Social Media and Web Sensing on Interior and Urban Design

Evangelos A. Stathopoulos
Information Technologies Institute
CERTH
Thessaloniki, Greece
estathop@iti.gr

Alexander Shvets
NLP Group
Pompeu Fabra University
Barcelona, Spain
alexander.shvets@upf.edu

Roberto Carlini
NLP Group
Pompeu Fabra University
Barcelona, Spain
roberto.carlini@upf.edu

Sotiris Diplaris
Information Technologies Institute
CERTH
Thessaloniki, Greece
diplaris@iti.gr

Stefanos Vrochidis
Information Technologies Institute
CERTH
Thessaloniki, Greece
stefanos@iti.gr

Leo Wanner
NLP Group, Pompeu Fabra University
Catalan Institute for Research and Advanced Studies
Barcelona, Spain
leo.wanner@upf.edu

Ioannis Kompatsiaris
Information Technologies Institute
CERTH
Thessaloniki, Greece
ikom@iti.gr

***Abstract*—Social media and websites provide an access to public opinions on certain aspects and therefore play an important role in getting insights on targeted audiences. Designers have been investigating how to use them for grasping social feelings and needs associated with the arrangement of spaces that surround people in everyday life to find inspiration and come up with ideas for adaptive designs. Following this, we propose a novel design-oriented tool-set that retrieves and analyses online public information from Twitter and focused content from relevant websites. We present the data collection pipeline and multilingual analysis algorithms like concept extraction and sentiment analysis on interior and urban design. Finally, we showcase an application of the proposed tool-set within two case studies.**

***Index Terms*—social media sensing, web sensing, information retrieval, textual concept extraction, textual sentiment analysis**

I. Introduction

The well-being of individuals depends on their surroundings as a matter of quality, artistic beauty and practicality. Designers and architects try to tackle these aspects by acquiring insights on people's opinions, feelings, culture and needs. By having social feedback, a creator of spaces may enhance a local community's morale when, for example, a public space is arranged accordingly or an office is designed in such a way escalating both individual performances and the workforce spirit. Such social interactions between designers and audiences used to occur physically but now, the digital era offers new opportunities.

The web and social media platforms are widely adopted by developed/developing communities and provide means of expressions with global reach. Big data are generated and consumed continuously by engaged users and enthusiasts. It is hard to monitor or classify such flows, however, in some cases it is feasible to aggregate content. Nowadays, tech giants more and more offer tools to aid in this direction such as application programming interfaces. Interested parties are able to develop software to connect with such APIs to retrieve items pertinent to specific keywords, accounts or trending topics, resulting in the formation of specialized data-sets which, however, consist of unstructured information in a textual form (often in different languages) requiring further analysis.

In this paper, we propose a design-oriented social sensing pipeline which is a seamless combination of smart data collection and natural language processing techniques that help to perform a large scale data analysis and knowledge extraction in the domain of art and architecture. The pipeline retrieves targeted textual content from online sources, such as thematic articles and user comments and posts containing information that can direct the design process. It further performs multilingual analysis of gathered collections extracting concepts and linking them to knowledge resources, detecting semantic structures and domain-specific relations, identifying emotions, sentiment topics and polarities, and assessing space imageability. The pipeline was developed within the MindSpaces project[1] that aims at creating means for art- and data-driven adaptive outdoors and indoors design.

II. Related Work

For the past years, information retrieval from web sources has been a hot research topic. There are efforts and tools on sensing the crowd's opinion like in [1] where software was developed to capture inter-activities related to public post to help in decision making. In [2] a method of grouping HTML tags and local parent headers was introduced where targeted content was retrieved from news article items. Platforms were introduced [3] capable of modeling social phenomena which develop incrementally in more than one steps and over a duration of time. Regarding social media in [4], they investigated how digital media and culture allows citizens to engage with, organize around and operate upon collective issues and be involved in co-creating the social fabric and built form of a

