or railway stations, the functionality aspects such as efficiency or the accessibility may play a more significant role. This suggests that the impact of particular *influential factors* differs significantly between different *building types*.

In this paper, we introduce a processing pipeline and experiments which mine the Social Web in order to measure (rank) popularity of architectural structure such as airports, bridges, churches, halls, and skyscrapers. We exploit the Web of data to correlate building rankings with corresponding features, in order to enable identification of statistically more popular architectural patterns.

Through this work, our main contributions are as follows.

- We present a method for ranking architectural structures that leverages Social Web sources and Linked Open Data – i.e., Flickr photos and their metadata as well as DBpedia.
- An approach to gain further insights into the perception of architectural structures, by bridging the gap between the Social and Semantic Web (correlation of structure features with facts from the Social Web).
- Influential factors and ground truth for ranking architectural structures as well as an empirical evaluation of models for generating accurate rankings.

2. PROBLEM DEFINITION

Through our work, we first aim to establish automated methods to compare and consequently rank architectural structures of varying types. Next, by correlating the rankings with structured data from DBpedia about building characteristics, we empirically demonstrate how successful architectural patterns can be automatically identified.

As a first step towards achieving this, we attempt to find answers to the question: ‘How does a building make one feel?’ We formalize these notions as follows. We define *influential factors* as the aspects that influence the perception of an architectural structure. Let $B := \{b_i; i = 1 \dots n\}$ be the set of buildings or structures, and $T := \{t_j; j = 1 \dots z\}$ be the set of building types (for example, churches, halls, skyscrapers). Given the set B of type $t \in T$, we aim to determine an optimal ranked subset, F , of *influential factors* which play a vital role in influencing building perception among people.

We thereby aim to analyze the varying influence of the factors in the set, $F := \{f_k; k = 1 \dots m\}$ on different building types in T . Let $Profile(b)$ be the building profile consisting of web data relevant to each building $b \in B$. We formulate the perception of a building b , as the normalized sum of sentiments expressed towards the building, with respect to the various influential factors.

$$Perception(b) := \frac{1}{|F|} \sum_{k=1}^m Senti^{f_k}(Profile(b)) \quad ,$$

where $Senti^{f_k}$ represents the sentiment score determined by using the influential factor f_k for the building b . Next, we present methods to automatically rank buildings of a particular type t based on the emotions that are invoked by the buildings among people, i.e., according to the perception of the buildings, $Perception(b)$. By exploiting the ranking of architectural structures thus generated, and correlating them with a set of characteristics $C := \{c_l; l = 1 \dots x\}$ of

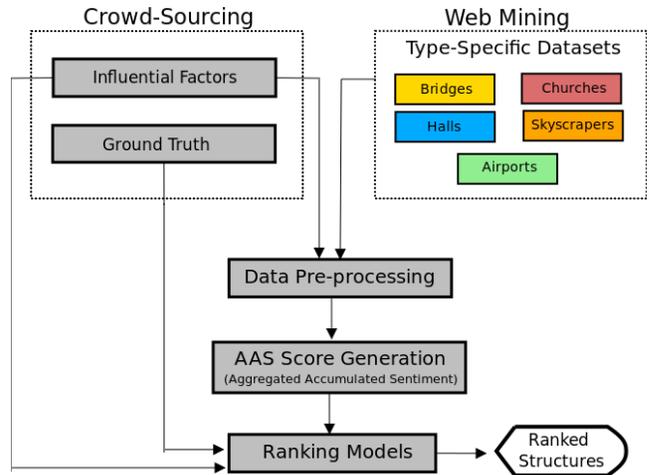


Figure 1: Pipeline of our approach for ranking architectural structures.

the buildings (extracted from DBpedia), we draw insights into architectural patterns. The characteristics in C map to DBpedia properties for the respective building b or building type t . We define an *architectural pattern* as a linear combination of mappings from building characteristics in C (eg., *architectural style*) to a particular value or value range (eg., *gothic*). We aim to identify successful architectural patterns, where ‘success’ is proportional to the positive perception of a structure.

3. APPROACH

In this section we explain our approach to rank architectural structures and mine successful architectural patterns.

3.1 Overview

We follow a threefold approach in order to rank structures based on their perception and consequently find patterns of well-perceived architectural structures. (i) First, we identify the influential factors. (ii) Next, we rank structures by crowdsourcing their popularity, in order to form the ground truth. In addition, we use automated methods for sentiment analysis and ranking. (iii) Finally, we correlate the influential factors with related structured data from DBpedia in order to identify well-perceived patterns for architectural structures. We define a well-perceived *pattern* as one that results in a high positive $Perception(b)$ value, for any structure $b \in B$ (for example, churches with a particular architectural style or skyscrapers with a height between x-y metres).

Figure 1 depicts our approach to combine crowdsourcing and Web mining methods, tailored to the type of architectural structures.

3.2 Crowdsourcing Influential Factors

Recent research works in the field of Neuroscience [19, 20], reliably suggest that neurophysiological correlates of building perception successfully reflect aspects of an architectural rule system that adjust the appropriateness of style and content. They show that people inadvertently rank buildings that they see, between the categories of either high-ranking (‘sublime’) or low-ranking (‘low’) buildings. How-

ever, what exactly makes a building likeable or prominent remains unanswered. Size could be an influential factor. At the same time, it is not sound to suggest that architects or builders should design and build only big structures. For instance, a small hall may invoke more sublime feelings while a huge kennel may not. This indicates that there are additional factors that influence building perception. In order to determine such factors, we employ crowdsourcing.

An initial survey was conducted with a primary focus on the expert community of architects, builders, and designers in order to determine influential factors. The survey administered 32 questions spanning over the background of the participants and their feelings about certain type of buildings, namely *bridges*, *churches*, *skyscrapers*, *halls* and *airports*. We consider these building types since they are the most commonly found building types across different cities, as observed from Emporis³, a real estate data mining company which is an authority on building data.

Within a two-day window, we received 42 responses from the expert community. The important influential factors that surfaced from the responses of the survey are presented below.

