Contemporary Aesthetics (Journal)

1-1-2018

Aesthetics in Digital Worlds

Ossi Naukkarinen
Aalto University School of Arts, Design and Architecture, ossi.naukkarinen@aalto.fi

Darius Pacauskas
Aalto University School of Ars, Design and Architecture, darius.pacauskas@aalto.fi

Follow this and additional works at: https://digitalcommons.risd.edu/
liberalarts_contempaesthetics

   Part of the Esthetics Commons

Aesthetics in Digital Worlds

Ossi Naukkarinen
& Darius Pacauskas

Abstract
Aesthetics is often seen as a philosophical academic discipline focusing on questions about art, beauty, the nature of aesthetic experiences, and many other related issues. However, there is another, non-academic side of aesthetics where similar issues are addressed. Non-academic cases of aesthetics, on the internet and elsewhere, far outnumber anything academic aestheticians can ever produce. However, how the picture of aesthetics looks like outside academia is not necessarily very actively considered in academic contexts, as professors, lecturers, and students tend to form their understanding of the field through scholarly literature and academic datasets. In this essay, we explore how close to or far away from each other the two environments presently are, especially in digital environments. For this, we applied computational text-mining techniques on Wikipedia, Google Trends, YouTube, Open Library books and Web of Science data. Results allowed us to compare both areas and describe differences and relations between the two. This article shows that there are various contexts where issues related to aesthetics are addressed, and that the overall picture of aesthetics is created by different kinds of actors and documents. Additionally, we suggest that digital tools can and should be used in disciplines such as aesthetics to create a more comprehensible and many-sided picture of the field, while taking into consideration the risk of using these tools too narrowly and over trusting them.

Key Words
aesthetics; altmetrics; data-mining; digital humanities; text-mining

1. Introduction

Aesthetics is often seen as a philosophical academic discipline focusing on questions about art, beauty, the nature of aesthetic experiences, and many other related issues. On the other hand, similar issues are also dealt with in non-academic and non-philosophical contexts.

The borderline between the two areas, academic and non-academic, has traditionally been blurry in aesthetics. Widely spread and interdependent ideas about art, beauty, aesthetic experiences, and other issues that are important in aesthetics have been developed by scholars, critics, artists, curators, designers, and others, and they affect each other. This can easily be seen when looking back at the history of the field, where many authors who are still important today have been operating in both areas and on their borderlines: Friedrich Schiller, Leo Tolstoy, Walter Benjamin, Arthur C. Danto, Umberto Eco, Susan Sontag, and Nicolas Bourriaud, to name some examples. In this respect, aesthetics seems to be a rather typical sub-field of humanities and can function as an example of humanities at large.

Despite traditional overlaps and borderline cases between academic and non-academic variations of aesthetics, nowadays the two areas often tend to be more separated than before. In performance evaluations of scholars and research groups done in universities, for example, it is typical to pay attention to publications listed in academic databases, such as Web of Science and Scopus, and ignore more journalistic essays, blogs, or tweets. Also, such non-academic activities as artists' statements and artworks are typically put aside even if they can have a strong impact on the aesthetic discourse at large. This kind of separation is not as strict in art and design universities, especially in the ones where so-called artistic research is being developed, but in our experience it is rather normal in universities focusing on sciences and humanities.[1] However, there is a risk that if we try to form a picture of contemporary aesthetics only through standard academic sources, our picture is biased. Moreover, putting too much emphasis on narrowly understood academic activities restricts and changes the traditional area of operation of humanistic scholars even at universities. In our opinion, it would be beneficial for the field if our approach would be broad, not narrow. In this essay, we explore some ways how new digital research tools can serve this kind of broadening.

Therefore, we would like to explore how the two areas, academia and actors outside of it, form interpretations about issues important in aesthetics broadly taken, and see how close to or far away from each other these two environments presently are. Here, 'academic' primarily refers to scholars affiliated to universities and to their publications that are included in scholarly databases. 'Non-academic' and 'outside academia' mainly refer to actors and their publications and other products typically presented in non-scholarly contexts. Of course, the borderline between these two is not absolutely tight, and some data sources such as YouTube include both academic and non-academic publications and other items. However, there are actors, publications, and datasets that are more academic and others that are less so. The point of this essay is not to define these areas and their borderlines in detail but to examine how the picture of a sub-field of humanities, aesthetics,

varies when seen through different contexts, and how different interpretations could be combined. How varied is the overall picture? What should be examined if we want to have an overview of aesthetics at large? And how could such overviews be created, for example, for the purposes of faculty evaluations in humanities?