[1] https://mindspaces.eu/

978-1-6654-9792-3/22/$31.00 ©2022 IEEE

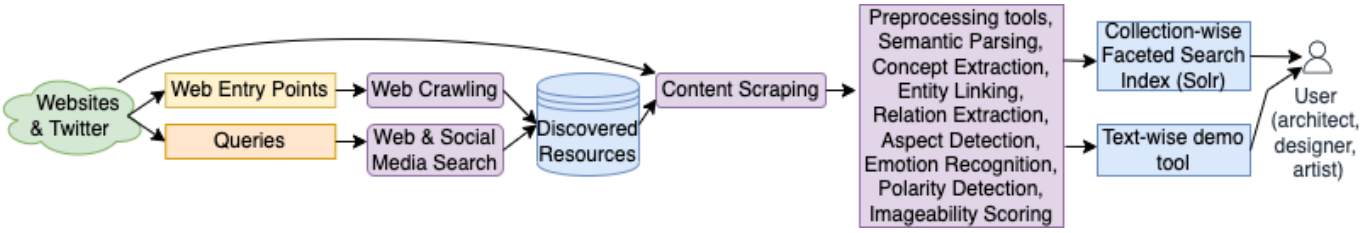

Fig. 1. The proposed MindSpaces social sensing pipeline

city. In [5], they tried to uncover inconspicuous places by using social media and street view images.

Some tools worth mentioning which follow a more commercial life-cycle than scientific are: Twazzup[2] which lists top trending content on responses, Social Mention[3] which aggregates top keywords, hashtags, users and extracts basic sentiments, HootSuite[4] and TweetReach[5] which provide keywords searching and statistics, and IceRokect and TweetDeck provide handling of Twitter user data.

III. Information Retrieval

The framework implemented for MindSpaces, consists of 2 basic elements: the Twitter crawler being a software which wraps around the freely accessible Twitter API[6] and an advanced version of easIE [6] which was developed for focused crawling on web-pages from targeted websites and for scraping multimedia content. The scope of the framework was not only to retrieve textual data from Twitter and Websites but also visual data such as images and form imagery datasets to boost the visual algorithms of the project. Social media and websites produce continuously content derived from their user base, either in the form of user posts following certain topics or by commenting personal opinions on article items. That type of textual content facilitates the artists, architects and designers to better understand public opinions about design aspects when manufacturing installations. As a result, usability questions arise when it comes to storing procedures and management of such data, making indexing and analysing data deriving from social media and websites a challenging task. Big data, variety in formats, variety in linguistics, infused noise and other issues demand for an intelligent crawler to tackle those problems and satisfy specific requirements.

A. Data Collection

In this work, 2 kinds of sources were crawled: public posts from Twitter and websites. Initially, the crawlers are able to extract public textual resources from open websites. For thematic websites, a scraping procedure follows in order to acquire only meaningful textual material concerning original posts and user comments when available. All the derived datasets consist of almost 400.000 entries and are used further in textual analysis algorithms as described in *Section V*. In conclusion, this framework may be considered as the first stage of a prepossessing pipeline which aggregates textual data.

As a starting point, the web entry points and search queries were defined. The web entry points consist URL addresses of online web domains, while the search queries include textual keywords which then are forwarded on to the Twitter API from the configuration file. There was no filtering based on hashtags as hashtags were treated as keywords. After retrieving items and before storing the information, several additional steps are present to pre-process the data. Multiple resources filtering techniques are included such as: language filtering based on English, Spanish or Catalan, removal of duplicate items based on ID, language matching among quotes, retweets and original posts and finally merging of text in case of complex quoted posts and retweet structures. As a final step, the raw data extracted are being stored in a NoSQL database which is further mapped and indexed by a Solr instance. The entire pipeline is depicted in Fig. 1.

1) Social Media Crawling: The sole social media choice to support case studies based on requirements was Twitter. The framework was designed to include additional functionalities apart from keyword searching which eventually remained unexploited after qualitative tests. Such functionalities are: retrieving tweets by using the streaming API, retrieving tweets within geocoordinate boundaries, monitoring and retrieving users' timelines, retrieving tweets by a list of IDs, retrieving tweets by an ID range, retrieving user objects by user IDs, retrieving user objects by screen names and searching for a place by either a query or an IP.

The Twitter crawler service has been online monitoring the collections of keywords in a dynamic manner for several weeks during the spring period. The scraping of textual content is executed in a round robin algorithmic approach where if the prior collection has nothing new to retrieve for via the API, it pauses temporarily and the service continues with the next collection of keywords. It generally acts in such repetitive cycles among all collections in a perpetual manner unless discontinued by the user. Another case where a temporal pause is forced, occurs when the Twitter API maximum bandwidth is reached, where the retrieval stops as long as the Twitter API developer key mandates. The twitter crawler was active for 5 weeks amassing a total of 19.402 MBs of raw twitter posts.