- The **history** associated with a building was identified to be an influential factor in terms of its affect on the people. There is a semblance of reverence towards historically significant buildings, and more often than not, they have a positive affect on people. For example, one expert who cited a *Neutral* affect from the Berliner Dom said, ‘Personally, I don’t like such baroque buildings but obviously it’s a valuable historical building and impressive’. Some other excerpts from various experts that indicate the influence of historical importance of buildings are : ‘it seems like an interesting old bridge’, ‘part of the history, beautiful historical architecture’, ‘the history of the building impresses me’.
- The imminent **surroundings** or the **built environment** of a building play a vital role in how the building itself is perceived. We observe that there is variance in what is perceived to be positive, between buildings that fit well into their surroundings and those that stand out. For example, one expert who cited a *Dislike* affect from the Fifth Avenue Presbyterian Church attributed his feelings to the reasoning, ‘it doesn’t suit the surroundings’. Some other excerpts from various experts that indicate the influence of the imminent surroundings of buildings are : ‘The big benefit is that plants are connected with a building-connection of nature and architecture’, ‘style contrast with the urban surroundings’ and so on.
- The **materials** used in the structure also influence the perception of the building.
- The **size** of a building influences its recognizability and/or visibility. This goes on to influence how the building is perceived.
- **Personal experiences** involving a building play a key role in influencing one’s feelings towards a building.
- The **level of detail**, which is an inherent part of a building’s structure is an important aspect to consider. We observe varying perceptions of intricate and complex work in the structure of a building. Some people are highly receptive of great craftsmanship, while others prefer more

³<http://www.emporis.com/>

Table 1: Trusted (TR) and Untrusted Responses (UR) from LimeService and CrowdFlower.

Building Type	# TR	# UR	# Participants
Airports	5,012	1,441	1,301
Bridges	1,357	0	76
Churches	2,085	0	79
Halls	2,880	641	1,664
Skyscrapers	7,166	370	4,276

minimalistic art work. This includes decoration and ornaments.

These influential factors correspond to the building types *bridges*, *churches*, *skyscrapers* and *halls*. However, we realize that when it comes to *airports*, people tend to acknowledge the importance of functional aspects of the buildings. By accounting for the functionality aspects that surfaced through crowdsourcing, and referring to Skytrax⁴ (a UK-based consultancy that runs an airline and airport review and ranking site), we have arrived at the following list of influential factors for airports.

- **Ease of access** to the airport (car, public transport connections, parking, etc.)
- **Efficiency** of movement/processing inside the airport (to and from gates/terminals, security, length of required paths/time from check-in to gate, etc.)
- **General design and appearance** (comfort, ambience, natural light, views)
- **Choice/availability** of shops, cafes, restaurants, etc.
- Seating/ resting/ relaxing /entertainment **facilities** in the airport
- Support for other **miscellaneous facilities** (like ATMs, disabled access, airline lounges, telephone access, wash-rooms, showers, etc.)
- **Size** of the airport

3.3 Crowdsourcing Ground Truth

We deployed surveys for each of the building types, in order to establish the ground truth in each case, using LimeService⁵ and CrowdFlower⁶. Respondents of the survey were presented with the buildings of the corresponding type (see Section 4.1 for dataset), and asked to rate them on a 5-point Likert scale from *Strongly Like* to *Strongly Dislike*, if they had been to the structure or seen it in person. Apart from this, participants were also asked to indicate the degree to which each of the influential factors determined from the initial survey (described in Section 3.2), played a role in their decision. These results were collected in another Likert scale from *Strong Influence* to *No Influence*. Table 1 shows the number of unique participants or workers that contributed to the survey(s) for the different building types.

A *response* is considered to be the set of answers pertaining to the corresponding questions of a building *b* of type *t*. The surveys for the *bridges* and *churches* were deployed on LimeService and shared through online social networks in order to trigger responses, while the surveys for the *airports*, *halls* and *skyscrapers* were deployed on the CrowdFlower platform with a monetary incentive for workers’ responses.

⁴<http://www.airlinequality.com/>

⁵<http://www.limeservice.com/>

⁶<http://www.crowdflower.com/>

Table 2: Influential factors for airports.

Influential Factor	Influence in Perception
Availability of shops, cafes, etc.	13.26%
Ease of access to the airport	14.58%
Efficiency of movement/processing	14.58%
General design & appearance	15.27%
Relaxing/Entertainment facilities	14.41%
Size of the airport	14.58%
Support for other miscellaneous facilities	14.03%

Table 1 reflects this variation in the number of participants as well as responses due to the type of platform used.

In order to maintain the integrity of drawing a comparison between results attained from these different platforms, we adjust for demographics, age and gender factors to avoid bias. In order to ensure that the workers provide valid responses devoid of any deception with ulterior motives, we intersperse the questions in the survey with test-questions that can help us detect bots or other malicious workers aiming to make quick money [6]. By doing so, we easily separate *trusted responses* from the *untrusted responses*. On the CrowdFlower platform there is a provision to create such test-questions, collectively called the *Gold Standard*. Since we utilize our personal and social networks to trigger responses for the surveys hosted on LimeService, we notice that the responses we receive are all trustworthy. This can be explained by the fact that there is no monetary incentive nor any other form of explicit incentive, implying that the workers provide responses without ulterior motives. As reflected in Table 1, we observe no untrusted responses for Bridges and Halls, the structures for which surveys were hosted on LimeService.

In addition, to prevent further bias in our crowdsourced surveys, we refrain from using images with filters or those which are edited to enhance the object in the image. We therefore use corresponding images obtained from Wikimedia Commons⁷ that are purely representational and devoid of any special touch-ups. We present images of all the architectural structures in equal resolutions, since prior work has shown that an image’s resolution and physical dimensions affect human’s aesthetic perception of it [4].