Especially, our aim is to critically examine the present possibilities provided by digital humanities for understanding fields like aesthetics and to clarify what we can expect from such approaches and what we cannot. There is a risk that digital approaches that are more and more often used in academic contexts lead to severely restricted results. However, when used critically and wisely and combined with other, more philosophical approaches, they can offer us new kinds of insight that could not be obtained in any other way. It is self-evident that as people spend more and more time in digital surroundings via their computers, tablets, phones, and other devices, their world consist of digital environments and things taking place there, aesthetic and otherwise. It is crucial to understand how such environments and their analyzing tools function.[2]

**2. Main results: aesthetics in plural**

For this article, we compared the picture of aesthetics provided by one of the major academic database, Web of Science (WoS), with aesthetics discussed in commonly used and large information sources that are largely non-academic, that is, Google Trends, YouTube, and Open Library books. For making the comparison possible, we created a kind of world map of knowledge based on English Wikipedia and further examined the characteristics of each of our information sources on this map. As far as we know, no such comparison has been done before, and a detailed description of the process is given in the methodology section below. Our main contribution is to introduce this approach and some of the novel possibilities it opens up. Within the limits of this text, it is unfortunately impossible to go deeper into case studies and analyze, for example, exactly how certain themes such as health, that, from our perspective, seem to be important outside academia could be better taken into account also in the academic discourse.

The main results indicate significant differences between different interpretations of aesthetics depending on through which database and set of documents we approach it. One interesting question is what kind of picture and created by whom should be seen as important, if we want to understand the field and its variations.

In each database, the documents that the database includes and that are related to aesthetics are connected to somewhat different categories of things. The clearest differences are between WoS (Figure 1a) and YouTube (Figure 1b). Their profiles are rather different. This can best be seen through a visual presentation created with the tool called LDAvis.[3] It shows different categories or topics as bubbles or circles. The size of each circle tells how widely each topic covers the documents, or rather the groups of words that form the documents, that are related to aesthetics in the given dataset.

The picture (Figure 1) shows the seventy-five most important topics related to aesthetics in these datasets.

Figure 1: Topic modeling results:
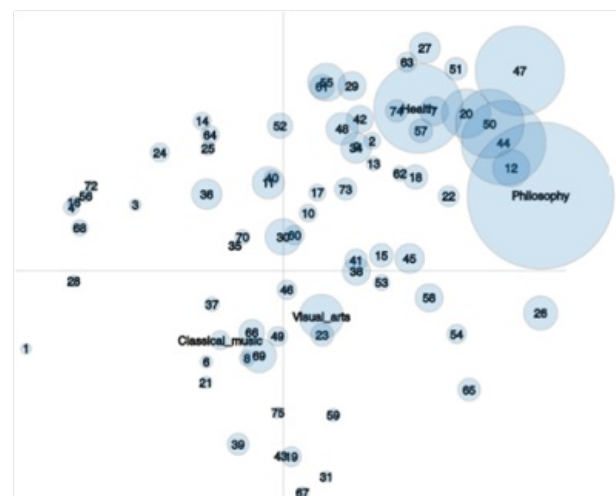

a) WoS (All databases),
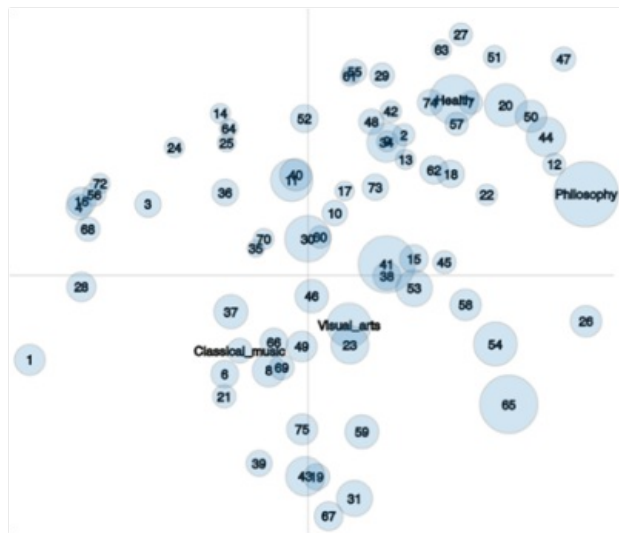
b) YouTube

Interactive image can be viewed here: http://dhoa.aalto.fi/twoworlds/#figure1.

The bigger the bubble is, the bigger the share of the documents (groups of words) that are related to aesthetics it contains. Topic modeling approach, which is described in more details later in this article, does not give names to such topics but only categorizes words according to their relations to each other when the number of topics is decided in advance. Human readers can then name such categories if they like. Here, for example, we have named the biggest category as "Philosophy." Some others could be called "Visual Arts," "Classical Music," and "Health." In fact, LDAvis tool provides an interactive scene where one can click the bubbles and find many other layers of additional information. That feature cannot be included into static documents but can be explored here (http://dhoa.aalto.fi/twoworlds/#figure1).

The documents that are related to aesthetics on YouTube, covering videos and their verbal descriptions, are much more evenly spread across all kinds of topics than in WoS, which covers mainly academic articles and has some more dominant topics. The dominant topics of aesthetics in WoS are also fairly close to each other on the upper right corner of the image, suggesting their close semantic relationship. This means that what people think of aesthetics and where they relate it to in YouTube is more versatile than in academia.