For collections of keywords, a survey by user partners on subjects about art and culture in urban environments was

[2]http://new.twazzup.com/
[3]http://socialmention.com/
[4]https://hootsuite.com
[5]https://tweetreach.com
[6]https://developer.twitter.com/en/docs/twitter-api/v1/tweets/search/overview

TABLE I
MAIN TOPICS AND KEYWORDS IN SPANISH AND CATALAN FOR TWITTER CRAWLING

Main topic in Catalan	**Keywords**	**Main topic in Spanish**	**Keywords**
Ordenació de la ciutat	Accessibilitat, arquitectura, ciutat, entorn urbà, espai urbà, estètica, industrial, medi ambient, mobilitat, sostenibilitat, Tecla Sala, urbanisme	Ordenación de la ciudad	Accesibilidad, arquitectura, ciudad, entorno urbano, espacio urbano, estética, industrial, medio ambiente, movilidad, sostenibilidad, Tecla Sala, urbanismo
Art	Artista, arts visuals, creació artística, obra d'art, patrimoni, emergent, art urbà	Arte	Artista, artes visuales, creación artística, obra de arte, patrimonio, emergente, arte urbano
Impacte social	Benefici social, comportament social, desenvolupament cultural, innovació social, interacció social, participació	Impacto social	Beneficio social, comportamiento social, desarrollo cultural, innovación social, interacción social, participación
Sensacions	Benestar, inspiració, sentiments, acústic, experimentació	Sensaciones	Bienestar, inspiración, sentimientos, acústico, experimentación
Llenguatges	Ciència, coneixement, creativitat, cultura, LHCultura, digital, difusió, disseny, exposició, indústria creativa, innovació, instal·lació artística, investigació, neurociència, realitat virtual, recerca, tecnologia	Lenguajes	Ciencia, conocimiento, creatividad, cultura, LHCultura, digital, difusión, diseño, exposición, industria creativa, innovación, instalación artística, investigación, neurociencia, realidad virtual, tecnología
Persones	Ciutadà, ciutadania, collectiu vulnerable, comunitat, dona, gènere, gent gran, joves, públic	Personas	Ciudadano, ciudadanía, colectivo vulnerable, comunidad, mujer, género, gente mayor, jóvenes, público
Convivència	Cohesió social, identitat, integració, multicultural, economia, pertinença, socialització	Convivencia	Cohesión social, identidad, integración, multicultural, economía, pertenencia, socialización
Aïllament de la comunitat	Diàleg, interacció, barrera lingüística, fragmentació, incomprensió, barrera, segregació, barri, collectiu, soledat	Aislamiento de la comunidad	Diálogo, interacción, barrera lingüística, fragmentación, incomprensión, barrera, segregación, barrio, colectivo, soledad
Integració de la comunitat	Immigració, conflicte social, racisme, xenofòbia, polarització, exclusió, cohesió, marginal	Integración de la comunidad	Inmigración, conflicto social, racismo, xenofobia, polarización, exclusión, cohesión, marginal
Accessibilitat de la comunitat	Gent gran, barrera arquitectònica, mobilitat, comunicació	Accesibilidad de la comunidad	Gente mayor, barrera arquitectónica, movilidad, comunicación
Inseguretat i manca de responsabilitat social	Inseguretat, robatori, soroll, baralla nocturna, il·luminació dels carrers, neteja, manteniment urbà, serveis socials	Inseguridad de la comunidad	Inseguridad, robo, ruido, pelea nocturna, iluminación de las calles, limpieza, mantenimiento urbano, servicios sociales

conducted to assemble a list in both Spanish and Catalan. The list was infused inside the configuration file of the Twitter crawling service, maintaining the grouped structure when retrieving raw Twitter posts via the API and storing it locally in a NoSQL database. The list of the collections of the keywords in the 2 selected languages is shown in Table I.