By accumulating the responses for each building b of type t and factoring in the scores from the 5-point Likert scale, we arrive at *normalized popularity-scores* for each building. The normalized popularity-scores are a reflection of the responses from the workers on the Likert scale, with respect to how they perceive the corresponding structures. We rank the buildings within each building type based on these scores, and adopt these rankings as the ground truth. Table 2 presents the various influential factors pertaining to *airports* and their corresponding influence in building perception. We observe that all the factors are almost equally influential. Interestingly, the aesthetic factor, ‘*general design and appearance*’, is marginally more influential in building perception than the other functionality related aspects.

⁷http://commons.wikimedia.org/wiki/Main_Page

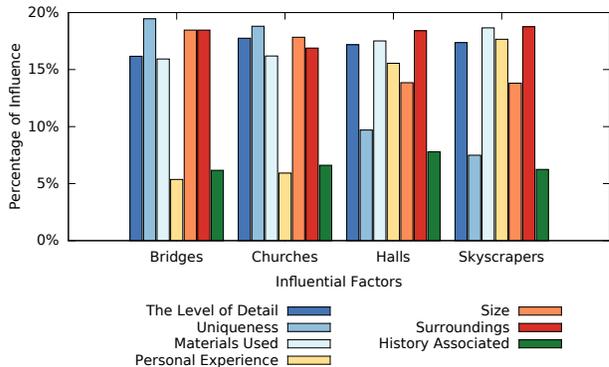


Figure 2: Comparison of influential factors.

The chart in Figure 2 shows a comparison between different building types with respect to the various influential factors. We observe that the ‘*uniqueness of a building*’ plays a significant role in case of *bridges* and *churches*, while it is less influential in the perception of *halls* and *skyscrapers*.

An observation that is common to all the building types is a significant influence of the ‘*surrounding built environment*’ in the perception of a building. This reiterates the mutual influence of a building on its built environment and vice-versa. Essentially, this means that as an extension, one can explore the correlation between a building, and other indices like ‘*well-being of a community*’ or ‘*the happiness index*’, by means of the impact a building has on its built environment.

Similarly, the influence of the *materials used* and the ‘*level of detail*’ are significant across all the building types we consider. The ‘*size*’ of a building, goes a long way in influencing its perception in case of *bridges* and *churches* as opposed to the relatively lower influence in case of *halls* and *skyscrapers*. Personal experiences of people with respect to *halls* and *skyscrapers* seem to influence their perception of the buildings significantly more than *bridges* and *churches*. Finally, the ‘*history associated*’ with a building plays a less influential role towards its perception. We believe this indicates that on average people are either not aware of the historic importance and bearing of most architectural structures, or that their understanding of the historic bearing does not affect their perception of the corresponding structures more significantly.

3.4 Building Perception : Ranking Models

In this section we present different ranking models for buildings, based on perception-related data extracted from the metadata of relevant Flickr images.

As shown in Figure 3, we propose to collect data for each of the buildings b in the set B . We create building profiles – i.e., a *Profile(b)* for each of the buildings b in the dataset – by merging the textual metadata from relevant Flickr images (title, description and comments) into a single representational unit, for each image corresponding to each building. This data will be used to generate feature vectors corresponding to each building b in the list. We will finally exploit different ranking models in order to rank the buildings.

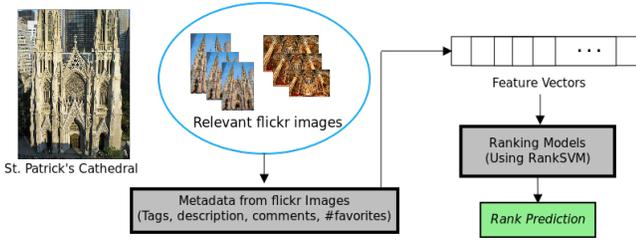


Figure 3: An example illustrating our approach towards the automated ranking of structures.

Sentic Feature Vectors

In order to extract the *emotions* from the Flickr metadata, we adopt a similar approach as the one introduced in [21] for emotion detection, which uses a term-based matching technique considering 8 basic emotions according to Plutchik’s psychoevolutionary theory of emotion [22], namely anger, anticipation, disgust, fear, joy, sadness, surprise, trust, apart from the positive and negative polarity associated with the words, to come up with a significant part of our feature vectors. To this end, we used the National Research Council Emotion Lexicon (EmoLex) [15] as prescribed in [21]. EmoLex is a large lexicon of words annotated with the associated emotions via the means of crowdsourcing [16]. Moving further, we create a profile for each building, $Profile(b)$, consisting of all the metadata from Flickr images relevant to the building. Then, we generate a sentic feature vector that represents the various dimensions of emotions contained in the profile of each building. This means that the components of the resultant vector portray each of the 8 emotions elicited by the profile for each building. These 8 components add up to 1 and each of them is a value ranging between 0 and 1. Apart from the 8 emotions, the polarity (positive and negative) features add up to 1 as well.

In addition, we assume that the normalized *number of favorites* for each building and the normalized *number of comments* for each building (accumulated from the metadata of flickr images relevant to the building) depict the interest of the people towards the building to some extent. This follows our intuition that the *favorites* indicate an approval of the buildings in the images, and can thereby be used as a significant feature to rank buildings automatically. The *number of comments* can also show the interest generated by the building in the picture. In the ranking models we employ, we follow the steps presented below.

- We compute the feature vectors for each of the buildings following the emotion detection procedure described in [21].
- We divide these feature vectors corresponding to all the buildings, into two sets (80%-20%), one for training the model and the other for testing the predictions of the learned model.
- We use RankSVM [8] to learn a model that can help to automatically rank the buildings based on their corresponding *associated emotions*, since it is a widely used in learning-to-rank tasks for Information Retrieval [14].

Automated Ranking Models

We employ different components of feature vectors, resulting in different ranking models. We adopt an intuitive and

exploratory combination of features, with an aim to produce accurate building rankings.

Frequency-based Models. The normalized *number of favorites* and the *normalized number of comments* for each building (accumulated from the metadata of Flickr images relevant to the building), independently form the basis of the *Frequency Models*. This means that according to the *Frequency Model*, each feature vector corresponding to a building consists of a single component; the normalized number of *favorites* or *comments*. This follows our intuition that the metadata from Flickr images – *favorites* and *comments* – indicate an approval of the content reflected in the images, in our case the buildings and can thereby be used as a significant feature to rank buildings automatically.