In Figure 1a, we looked at the aesthetics related publications within the whole scientific database of WoS, but if we restrict our view only to aesthetics publications that are within the Arts & Humanities category of WoS and in journals specializing on philosophical aesthetics, the view will show that documents are even more philosophy oriented (Figure 2a). Similar assumptions arise when we look at how aesthetics is categorized in Open Library books (Figure 2b). Of course, our analysis thus far only suggests this idea and it would require additional layers to strengthen the initial results.

Figure 2: Topic modeling results:
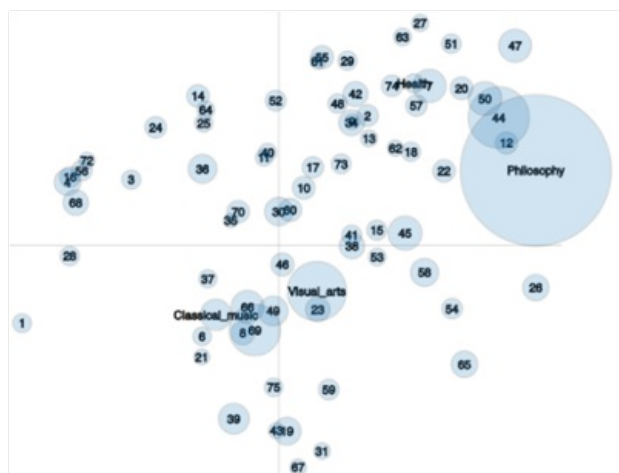
a) WoS (A&H)
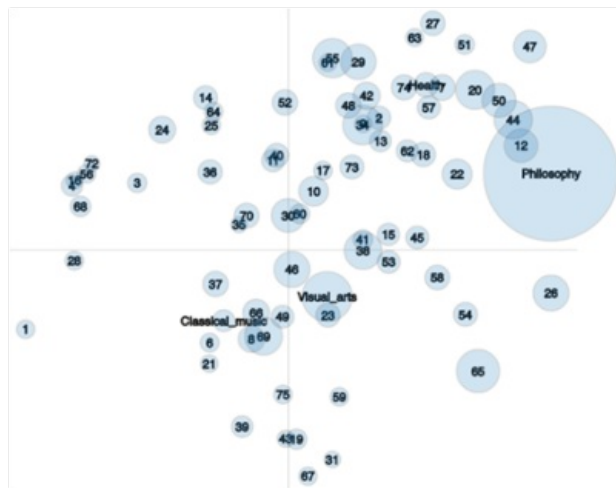
b) Open Librarybooks (ALL)

Interactive image can be viewed here: http://dhoa.aalto.fi/twoworlds/#figure1.

Comparing the top topics in different datasets can also be shown as a table (Table 1). Here, it is plain to see that even some top-five topics are rather surprising from the point of view of the academic philosophical aesthetics.

Table 1: Scores and most important words for topics[4]

YouTube

Topic Score	Topics	Dominant Words in Topics
3.53	Philosophy	idea, think, describe, view, theory, concept
2.81	Topic_65	tell, friend, want, love, leave, feel
2.70	Topic_41	http, online, video, user, internet, free
2.02	Health	cancer, disease, treatment, blood, patient, cause
1.77	Topic_30	fashion, wear, model, clothe, dress, designer

Google Trends

Topic Score	Topics	Dominant Words in Topics
12.66	Philosophy	idea, think, describe, view, theory, concept
5.01	Health	cancer, disease, treatment, blood, patient, cause
4.84	Topic_44	self, social, psychology, individual, behavior, disability
3.63	Topic_11	medical, hospital, health, medicine, care, nurse
3.03	Topic_30	fashion, wear, model, clothe, dress, designer

Open Library books (Titles and Descriptions)

Topic Score	Topics	Dominant Words in Topics
9.11	Philosophy	idea, think, describe, view, theory, concept
8.59	Topic_38	student, education, teacher, class, teach, campus
5.84	Topic_50	technology, information, management, system, networks, data
4.79	Topic_10	library, volume, edition, collection, page, archive
4.78	Topic_44	self, social, psychology, individual, behavior, disability

Open Library books (paragraphs from full texts)

Topic Score	Topics	Dominant Words in Topics
17.06	Philosophy	idea, think, describe, view, theory, concept
2.33	Visual arts	artist, museum, paint, gallery, exhibition, art
1.73	Topic_65	tell, friend, want, love, leave, feel
1.47	Topic_34	window, roof, wall, building, architect, floor
1.44	Topic_55	forest, reserve, conservation, natural, environmental, nature

Web of Science (Arts & Humanities database)

Topic Score	Topics	Dominant Words in Topics
22.36	Philosophy	idea, think, describe, view, theory, concept
3.67	Topic_44	self, social, psychology, individual, behavior, disability
3.42	Visual_arts	artist, museum, paint, gallery, exhibition, art
2.59	Topic_69	novel, writer, author, literature, poet, poetry
1.13	Topic_45	political, movement, union, communist, party, protest

Web of Science (all databases)

Topic Score	Topics	Dominant Words in Topics
21.04	Philosophy	idea, think, describe, view, theory, concept
7.74	Health	cancer, disease, treatment, blood, patient, cause
7.71	Topic_47	model, method, measure, size, sample, analysis
6.93	Topic_44	self, social, psychology, individual, behavior, disability
4.57	Topic_50	technology, information, management, system, networks, data