2) Web Scraping: Additionally, a website crawler performing focused crawling was developed and deployed to retrieve multimedia from targeted websites. This framework is tracking specific technical parts of a web-page by focusing on building blocks of properly defined websites. Plenty of choices were available for thematic websites, however, we chose only a few based on well-defined structures. The main focus was on scraping original posts from websites along with user comments following a dynamic procedure called deep crawling, concluding in retrieving every available public content through iterative visits in every web-page of the targeted website. By inserting the CSS selectors about the central page table, the next item and the next page inside the configuration file of the framework, deep crawling is achieved. In total, 173.02 MBs of textual data were collected. The 7 selected websites are enumerated with details in Table II:

IV. MULTILINGUAL TEXT ANALYSIS

Text analysis pipeline includes a wide range of primary pre-processing functions from sentence splitting to UD-based parsing and a set of high-level techniques listed in Fig. 1.

Concept extraction lists entities that occur in a text. Entity linking finds articles corresponding to extracted entities in cross-lingual lexical and knowledge resources such as DBpedia[7] and Babelnet[8]. Semantic parsing represents the text in a set of predicate-argument constructs for knowledge graphs. Relation extraction aims at detecting domain-specific entities with the values of their attributes. Aspect detection classifies a text with respect to a set of topics of opinions inherent in the social media texts and highlights respective aspects. Emotion recognition identifies the sentiment of a sentence within emotive classes "happy", "sad", "displeasure", and "other", while polarity detection operates over generic coarse-grained categories "positive", "negative", and "neutral". Imageability assessment detects the most memorable concepts in a text and assigns a numeric score that reflects how easy it is to recreate a place discussed in a text as a mental image.

For concept extraction, we use our model[9] proposed in [7] which is the modification of the neural biLSTM pointer-generator network. It scans the input sequence and identifies words relevant to the outcome at each state by consulting the internal pre-learned vocabulary and observing the wide context of each word at once with a copy attention mechanism that allows picking up domain-specific out-of-vocabulary concepts as-is without the necessity in re-training.

[7]https://wiki.dbpedia.org/
[8]https://babelnet.org/
[9]https://github.com/TalnUPF/ConceptExtraction

TABLE II
WEBSITES COLLECTION FOR CRAWLING WITH EASIE

Websites	Short Description
https://elfar.cat/	A digital archive of a monthly newspaper circulated in the broader area of L'Hospitalet whose topics include a wide variety of interests for the citizens of that area.
https://www.llobregatdigital.cat/	A news portal of L'Hospitalet with interviews and opinions about culture, economy, society, health and politics.
http://www.estrellalh.com/	Proximity means of communication centered in the city of L'Hospitalet de Llobregat (Barcelona). An initiative of the association Foment de la Informació Crítica (FIC-LH), with the aim of promoting knowledge and exchange of information among the citizens of L'Hospitalet.
http://localmundial.blogspot.com/	A personal blogspot from Manuel Domínguez that presents topics about history, society, politics and economy for the region of L'Hospitalet.
https://www.dezeen.com/	Dezeen is the world's most popular and influential architecture and design magazine, and the winner of numerous awards for journalism and publishing.
https://www.danieldavis.com/	A researcher's personal point of views with interests in two main themes: how technology influences architecture, and how architecture influences people.
https://www.archdaily.com/	A platform to collect and spread the most important information for architects seeking to build a better world.

The purpose of the semantic parsing component is to transform a text into deep predicate-argument structures where all the language-specific and syntax-oriented information is removed so that the final representation of the text becomes language-independent. The outcome is obtained through the application of graph-transduction grammars to the surface-syntactic structures.

For relation extraction, we trained a BERT model [8] on a generic question-answering data-set[10] to tackle open-domain attributes. To detect attributes specific to this paper, we prepended to each sentence one of the questions "Which shape is it?", "Which material is it?", and "Which colour is it?" and used the resulting sequence as the input to the model. Once the attribute is found, the second attribute of the relation is taken by a set of rules over syntactic dependencies.

For aspect detection, we selected an unsupervised neural attention model [9] that aims at discovering interpretable aspects with diverse and coherent descriptors. The model assumes coherence by exploiting the distribution of word co-occurrences through the use of neural word embeddings. In addition, an attention layer is used to de-emphasize irrelevant words. We identified and fixed the set of topics equal across the languages in collected data to be able to ground the new coming texts uniformly. We maintain a separate model for each language to account cultural differences.

Emotion recognition is based on a biLSTM model that was trained to learn the word representations from a large unlabeled emotive corpora and further fine-tuned on labeled dialogues from the SemEval-2019 EmoContext dataset[11]. We adjusted the model for individual short text classification by leveraging only one of the biLSTM units for the input text where emotional statements appeared during the training.