Polarity Model. Corresponding to each building, the *Polarity Model* utilizes feature vectors with two components; the positive and negative polarities.

Enhanced Sentic Model. In the *Enhanced Sentic model* the feature vectors comprise of 12 features. Apart from the 10 sentic feature dimensions – 8 basic emotions and polarity – we also include for each building the corresponding number of favorites and number of comments, both of them normalized.

Filter Model. The *Filter Model* also comprises of 12 features. It uses the influential factors we determined earlier in Section 3.3 and filters the data profiles $Profile(b)$, that we create for each building b .

As a first step, we build a *Bag of Words (BoW)* for each influential factor $f \in F$, corresponding to the building type $t \in T$. In order to do so, we use the Natural Language Toolkit (NLTK) WordNet package for Python [1].

Using the WordNet package, we can derive the related words corresponding to each influential factor through the WordNet synsets. A synset or synonym ring is defined as a set of one or more synonyms that are interchangeable in some context without changing the truth value of the proposition in which they are embedded. For example, for the Influential Factor, *Size of the building/structure*, we use the WordNet synsets to derive a *BoW* that are related to ‘size’.

We also use the *Big Huge Thesaurus*⁸ API in order to extend the *BoW*. The Big Huge Thesaurus is leveraged to extract synonyms, antonyms, related terms, similar terms and user suggestions in order to further extend the *BoW*.

In the second phase, we exploit the extended *BoW*, in order to filter the building data profiles $Profile(b_i)$, that we created for each building b_i of building type t_j . By doing so, we further prune the data by getting rid of potential noise. Figure 1 depicts this vital role played by the influential factors during the pre-processing stage.

Weighted Model. The *Weighted Model* is an extension of the *Filter Model*. Here, we consider the degrees of influence of each influential factor corresponding to the building type. First, we generate the sentic feature vectors for all buildings in the dataset, after pruning the building data profiles $Profile(b)$ corresponding to each influential factor. Then, the feature vectors are weighted with respect to their percentage of influence (depending on the building type), normalized, and combined.

⁸<http://words.bighugelabs.com/>

⁹In Table 3, $dbprop:direction \in \{\text{north, south, east, west, northeast, northwest, southeast, southwest}\}$.

Table 3: DBpedia properties that are used to materialize corresponding influential factors.

Influential Factors	Airports	Bridges	Churches	Halls	Skyscrapers
History Associated, Materials Used, Size, Level of Detail, Surroundings	dbpedia-owl: runwaySurface, dbpedia-owl: runwayLength, dbprop:cityServed, dbpedia-owl: locatedInArea, dbprop:direction ⁹	dbprop:architect, dbpedia-owl: constructionMaterial, dbprop:material, dbpedia-owl: length, dbpedia-owl: width, dbpedia-owl: mainspan	dbprop: architectureStyle, dbprop: consecrationYear, dbprop:materials, dbprop:domeHeightOuter, dbprop:length, dbprop:width, dbprop:area, dbpedia-owl:location, dbprop:district	dbpedia-owl: yearOfConstruction, dbprop:built, dbprop:architect, dbprop:area, dbprop:seatingCapacity, dbpedia-owl:location	dbprop:startDate, dbprop:completionDate, dbpedia-owl:architect, dbpedia-owl:floorCount

Table 4: Coverage of DBpedia properties representing *size* for different architectural structures in our dataset.

Airports	Bridges	Churches	Halls	Skyscrapers
runwayLength: 95%	length: 67.79%	architectureStyle: 36.69%	seatingCapacity: 65.67%	floorCount: 91%

The resulting weighted feature vectors are then used to train and test the model. In this way, the influential factors identified for each building type play a crucial role in the performance of the model itself.

As described earlier, we formulate the perception of a building as $Perception(b)$, and employ our ranking models to arrive at building rankings.

3.5 Mining the Web to Correlate Influential Factors with Relevant Structured Data

Having overcome the first hurdles of establishing the influential factors for different types of structures, and then generating rankings of structures based on their corresponding perception, the next challenge is to consolidate and correlate the influential factors with additional relevant information that can be extracted from DBpedia. Our approach to derive patterns in the perception of well-received structures is depicted in the Figure 4.

We exploit structured data from the DBpedia knowledge graph in order to correlate the influential factors with concrete features. Table 3 depicts some of the properties that are extracted from the DBpedia knowledge graph in order to correlate the influential factors corresponding to each structure with specific values. By doing so, we can analyse the well-received patterns for architectural structures at a finer level of granularity, i.e., in terms of tangible properties. In order to extract relevant data from DBpedia for each struc-

ture in our dataset, we first collect a pool of properties that correspond to each of the influential factors as per the building type as shown in Table 3. In the next step, by traversing the DBpedia knowledge graph leading to each structure in our dataset, we extract corresponding values for each of the properties identified. The properties thus extracted semi-automatically, are limited to those available on DBpedia. In addition, it is important to note that not all structures of a particular type are described through the same property set, where descriptions vary specifically with respect to their completeness. Therefore, although all the identified values accurately correspond to the structure, the coverage itself is restricted to the data available on DBpedia (see Table 4).

4. EXPERIMENTAL EVALUATION

In this section, we present our dataset for experiments, evaluate the performance of our ranking models, and discuss our results.

4.1 Dataset

As described in Section 3.3, we create building-type specific datasets and generate a new ground truth by exploiting crowdsourcing platforms like CrowdFlower and LimeSurvey. For our experiments, we consider the following architectural structure types: airports, bridges, churches, halls, and skyscrapers. The dataset we thereby created, consists of structures in the 10 biggest cities in Germany and USA.¹⁰

To ensure little variance in terms of the number of images per building, we only consider those buildings which correspond to at least a threshold number of images in our dataset. Table 5 depicts the number of images, favorites and comments corresponding to each building type.