Detailed table with sorting by column option can be viewed here:

Some of the topics that seem to be important for aesthetics in YouTube are marginal in WoS. For instance, in YouTube Topic_65 cluster words such as 'friend,' 'love,' want,' 'leave,' and 'feel' with aesthetics. We cannot just think that this sounds weird considering what kinds of themes traditional textbooks of philosophical aesthetics cover, or that we can forget such themes because they seem to be irrelevant. Maybe they are relevant for the majority of non-academics, and maybe the majority really connects aesthetics with such themes that could perhaps be called 'human relations.' What does this actually mean in and for aesthetics, in and outside academia? This is a question that would require more analysis, but the important thing for now is that such novel questions would be impossible to find without large-scale data mining. If we want to broaden our understanding of the field, new digital tools seem to be useful in this process. These initial results suggest, at least, that academics should maybe open up their horizons and pay more attention to things that the rest of the world seem to think are aesthetically interesting and important.

Another issue worth considering, especially for philosophical aestheticians, is that both in and outside academia related to health and medical issues seem to be pivotal for aesthetic discussions, for example, topics "Health," 44, and 11. This, too, is something that is rarely mentioned in philosophical books and articles on aesthetics, let alone analyzed in more detail. Is this ignorance worth preserving or should non-academics, empirical scientist, and philosophers of aesthetics start to learn from each other? Of course, there are some rare cases where this has already been done, broadly taken, such as Anjan Chatterjee's book, *The Aesthetic Brain*, but it might be a good idea to notice this also in other contexts.[5]

It may also come as a surprise that, say, the topic of visual arts only typically covers some percentage of the documents related to aesthetics. True, when combined with topics that are related to other art forms, such as music, their total coverage is fairly high but it is still quite clear that aesthetics by no means can be seen as the philosophy of art alone, even if, since Hegel, aesthetics and philosophy of art have sometimes been seen as identical.

The results suggest that in all the databases that we analyzed, the biggest topic group includes documents that are related to philosophy. So, in all of these contexts, one strong strand of aesthetics is aesthetics seen as a philosophical field of activities. In WoS, it covers more than 20% of documents in aesthetics and even in Google Trends, more than 12%. In this category, aesthetics is clustered together with words such as 'idea,' 'think' and 'theory.'

However, the size of a single category does not directly reveal the importance of the area itself. If we add some other aspects to be used as interpretation tools, the picture changes. In YouTube, for example, we can find topics that alone are smaller than philosophy, but together they seem to form a group of categories that are related to different types of music. Combined, they cover more: Topic_31 is 1,1%, classical music is 0.5%, Topic_39 is 0.6%, Topic_49 is 0.8%, Topic_59 is 1%, topic 75 is 0.8%, which equals 4.8% versus 3.5% of philosophy topic coverage. This, too, would be impossible to know without computational tools.

Moreover, we could also look at the matter from a different perspective and pay attention not only to what kinds of documents all sorts of aestheticians tend to create but what kind of documents users tend to consume and enjoy. If we examine not only the number of documents (videos) in YouTube, but also their views, comments, likes, and dislikes, and thus create an alternative to ordinary scientific measurements, that is, use so-called altmetrics, it seems that, in average, a single document gets more attention in, say, classical music than in the big group of philosophy (Table 2).[6] So, it might be that in YouTube these documents have a stronger role in defining the overall picture of aesthetics than philosophy videos, which are numerous but are not viewed and "liked" as often. This becomes close to paying attention not only to the number of publications of, say, a scholar but also to the number of references they get, resulting in their h-index.

Table 2: Altmetrics from YouTube dataset [7]

Topic	Topic Score	Views in millions	Views per doc	Comments per doc	Likes per doc	Dislikes per doc	Number of docs.	Coeff
Classical Music	0.51	56.2	36,475	44	370	17	1,542	94,211
Health	2.02	258.8	37,654	35	303	16	6,873	84,896
Topic 45	0.44	50.7	33,579	25	330	11	1,510	84,824
Topic 11	1.48	201.5	32,998	31	274	17	6,106	75,682
Topic 35	0.25	6.8	22,280	34	299	9	303	68,948

Detailed table with sorting by column option can be viewed here:

Looking at what users tend to like related to aesthetics in Open Library books, we can see that philosophy is playing a big role here (Table 3). If we compare the so-called YouTube altmetrics with the Open Library one, we can assume

that these two environments are surrounded and created by different kinds of communities. Books are more often probably consumed by academics, while YouTube videos by the "general mass" or maybe by a number of distinctive communities united under one umbrella. However, it is an assumption that cannot be confirmed by the data we gathered, but it seems clear that users surrounding these platforms differ in taste and understanding of aesthetics.