Regarding polarity detection, we used open-source implementations for Spanish and Catalan. The model for Spanish was obtained by pre-training deep bidirectional transformers on Spanish Twitter with further fine-tuning on the data of the Sentiment Analysis workshop (TASS, 2020)[12] following the ELiRF-UPV model proposed at the workshop [10]. Catalan as a low-resourced language has a smaller number of pre-trained models. We leveraged a simple implementation of Naïve Bayes model trained with 50,000 tweets labeled using a lexicon-based approach [11].

The imageability assessment is performed over extracted concepts using psycholinguistic dictionaries as presented in [12]. One score is calculated as an average imageability of all concepts, another score - as an average of only the top-three imageable concepts, and the overall score is taken as a harmonic mean of both latter scores.

There are two endpoints through which creatives (architects, designers and artists) can interact with the results of the pipeline. Firstly, it is a demo tool[13] that allows users to analyze texts independently one by one. Secondly, extracted features are organized as a faceted search index Solr[14] that enables the functionality of querying information and performing data analysis on a collection level.

V. APPLICATION

We applied the developed information retrieval and analysis pipeline within two pilot case studies of MindSpaces. The first case study consists in identifying societal needs and sentiments in relation to the city of L'Hospitalet in Spain so that architects and artists could help to find ways to address identified issues throughout the urban design process. The second case study is dedicated to workplace design. The pipeline assists in the detection of contextual sentiment in relation to specific design features in workplaces (which design features do people prefer or dislike for creating privacy or for collaborating, and so on).

A. Social media sensing for outdoor design

Within the first case study, several Twitter and news collections have been processed and the results of analysis have been indexed using Solr to let users perform analytics activities. The currently running index has the information extracted from 55,000 messages in Catalan, 155,000 messages in English, and 190,000 messages in Spanish. The frequency of topics

[10]https://rajpurkar.github.io/SQuAD-explorer/
[11]https://github.com/sismetanin/emosense-semeval2019-task3-emocontext
[12]http://tass.sepln.org/2020/
[13]https://taln.upf.edu/mindspaces/demo/index.html
[14]https://solr.apache.org/

relevant to the purposes of the case study as classified by the aspect detection component is shown in Fig. 2.

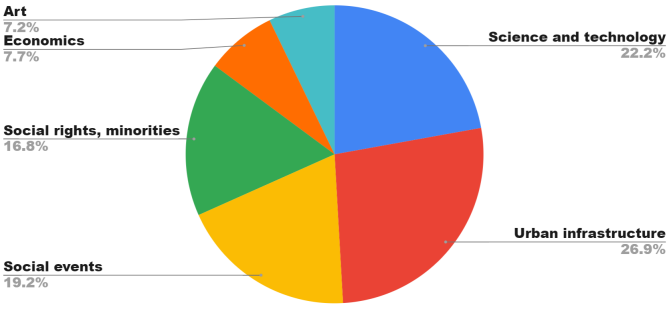

Fig. 2. Topics identified by aspect detection component in Twitter collections

Solr supports complex queries that can return statistics over individual values or get the number of co-occurrences of some attributes extracted from texts. For example, it is possible to count how strong some concept is associated with different sentiment polarities in a collection of texts devoted to some particular topic (Fig. 3). With this, based on the processed public opinions about outdoor places in the city of L'Hospitalet, users can get closer to understanding the importance of social problems such as community isolation, community integration, accessibility, insecurity and lack of public spirit.


```
"facet_pivot": {
    "ta.index.dbpedia,ta.index.emotag": [
        {
            "field": "ta.index.dbpedia",
            "value": "http://es.dbpedia.org/resource/Accesibilidad",
            "count": 540,
            "pivot": [
                {
                    "field": "ta.index.emotag",
                    "value": "NEU",
                    "count": 458
                },
                {
                    "field": "ta.index.emotag",
                    "value": "NEG",
                    "count": 225
                },
                {
                    "field": "ta.index.emotag",
                    "value": "POS",
                    "count": 214
                }
            ]
        },
```


Fig. 3. Outcome of a query that counts emotions related to "accessibility"