We merge the textual metadata from the Flickr images (title, description and comments), for each image corresponding to each building, b . This constitutes the building profile, $Profile(b)$, for each building.

4.2 Performance of Ranking Models

We evaluate our ranking models in order to observe their performance. To this end, we create 10 splits in order to reasonably gauge the performance of the model from 10 rounds of learning (training) and consequent predictions (testing).

¹⁰We choose these countries due to their high social media traffic.

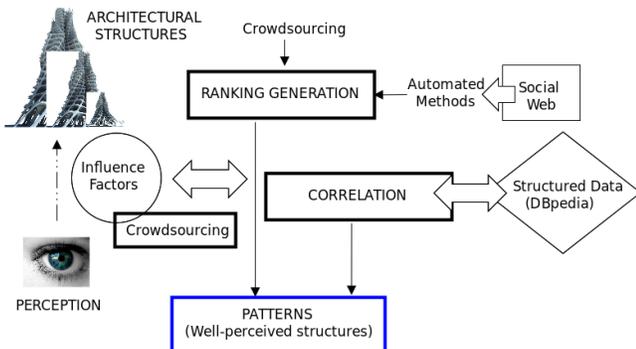


Figure 4: Approach to derive patterns for well-perceived architectural structures.

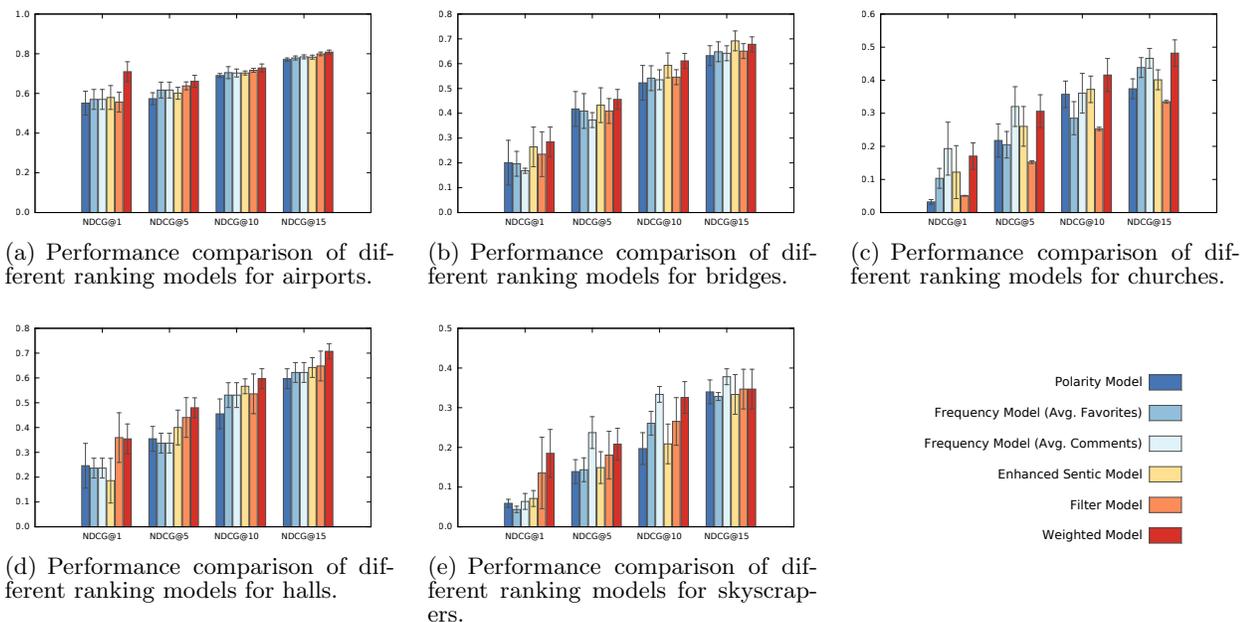


Figure 5: Performance comparison of different ranking models for different building types.

Table 6: Performance comparison of different ranking models for halls.

Avg. NDCG@	Polarity Model	FM (Avg. Favorites)	FM (Avg. Comments)	Enhanced Sentic Model	Filter Model	Weighted Model
1	0.2462	0.2366	0.2366	0.1860	0.3595	0.3544
5	0.3547	0.3372	0.3372	0.4003	0.4405	0.4799
10	0.4552	0.5308	0.5308	0.5664	0.5359	0.5971
15	0.5970	0.6219	0.6219	0.6421	0.6482	0.7073

Table 5: Type-specific dataset characteristics.

Building Type	# Buildings	# Images	# Favorites	# Comments
Airports	100	32,757	28,139	18,819
Bridges	59	12,050	19,281	25,677
Churches	139	28,683	20,857	37,036
Halls	67	20,178	11,676	14,271
Skyscrapers	178	61,538	138,899	183,051
Total	543	155,206	218,852	278,854

In order to evaluate the performance of the ranking models, we use the Normalized Discounted Cumulative Gain (NDCG) metric [14]. NDCG is a commonly used metric to judge the performance of an algorithm on training data and to compare the performance with other machine-learned ranking algorithms. Furthermore, by computing NDCG at different levels we can gain insight into the quality of the trained models.

The histograms in Figure 5 present the performance of our ranking models for the different building types. We plot NDCG values (averaged from 10 rounds of training and testing) at all levels.

We find that across the different building types, the *Weighted Model* exhibits better performance than other models for almost all NDCG levels (as illustrated in Table 6 for halls). We infer that this performance gain is due to the weighted combination of feature vectors corresponding to a building, according to the influential factors. The cases bearing exceptions are discussed further below.

In Figure 5(a), we observe high NDCG values at all levels. This can be attributed to our observation that meta-data from *airport* images on Flickr are highly rich with relevant emotion-contexts. The *Weighted Model* outperforms the other models at NDCG@1. It marginally outperforms the other models at all the other levels of NDCG measured.