Table 3: Altmetrics from Open Library books dataset:[8]

Topic	Topic Score	Views	Views per doc	Docs
Philosophy	17.06	3,920,800	71	55,080
Topic_55	1.44	242,005	44	5,481
Topic_20	1.43	222,867	42	5,368
Topic_65	1.73	362,472	41	8,801
Topic_26	1.21	246,656	39	6,391

Detailed table with sorting by column option can be viewed here: http://dhoa.aalto.fi/twoworlds/#table3.

Another size-related issue is that clearly the number of people contributing and amount of documents related to non-academic databases are much higher than in the academic ones, even if we could cover only a fragment of potentially relevant non-academic databases for this article. Does this mean that what is bigger is more important than what is smaller? Data mining, as such, cannot answer this. It only gives us results that describe those aspects of databases that can be quantified and, as such, does not take a stand on issues of qualitative value, norms, or the like. It gives us materials that could help us think of such issues from fresh perspectives. Smaller might be more valuable, in some respects, as was shown in the previous paragraphs. Yet, the fact that the volume of non-academic contributions, all in all, is much higher, for example, around 800,000 YouTube videos that include the word 'aesthetics,' compared to around 20,000 of WoS entries, suggests that, in the field of aesthetics, large, non-academic contributions are more common and widely spread and so should be taken into consideration also in academia, if we hope to form a comprehensive picture of the whole.

However, at the moment it is quite clear that analyzing datasets, academic and otherwise, tells only of certain aspects of the field of aesthetics. The results that we achieved are based on verbal data only. For now, it is not possible to analyze pictures or sounds directly, and they cannot even be found if they do not have verbal metadata attached to them, calling them aesthetics; or, rather, it is not possible for purposes of this kind of analysis where the nature or a whole cultural field is the target. The tools we used, and many others, cannot identify a picture as an interesting contribution to the discussion about aesthetics if no one has named it as such. Yet, it is quite possible to think that, say, many well-known visual works of conceptual art, such as Joseph Kosuth's classical *One And Three Chairs* or Ai Wei Wei's *Dropping A Han Dynasty Urn*, address issues that are central to aesthetics at large. Also, many fashion and design blogs, sometimes created by people working in art and design universities, are undoubtedly interesting for aesthetics, even if the main aspects of them are visual, not verbal. Standard academic analyzing tools cannot necessarily find them. Pictures and sounds can be digitally analyzed for other purposes, of course. For example, the algorithm called *EMI* (Experiment in Musical Intelligence) can analyze and compose music, and there are tools that can identify visual forms without verbal metadata, such as Google image search.

Moreover, the verbal data available is typically restricted to titles, keywords, abstracts, and other short descriptions of larger contents, and no full-text analyses can be done. All this is normally only in English, in the most widely used databases. Through them, we cannot attain any information about aesthetics done in other languages. So, even if the amount of data is very big, it should not make us believe that a good grasp of, say, Korean, Swedish, or Italian aesthetics could be reached through it. Even Mandarin and Spanish are missing, even if globally there are more speakers than of English. These tools are useful and reveal information that cannot be found by any other means, but the results must still be critically examined and they cannot replace more traditional approaches.

An additional restriction is that even if we were satisfied with results based only on English, the type of approach we tested is heavily dependent on the very word 'aesthetics.' We can only cover documents that include that word in one form or another, in the title, key word list, verbal video description, or the like. Still, it is most probable that there are plenty of documents that are quite relevant for aesthetics but do not use that terminology, even if they are verbal. This is a very important element in the equation, and it is not quite clear how much our results tell about the field of aesthetics at large or only about occasions where the word 'aesthetics' is used. Above all, both options are referred to without taking the final stand on the matter.

Let's look at this scenario to further illustrate the situation. Now we know that the word 'aesthetics' is typically related to certain clusters of words that form topics. Based on this, we could assume that if we removed the word

'aesthetics' itself and used the discovered clusters of words, we could find documents that relate to aesthetics without using the word 'aesthetics.' For instance, with such clusters, we could go through whichever dataset and see how many documents cover a topic that combines words such as artist, museum, paint, and gallery. However, we should not to forget that the topics we have got were based on the 'aesthetics'-related dataset, and thus giving just any general dataset would bring results about that topic but not necessarily about aesthetics. In our case, searching with words related to museum or the classical music topic would discover documents that discusses museum and classical music, but not necessarily related to aesthetics, unless we think that museums and music are always sub-categories or the like of aesthetics. In other words, using topic modeling on aesthetics-related documents will bring us topics that probably relate to aesthetics, but oppositely using those generated topics on any other dataset will not necessarily discover aesthetics-related information.

Also, for some other topics that are clearly relevant for aesthetics when that word is used, say, the one that we called 'philosophy,' the cluster of the most frequent terms seems to be of the type that, by using only them and excluding 'aesthetics' from the list, we would probably end up covering a mass of documents that are relevant for philosophy but not for aesthetics.

In fact, the initial results that we achieved through data mining academic and non-academic databases are much more detailed and numerous than can be presented in one single article. These kinds of results are new, and they also generate novel questions that emphasize that using computational approaches refreshes and broadens the possibilities of scholars in aesthetics and thus should be used. However, as they are not yet very well known in aesthetics, it is good to spend some space for introducing how they can be used. So, how did we actually achieve these results, and how could similar methods be used elsewhere?