B. Web sensing for indoor design

The possibility of applying queries of various complexity with multiple filters over raw results of textual analysis also gives a high degree of freedom to creatives in formulating and testing their hypotheses in an attempt to find inspiration looking at the data from different perspectives. Thus, a series of research questions was formulated by the user partners of the MindSpaces consortium in order to perform an experiment on data exploration over constructed search indices within the second case study. First, a set of generic questions that had to be applied to various categories of concepts was provided: (i) How often a word/concept comes up; (ii) How often a word/concept comes up (related to office spaces); (iii) How much a word/concept/phrase comes up in relation to another word/concept/phrase (ranked list most to least); (iv) How often a word/concept/phrase comes up in negative or positive context; (v) What design elements are attributed positively/negatively with (ranked list most to least). Each category was given by architects participating in the case study with a definite list of concepts (5 to 40 per category):

- general: office, commercial space, corporate, open-plan;
- type of work: software/tech, finance, creative (architecture, graphic design, advertising), co-working;
- destinations / elements: desk, bookshelf, coffee machine, private office, stair, etc.;
- lighting: natural light, skylight, lamp, warm light, etc.;
- space: high ceiling, ceiling height, spatial proportions;
- sentiment / internal state / descriptive words: collaboration, visibility, focus, private, innovation, etc.;
- design parameters: layout / organization, furniture, partitions / privacy, materials / material palette, etc.

The queries to the faceted search index were successfully constructed for all of the questions and adequate results were obtained. The returned lists for the third and the fifth questions were ranked using TF-IDF weighting and up to 200 concepts were taken including newly identified, not listed by the architects. The returned results are shown in Fig. 4.

Fig. 4. Concepts associated with given topics grouped within faceted search

Second, for a given set of occupations (engineers, creatives, designers, etc.), there was a request to find associated materials, colors, shapes and arbitrary concepts when text is related to specific design parameters such as furniture, productivity, etc. Finally, user partners investigated preferences of workers of different occupations with respect to various space arrangements such as "quiet private spaces to work", "active social spaces to work", "partitions around the desk", and so on. The exploratory search returned highly interpretable results. We combined outcomes of several queries in Fig. 5 and Fig. 6.

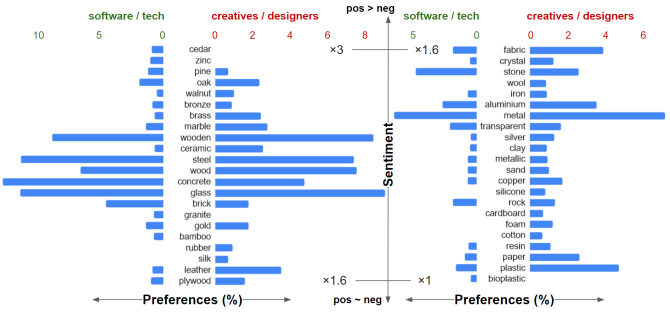

Fig. 5. Extracted relations and sentiments intersected by faceted search

	creatives / designers	software / tech
1	active social spaces to work	active social spaces to work
2	quiet private spaces to work	white spaces
3	white spaces	quiet private spaces to work
4	dark spaces to work	more natural light
5	colourful spaces	dark spaces to work
6	more natural light	colourful spaces
7	open space	open space
8	partitions around the desk	partitions around the desk
9	more space at desks	small collaboration spaces for informal meetings
10	small collaboration spaces for informal meetings	more space at desks
11	denser settings	denser settings
12	personalized desk area	activity-based-working / hot desking
13	activity-based-working / hot desking	personalized desk area

Fig. 6. Preferences in workspace arrangements based on statistics from Solr

Fig. 5 shows, on one hand, the usage of different materials in office spaces associated with different types of workers (horizontal axis) and, on the other hand, the sentiments about these materials (vertical axis[15]). The materials are sorted along the vertical axis by the proportion of positive over negative emotions. Thus, the top (top-left) materials gained three times more positive emotions than negative, while the bottom (bottom-right) materials received almost equal numbers of positive and negative emotions. Fig. 6 demonstrates the discrepancies in preferences in various workspace arrangements. The ranking of parameters associated with different occupations is similar and only several parameters change their position in at most two ranks. The outcomes of the experiment influenced the design process for creating alternative configurations of office spaces as 3D models to be used within simulations in VR.