The Figure 5(b) presents the performance of our ranking models for the building type, *bridges*. In case of the *bridges*, we find that the *Enhanced Sentic Model* performs better than the other models at NDCG@15. The *Weighted Model* results in a better performance than the *Frequency-based Models* and the *Polarity Model*. The chart in Figure 5(c), depicts the case of *churches*, where the *Frequency-based Model*, with average number of comments as feature, outperforms the other models at the NDCG levels 1 and 5. The *Weighted Model* performs the best at the remaining NDCG levels measured. In case of *halls*, as shown in Figure 5(d) we observe that the *Weighted Model*, followed by the *Enhanced Sentic Model* performs better than *Frequency-based models* as well as the *Polarity Model*. Figure 5(e), presents the performance of the ranking models pertaining to *skyscrapers*.

An important revelation is that simple models based on reliable features like the normalized number of *favorites* and *comments* can perform fairly well. However, we need sophisticated models like the *Weighted Model* in order to attain a higher and more stable performance across different types of structures.

We observe a clear variance in the performance of the models across the different types of architectural structures,

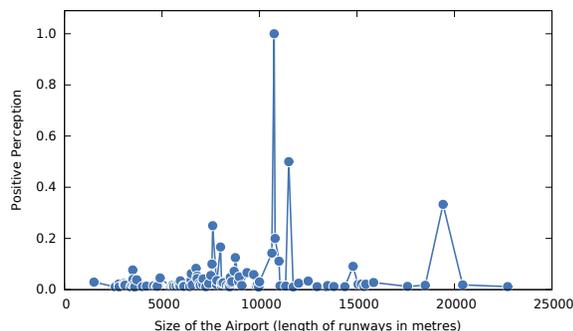


Figure 6: Influence of Size (total length of runways) in the perception of airports.

as depicted by the error bars representing the standard error. We attribute these differences to the varying importance of different emotions (which are used as features in training the models) with respect to different structure types. In addition, it is assumed that the architectural relevance of comments vary heavily among building types. For instance, while in case of churches, Flickr images and comments might likely be about the building itself, in case of bridges or airports, a large proportion of comments (and extracted sentiments) might indeed relate to other aspects. While comments and extracted sentiments might refer to aspects independent of the depicted building (for instance, the photographic quality or an event taking place at the depicted venue), additional pre-processing is required in order to better select social media of relevance for the task at hand. The crowdsourced ground truth for different architectural structures and the detailed performance of our automated ranking models are additionally published for reference.¹¹

5. CONSOLIDATION OF PATTERNS: PROOF-OF-CONCEPT

By correlating the influential factors to specific DBpedia properties, we can identify patterns for well-perceived architectural structures. In order to demonstrate how such observed patterns for architectural structures can be consolidated, we choose the influential factors, *Size* of the structure and *Level of Detail*. Although this approach can be directly extended to other influential factors and across different kinds of architectural structures, due to the limited space we restrict ourselves to showcasing these influential factors.

Airport ‘size’ is traditionally judged either by the number of operations (takeoffs and landings, runways) or the passenger traffic (number of passengers who fly in or out of the facility)¹². Characteristics of major airports include two or more long runways capable of handling the larger jet airliners. The length of the runways are a fair indicator of the size of an airport. We observe that for each airport, we can extract indicators of size using the DBpedia property `dbpedia-owl:runwayLength`. We extract the length of the runways for each airport in our dataset in order to analyse and determine the well-received pattern for airports with respect to their size. The graph in Figure 6 shows how

the popularity, i.e., the positive perception (as a factor of rank) of airports varies with their size. We observe that airports possessing runways with a length between 7,000-12,000 metres are generally well-perceived by people (higher $Perception(b)$).

Similarly, in case of bridges the influential factor ‘size’ can be represented using the DBpedia properties `dbpedia-owl:length`, `dbpedia-owl:width` and `dbpedia-owl:mainspan`, for halls we can use the DBpedia properties `dbprop:area` and `dbprop:seatingCapacity`, while we can use `dbpedia-owl:floorCount`, and `dbprop:height` to consolidate the well-perceived patterns for Skyscrapers. We thereby extract corresponding property values for each structure in our dataset using the DBpedia knowledge graph.

Figure 7(a) shows how the popularity, i.e. the positive perception of bridges varies with their size (in terms of length of the bridge). It is interesting to note that long bridges are not necessarily perceived well. We note that some bridges with length less than 1000 metres are perceived very well by people ($Perception(b_i) > 0.5$).

The plot in Figure 7(b) shows that skyscrapers having 25-65 floors form the crux of the most well-perceived skyscrapers. We observe that halls with a seating capacity between 1000-4000 people are well-perceived with the positive perception between 0.1 and 1.

For churches, we demonstrate the consolidation of patterns with respect to the influential factor *Level of Detail*. The `dbprop:architectureStyle` is a good measure of the detail in the structure. We thereby correlate the influential factor *Level of Detail* with the architecture styles using `dbprop:architectureStyle` in the DBpedia graph. By doing so, the churches in our dataset are mapped to 15 different architectural styles. The 3 most popular styles are found to be ‘Gothic Revival’, ‘Romanesque’, and ‘Gothic’.

Caveats and Limitations. In this paper, we have shown how architectural patterns can be mined by correlating structure features with properties from DBpedia. It is very important to note that the architectural patterns observed and presented here are based on merely a single dimension (i.e., size or level of detail). We have already shown that perception of an architectural structure involves multiple factors. In order to establish more concrete, meaningful and thorough architectural patterns, in our future work we will consider the remaining influential factors in a similar manner for each type of structure.

We have taken explicit precautions to ensure the reliability of using crowdsourcing to rank buildings based on their appearance and aesthetic appeal. We manifest this through different influential factors. As described earlier in Section 3.3, we take measures to counter possible biases in the responses received from crowd workers. However, given that some architectural structures may have binding cultural relevance to particular workers, and without detailed information regarding the backgrounds of various workers, we acknowledge that it may be infeasible to account for such biases to an accurate extent.

While we have also taken steps to ensure that the images representing the buildings are the central subjects and devoid of additional effects, it was beyond the scope of our work to account for biases due to other photometric features such as hue, brightness, and so forth.