3. Methodology

The basic process of the data collection aimed at capturing the topics (themes) that aesthetics-related text publications are covering in academia and outside. We examined five different kinds of data (table 4):

(1) Wikipedia articles

(2) Web of Science (WoS) scientific database

a) WoS, all databases

b) WoS, Arts & Humanities databases

(3) Google Trends

(4) YouTube

(5) Open Library books

a) Titles and description from Open Library books

b) Paragraphs from full texts from Open Library books

First, all possible topics needed to be identified. For that we used (1) English Wikipedia articles. English Wikipedia is the largest online encyclopedia created and managed by volunteers. Any big textual dataset could be chosen for this purpose, such as news portal articles or movie reviews, but Wikipedia is arguably the best known.

Later, based on the Wikipedia-generated topics, we retrieved topics that present interests of different aesthetics communities, including academia and non-academia. To represent academia we used (2) WoS scientific database. To represent other areas, data from (3) Google Trends, (4) YouTube, and (5) Open Library books were used.

3.1 Data collection

Except data from Wikipedia that was retrieved in total, the rest of datasets were retrieved by using the keyword 'aesthetics.' From each data entry we took its title and description. An exception was made for Open Library data where, in addition to titles and descriptions, we also took a paragraph from each full text book where the word 'aesthetics' was mentioned.

Table 4: Datasets used:

Dataset	Title	Description	Entries in dataset / Entries used	Purpose	Scope
Wikipedia	Article's name	Article's text	3 671 353 / 3 671 353	Creating topics	Whole English Wikipedia
WoS (A&H)	Publication's title	Abstract	9 360 / 9 360	Finding related topics	Keyword 'aesthetics' + aesthetics journals in Arts & Humanities database
WoS (Whole)	Publication's title	Abstract	18 536 / 18 536	Finding related topics	Keyword 'aesthetics' + aesthetics journals in all WoS databases
Google Trends	Related topic to the query	Wikipedia article's summary related to the title	85 / 76	Finding related topics	Keyword 'aesthetics' for each year
YouTube	Video Title	Video description	122 087 / 69 141 downloaded (771 000 was available)	Finding related topics	Keyword 'aesthetics'
Open Library books (Aesthetics)	Book Title	Book description	73 663 / 37 097 downloaded (111 000 was available)	Finding related topics	Keyword 'aesthetics' in titles and description
Open Library books	Book Title	A paragraph from full text where "aesthetics" is mentioned	73 663 / 71 105 downloaded (111 000 was available)	Finding related topics	Keyword 'aesthetics' in full text

3.1.1 Data collection: Wikipedia

English Wikipedia articles were collected from a Wikipedia-released data dump covering 3,671,353 articles created up to 2017 February. It was used to categorize a variety of topics, in this case 200, regardless of their relation to the area of aesthetics, under which comparison would be done.

3.1.2 Data collection: Web of Science

We used two different WoS based datasets: one consisting of works that were published in the category of Arts & Humanities and another including all fields. For instance, there are plenty of publications that use the word 'aesthetics' in chemistry or technology. With the first dataset, we captured aesthetics that relates solely to humanities; with the second, to all possible fields of study. This approach is comparable to the data collection principle that was used in the article by Ossi Naukkarinen and Johanna Bragge in 2016.[9]

3.1.3 Data collection: Google Trends

In addition to trendy topics for the desired period, Google Trends can also show related topics for the keywords that are used. This option is described as "*Users searching for your term also searched for these topics.*We treated these related topics [D9] as titles. However, the description of these topics and titles was not provided by Google Trends. Thus, in order to compare this dataset with other ones, we needed to find a description for each of the titles gathered from Google Trends. The description for each of the eighty-five titles was gathered from Wikipedia while inputting the title as a search word and getting a summary of the related article. Thus, the summary was treated as a description. Nine out of the eighty-five entries did not receive a summary, due to no Wikipedia entry, or there were too many entries related to the same keyword, for example, 'band,' 'health,' or 'value.' Thus, these entries were removed from a further analysis.

3.1.4 Data collection: YouTube

For getting YouTube-related data, we used the application programming interface (API) provided by Google. The results covered approximately 800,000 entries but, due to the slow process, we were able to gather around 122,087 of them. However, there were videos whose verbal descriptions did not include enough English words to make an analysis of, or they did not include the word, 'aesthetics.' Thus, we removed those from a further analysis and had data of 69,000 videos.

3.1.5 Data collection: Open Library books

To collect book-related, and also journal-related, entries from Open Library (www.openlibrary.org), we developed a web crawler. We excluded patent registry data. The key term 'aesthetics' resulted in around 111,000 entries, and we were able to gather 73,663 of them. The term was searched throughout all the text, and to form one dataset we removed entries that didn't include 'aesthetics' in the title or in their description, and we also excluded cases where the description of the book was missing. For another dataset we took all entries with a paragraph from book that included the word 'aesthetics.'