VI. Conclusion

In this paper, an information retrieval and analysis framework was proposed. Targeted data-sets were formed, and natural language processing techniques such as concept extraction and sentiment analysis were applied to sense public opinions on matters of interior and urban design. Insightful results obtained within two applications were demonstrated.

The evaluation of crawlers, content and text algorithms occurs in the next months. The first case study describes an outdoors urban environment in L'Hospitalet de Llobregat, Spain concerning the renewal of an urban outdoors space via art-inspired design solutions promoting its cultural and environmental assets, improving flow and functionality for increased social interaction, tourism and economic activity. The second case study is about inspiring workplaces, extracting spatial and functional needs of modern work-spaces and accommodating smart sensing modifications. Users experience innovative, art-inspired designs through VR environments where data are measured and integrated to modify work-spaces, to elicit increased worker engagement, inspiration, interaction and productivity, while also improving functionality.

The presented pipeline can be used within research, technological and artistic projects where the focus on data exploration plays a significant role in the overall process towards understanding end-users needs, preferences, and sentiments. The models can be ported to other related domains with a possibility of fine-tuning with a moderate amount of data.

[15]For the sake of space, the data along the vertical axis split into two columns. The top part is on the left side, and the bottom part is on the right.

Acknowledgment

This work has been supported by the EC-funded project MindSpaces (H2020-825079)

References

[1] F. Erlandsson, R. Nia, M. Boldt, H. Johnson and S. F. Wu, "Crawling Online Social Networks," 2015 Second European Network Intelligence Conference, 2015, pp. 9-16.
[2] H. J. Carey and M. Manic, "HTML web content extraction using paragraph tags," 2016 IEEE 25th International Sympo- sium on Industrial Electronics (ISIE), 2016, pp. 1099-1105.
[3] V. Kagan and V. S. Subrahmanian, "Understanding Multi-Stage, Multi-Modal, Multimedia Events in Social Media," 2018 International Workshop on Social Sensing (SocialSens), 2018, p. 4.
[4] De Lange, M., & De Waal, M. (2017). Owning the city: New media and citizen engagement in urban design. In Urban land use (pp. 109-130). Apple Academic Press.
[5] Zhang, F., Zu, J., Hu, M., Zhu, D., Kang, Y., ... & Huang, Z. (2020). Uncovering inconspicuous places using social media check-ins and street view images. Computers, Environment and Urban Systems, 81, 101478.
[6] Vasiliki Gkatziaki, Symeon Papadopoulos, Richard Mills, Sotiris Diplaris, Ioannis Tsampoulatidis, and Ioannis Kompatsiaris. 2018. EasIE: Easy-to-Use Information Extraction for Constructing CSR Databases From the Web. ACM Trans. Internet Technol. 18, 4, Article 45 (November 2018), 21 pages. DOI:https://doi.org/10.1145/3155807
[7] Shvets, A., and Wanner, L. (2020). "Concept Extraction Using Pointer-Generator Networks and Distant Supervision for Data Augmentation". In 22nd International Conference Knowledge Engineering and Knowledge Management (EKAW), LNCS, vol. 12387, 120–135. Springer, Cham.
[8] Devlin, J., M. W. Chang, K. Lee, and K. Toutanova. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics, 2019: 4171-4186.
[9] He, R., Lee, W. S., Ng, H. T., and Dahlmeier, D. "An unsupervised neural attention model for aspect extraction". In ACL, 2017: 388-397.
[10] González-Barba, J.Á., Arias-Moncho, J., Hurtado Oliver, L.F. and Pla Santamaría, F., 2020, September. Elirf-upv at tass 2020: Twilbert for sentiment analysis and emotion detection in spanish tweets. In Proceedings of the Iberian Languages Evaluation Forum, CEUR, pp. 179-186.
[11] Balaguer, P., Teixidó, I., Vilaplana, J., Mateo, J., Rius, J. and Solsona, F., 2019. CatSent: a Catalan sentiment analysis website. Multimedia Tools and Applications, 78(19), pp. 28137-28155.
[12] Pistola, T., Georgakopoulou, N., Shvets, A., Chatzistavros, K., Xefteris, V., Táboas, A., Koulalis, I., Diplaris, S., Wanner, L., Vrochidis, S., and Kompatsiaris, I. (2022). Imageability-based Multi-modal Analysis of Urban Environments for Architects and Artists. In the 2nd International Workshop on Fine Art Pattern Extraction and Recognition (FAPER'22).