¹¹<http://data-observatory.org/building-perception/>

¹²http://virtualskies.arc.nasa.gov/airport_design/3.html

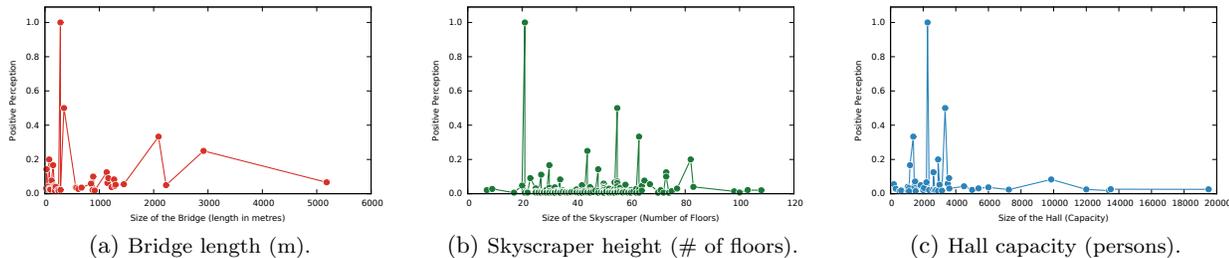


Figure 7: Influence of size in the perception of architectural structures.

6. RELATED WORK

Prior research works have established the fact that architectural structures play an important role in influencing the built environment and consequently the well-being of a community. Leyden et al. show that the design and conditions of cities are strongly associated with the happiness of residents in 10 different urban areas [13]. Lathia et al. reflect on community well-being from urban mobility patterns [12]. Bill Hillier introduced *space syntax*, a science-based, human-focused approach that investigates relationships between spatial layout and a range of social, economic and environmental phenomena [7]. These phenomena include patterns of movement, awareness and interaction; density, land use and land value; urban growth and societal differentiation; safety and crime distribution. In contrast to these works, in our paper we present a quantitative approach towards gauging the perception of architectural structures.

There has been a large amount of research concerning employing the wisdom of the crowds, to solve tasks which require a large amount of human input or computation. Such works have also spanned across various domains. The authors of [18] suggest making crowdsourcing an integral part of the workflow for Galleries, Libraries, Archives and Museums (GLAMs). Quercia et al. [23, 24] crowdsource perceptions of beauty, quiet and happiness across the city of London by using Google Street View images and make use of gamification in order to build a recognizability map of the city. Similarly in our work, we rely on experts to identify influential factors that affect building perception, and employ crowdsourcing to build our ground truth.

There has been a fair amount of research work in the domain of sentiment extraction and analysis from web data sources. For example, Kennedy et al. show that Flickr tags and other metadata can be used to enhance and improve our understanding of the world [9]. The authors of [27] study the connection between sentiment of Flickr images expressed in the corresponding metadata and their visual content. Complementing the prior works with regards to exploiting sentiments expressed by people over Social Media in varying domains, in our work we focus on leveraging these signals to improve perception-based ranking models for architectural structures and their effect on the surrounding environment.

In the context of ranking *architectural structures*; over the last decade and more, there has been a considerable amount of research done with an aim towards determining the efficiency or sustainability of a building, and comparing buildings on criteria pertaining to these features [2, 3, 26]. Roulet et al designed a multi-criteria rating methodology for buildings with the purpose of ranking or rating office buildings

and retrofit scenarios of the same building according to an extended list of parameters [25]. Similar works have focused on the Indoor Environment Quality of different buildings as a means of comparison and/or ranking buildings. In order to design appropriate evaluation and rating methodologies for buildings, we need to take into consideration a number of characteristics. In general, many parameters and criteria are considered to access the buildings by each of these methodologies. The criteria may include visual and acoustic comfort, cost and energy efficiency, impact on the environment, perceived health and so on. Energy efficiency however, is considered to be the main factor in almost all recent building rating schemes [11]. Apart from energy efficiency and the other characteristics already mentioned, buildings are also rated and compared based on their safety provisions. Most often this has to do with fire-safety measures [10, 17, 30]. These works however, only consider the functional aspects of architectural structures and fall short with respect to gauging the aesthetic elements. Finally, in this work we build upon our previous work regarding extracting architectural patterns from Web data [5].

7. CONCLUSIONS AND FUTURE WORK

One of our main contributions is the pipeline we designed that can be tailored to specific architectural structure types in order to allow the measurement of public perception of structures. Alongside this, the influential factors and the ground truths established for different types of architectural structures are key contributions of our work. An interactive visualization supports further deliberation.¹³

Through our experiments, we find that in the task of ranking structures based on their associated perception extracted from Web data, a big challenge is to ensure the relevance of the extracted text to the structure-type that we are interested in. We are led to believe that pruning relevant data to closely fit the corresponding structure types will have a positive impact on ranking performance. In this respect, filtering mechanisms which consider the most fine-grained type possible (for instance, airport instead of building), seem the most promising. This is due to the insight that different types are usually influenced by different factors, as identified through our crowdsourcing activities. To this end, influential factors can provide a means to tailoring NLP-based filtering methods.

A broad range of architectural insights can be facilitated as a result of rankings thus generated. We demonstrate this by correlating with building characteristics extracted from DBpedia. Our models and methods can help in analysing

¹³<http://bit.do/building-perception-airports/>

the evolution of the popularity of a building. Apart from architects, builders, magazines, News Channels, building corporations or other parties interested in building rankings, can greatly benefit from this approach; by eliminating a large amount of human costs, otherwise required to arrive at such rankings. In addition, our approach to crowdsource the influential factors further reduces the manual labour and need for cumbersome human intervention. In many cases, influential factors with respect to different structure types are not known apriori.

In the imminent future, one direction for investigations is the correlation of building with additional structured data, as prototypically implemented in Section 3.5. With respect to mining architectural patterns, we will extend our work to cover a rigorous analysis that can help us mine patterns with multiple facets. For example, to mine patterns like ‘skyscrapers with x size, y uniqueness, and z materials used are best perceived’.