3.2 Analysis

Analysis consisted of three parts. During the first phase, the model was prepared and all topics were identified, based on English Wikipedia articles (Figure 3). In the second phase, different kinds of datasets were analyzed to see which datasets covered which topics, and to what extent (Figure 4). In the third part, we analyzed two datasets, YouTube and Open Library, to look into alternative metrics for seventy-five topics.

3.2.1 First phase of analysis

As the main analysis tool, of particular importance for the first stage, we chose a topic modeling approach.[10] The reasons behind this choice lie in the balance of ease of use and usefulness.

To begin, the data (text) was pre-processed. Punctuation was removed and words were transformed to lowercase and lemmatized, which means grouping the inflicted versions of a word together, for example, 'aesthetics' became 'aesthetic.'[11]

Topic modeling requires specific input for the analysis to form the model based on which documents will be segregated to topics. First, it requires data that is separated into a set of documents, that is, the text shouldn't be one continuous piece, it should have clear starting and ending points, and there should be multiple documents. It is important for the workflow of topic modeling, as the main purpose of modeling is to group words based on their co-occurrence in different documents. For instance, if words such as 'Kant' and 'philosophy' often appear in the same documents, they will probably be included in the same topic.[12]

Figure 3. 1st phase of analysis: Gathering all topics that the world is interested in, as seen through Wikipedia.

An interactive image can be found here http://dhoa.aalto.fi/twoworlds/#figure2.
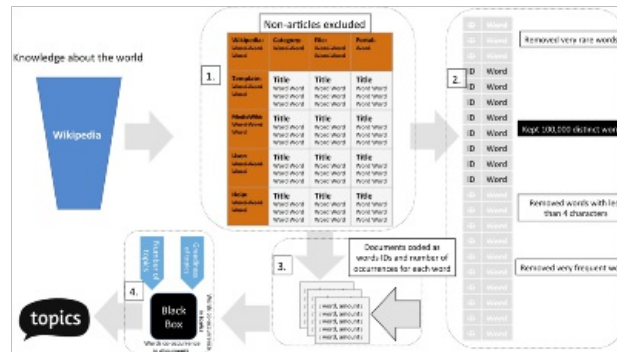
Another requirement for topic modeling is a predefined dictionary, that is, a set of words that is included into analysis. A dictionary defines which words the tool will take into account while iterating through the given documents. For example, if one decides to leave only nouns in the dictionary the computer does not notice verbs.[13]

The next requirement is to transform each document into a bag of words according to their existence in the dictionary and times of their occurrence or the amount each word is repeated in each document. This process is done for each of the documents, with all words from the dictionary.[14]

Furthermore, there are two parameters that are usually left for the user to be decided: (1) the number of topics we want to see as the final outcome, and (2) the hyper parameter *alpha* that defines the distribution of different words throughout topics. *Alpha* parameter defines the so-called greediness of the each topic. If it is too high, one topic can become very big in size, that is, the result will be that most of the documents will be about one topic. If the *alpha* parameter is very low, all the topics will be similar in size and might include words that do not relate to each other in any meaningful way. When these two parameters are seen to be appropriate depends on the researchers' experience of their field and on their beliefs about the phenomena they are studying.[15]

3.2.2 Second phase of analysis

Data entries of each dataset needed to be prepared in a similar way, as was done in the first phase of analysis. Each document was pre-processed, removing punctuation marks, numbers, lowercasing, and lemmatizing words. Later on, we used the same dictionary that was created in the first phase of analysis, for transforming documents to a bag of words. The final analysis stage was different from the first phase of analysis. The aim was to use the generated topics and check the likelihood of documents belonging to those topics. It was done by taking documents from a dataset and checking to what extent every word from a particular document belongs to one or another topic.

Figure[13] **4.** The second phase of analysis: identifying which of the gathered topics each community is interested in.

An interactive image can be found here
http://dhoa.aalto.fi/twoworlds/#figure2.
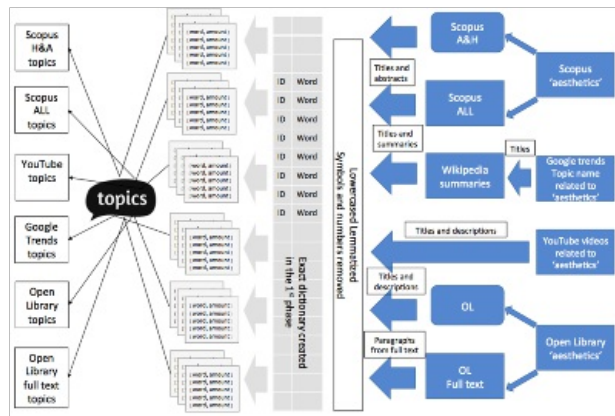
3.2.3 Third phase of analysis

Data entries from YouTube (title and description), and Open Library books (paragraphs from full texts) followed the same preparations as the first and second phases of analysis. We got the topics distributions for all the documents in the datasets, and summed alternative metrics or views, comments, "likes," and "dislikes" from each of the YouTube video (document) under the particular topic, and the amount of views of each of the book from Open Library books.