Acknowledgments. This work is partly funded by the European Union under FP7 grant agreement 600908 (DURAARK). We thank Giorgos Katsimpras and Mohammad Alrifai for their efforts in helping us with this work.

8. REFERENCES

- [1] S. Bird. Nltk: the natural language toolkit. In *Proceedings of the COLING/ACL on Interactive presentation sessions*, pages 69–72. Association for Computational Linguistics, 2006.
- [2] P. M. Bluysen, C. Cox, N. Boschi, M. Maroni, G. Raw, C. Roulet, and F. Foradini. European project hope (health optimisation protocol for energy-efficient buildings). In *Healthy Buildings*, volume 1, pages 76–81, 2003.
- [3] W. Chow. Proposed fire safety ranking system eb-fsrs for existing high-rise nonresidential buildings in hong kong. *Journal of architectural engineering*, 8(4):116–124, 2002.
- [4] W.-T. Chu, Y.-K. Chen, and K.-T. Chen. Size does matter: how image size affects aesthetic perception? In *Proceedings of the 21st ACM international conference on Multimedia*, pages 53–62. ACM, 2013.
- [5] U. Gadiraju, R. Kawase, and S. Dietze. Extracting architectural patterns from web data. In *Proceedings of the ISWC 2014 Posters & Demonstrations Track a track within the 13th International Semantic Web Conference, ISWC 2014, Riva del Garda, Italy, October 21, 2014.*, pages 461–464, 2014.
- [6] U. Gadiraju, R. Kawase, S. Dietze, and G. Demartini. Understanding malicious behavior in crowdsourcing platforms: The case of online surveys. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI 2015, Seoul, Republic of Korea, April 18–23, 2015*, pages 1631–1640, 2015.
- [7] B. Hillier. *Space is the machine: a configurational theory of architecture*. 2007.
- [8] T. Joachims. Optimizing search engines using clickthrough data. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD ’02*, pages 133–142, New York, NY, USA, 2002. ACM.
- [9] L. Kennedy, M. Naaman, S. Ahern, R. Nair, and T. Rattenbury. How flickr helps us make sense of the world: context and content in community-contributed media collections. In *Proceedings of the 15th international conference on Multimedia*, pages 631–640. ACM, 2007.
- [10] M. Kobes, I. Helsloot, B. de Vries, and J. G. Post. Building safety and human behaviour in fire: A literature review. *Fire Safety Journal*, 45(1):1–11, 2010.
- [11] M. Kordjamshidi. *House Rating Schemes*. Springer, 2011.
- [12] N. Lathia, D. Quercia, and J. Crowcroft. The hidden image of the city: sensing community well-being from urban mobility. In *Pervasive Computing*, pages 91–98. Springer, 2012.
- [13] K. M. Leyden, A. Goldberg, and P. Michelbach. Understanding the pursuit of happiness in ten major cities. *Urban Affairs Review*, 47(6):861–888, 2011.
- [14] T.-Y. Liu. *Learning to Rank for Information Retrieval*. Springer, 2011.
- [15] S. M. Mohammad and P. D. Turney. Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 26–34. Association for Computational Linguistics, 2010.
- [16] S. M. Mohammad and P. D. Turney. Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, 2012.
- [17] E. Ng. *Designing high-density cities for social and environmental sustainability*. Earthscan, 2010.
- [18] J. Oomen and L. Aroyo. Crowdsourcing in the cultural heritage domain: opportunities and challenges. In *Proceedings of the 5th International Conference on Communities and Technologies*, pages 138–149. ACM, 2011.
- [19] I. Oppenheim, H. Mühlmann, G. Blechinger, I. W. Mothersill, P. Hilfiker, H. Jokeit, M. Kurthen, G. Krämer, and T. Grunwald. Brain electrical responses to high-and low-ranking buildings. *Clinical EEG and Neuroscience*, 40(3):157–161, 2009.
- [20] I. Oppenheim, M. Vannucci, H. Mühlmann, R. Gabriel, H. Jokeit, M. Kurthen, G. Krämer, and T. Grunwald. Hippocampal contributions to the processing of architectural ranking. *NeuroImage*, 50(2):742–752, 2010.
- [21] C. Orellana-Rodriguez, E. Diaz-Aviles, and W. Nejdl. Mining emotions in short films: user comments or crowdsourcing? In *Proceedings of the 22nd international conference on World Wide Web companion*, pages 69–70. International World Wide Web Conferences Steering Committee, 2013.
- [22] R. Plutchik. *A General Psychoevolutionary Theory of Emotion*. Academic press, New York, 1980.
- [23] D. Quercia. Urban: crowdsourcing for the good of london. In *Proceedings of the 22nd international conference on World Wide Web companion*, pages 591–592. International World Wide Web Conferences Steering Committee, 2013.
- [24] D. Quercia, J. P. Pesce, V. Almeida, and J. Crowcroft. Psychological maps 2.0: A web engagement enterprise starting in london. In *Proceedings of the 22nd international conference on World Wide Web*, pages 1065–1076. International World Wide Web Conferences Steering Committee, 2013.
- [25] C.-A. Roulet, F. Flourentzou, H. Labben, M. Santamouris, I. Koronaki, E. Dascalaki, and V. Richalet. Orme: A multicriteria rating methodology for buildings. *Building and Environment*, 37(6):579–586, 2002.
- [26] C.-A. Roulet, N. Johner, F. Flourentzous, and G. Greuter. European project hope (health optimisation protocol for energy-efficient buildings). In *Healthy Buildings*, volume 1, 2003.
- [27] S. Siersdorfer, E. Minack, F. Deng, and J. Hare. Analyzing and predicting sentiment of images on the social web. In *Proceedings of the international conference on Multimedia*, pages 715–718. ACM, 2010.
- [28] C. Sitte. *City planning according to artistic principles*. Rizzoli, 1986.
- [29] L. H. Sullivan. *The autobiography of an idea*, volume 281. Courier Dover Publications, 1956.
- [30] K. T. Tsang. Stochastic quantitative fire risk assessment on old buildings in hong kong. 2012.