Out of 200 initial topics, we eliminated topics that do not bring any remarkable value in understanding the field of aesthetics and left only those that have a considerable role at least in one of the datasets. We ended at seventy-five topics. Limiting topics was a choice to allow better information processing for our readers. Our graphs (Figures 1 and 2) contain seventy -five bubbles instead of 200.

4. Conclusion

We have shown that there are various contexts where issues related to aesthetics, or 'aesthetics,' at least, are addressed; that different pictures of the field are created by different kinds of actors and documents; and that different contexts differ from each other. As such, this is not surprising, but with computational methods one can drill deeper in this theme. We begin to see what kind of differences there are and what kinds of things are typical for different sub-areas, and we may also find discourses and themes that are clearly important for some corners of the field but that we have ignored thus far. This, in turn, may provide us new themes and materials to study. For example, the aesthetics of health and human relations seem to be important elsewhere, and that could motivate philosophical, academic aestheticians to work more actively around them. Where this could take us remains to be seen.

We have also suggested that now when contemporary digital or computational analyzing tools are available, they can and should be used in disciplines such as aesthetics to create a more comprehensible and many-faceted picture of the field and to open up new, unforeseen questions to be addressed in the next stages of the analysis. This, we think, is recommendable when more and more of our activities are becoming digital, taking place through computers, smart phones, pads, and other similar tools.

However, we emphasize that in some contexts, like academic evaluations, there is a risk of using databases and other digital approaches too narrowly and trusting them too easily. This is particularly important in such humanistic fields as aesthetics, where other, less academic or scientific activities have traditionally played an important role in developing ideas and practices. We need tools that can cover both academic and non-academic strands of the discussion. Here, we have opened some initial routes that could be followed further.

Aesthetics, here, functions as an example of a question that is important for the arts and humanities at large. Our suggested inquiry can also be applied to other areas, such as ethics. Here, we only focus on instantiations of aesthetics in digitized environments, without claiming that non-digital cases of aesthetics did not exist. However, digital environments play a bigger and bigger role in our lives all the time, and that is why they call for careful and many-sided analyses. We assume that especially different types of altmetrics combinations will be needed because they can open up more varied kinds of views on the examined phenomena than single approaches. Of course, altmetrics would not only provide new insights but also open additional problems. For example, if we wanted to use data owned by the so called GAFA companies (Google, Apple, Facebook and Amazon) can we have it, and how would such giant corporations restrict and guide our operations? These companies, in the end, are the ones that affect our lives very strongly, aesthetically and otherwise. They should be analyzed, but is that really possible?

In the end, we would also like to emphasize that even if we had the best possible computational tools to help analyze the largest datasets available, that would only tell something of aesthetics within these datasets. True, we all are more and more intertwined with the digital tools that we use and may feel that we are completely merged with them, and we ourselves may be just very complicated sets of algorithms. But even as such, we cannot help feeling that there are lots of aesthetically important things that can only be experienced with our very physical, analog bodies and senses. No computational analysis can substitute that for us as experiencing and feeling human beings. There are lots of tools that can help us analyze academic and non-academic knowledge. However, we should be cautious when applying those tools and not give the primary investigator's role to the computer. It should only take the position of a research assistant.[16]

Ossi Naukkarinen
ossi.naukkarinen@aalto.fi

Ossi Naukkarinen, PhD, is Professor of Aesthetics at the Aalto University School of Arts, Design and Architecture, Finland. He has previously published articles in *Contemporary Aesthetics* on mobile aesthetics (2005), artification (2012), and everyday aesthetics (2013 and 2017).

Darius Pacauskas
darius.pacauskas@aalto.fi

Darius Pacauskas, PhD, is Postdoctoral Researcher in the Aalto University School of Arts, Design and Architecture, Finland.

Published June 12, 2018.

Endnotes

---

[1] See Mika Hannula, Juha Suoranta, and Tere Vadén,*Artistic Research Methodology: Narrative, Power and the Public* (New York: Peter Lang, 2014); Maarit Mäkelä and Sara Routarinne, eds.,*The Art of Research: Research Practices in Art and Design* (Helsinki: University of Art and Design Helsinki, 2006).

[2] This article can be seen to form a pair with the article "Aesthetics in the Age of Digital Humanities," *Journal of Aesthetics and Culture* (Vol. 8, 2016, Issue 1), at http://www.tandfonline.com/doi/full/10.3402/jac.v8.30072 (10 April 2018) where Ossi Naukkarinen and Johanna Bragge analyzed the field of aesthetics through Web of Science. This time, we included not only WoS but also data gathered from other important sources. Even if the data mass is much bigger than in the article by Naukkarinen and Bragge, we will show that the achieved results must still be taken *addito salis grano*.

[3] See Carson Sievert and Kenneth E. Shirley, "LDAvis: A method for visualizing and interpreting topics" in *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces* (Baltimore, Maryland, USA, June 27, 2014), 63–70. https://nlp.stanford.edu/events/illvi2014/papers/sievert-illvi2014.pdf (10 April 2018) for better understanding of the tool.

[4] Topic scores relates to the size of the topic, or size of the bubble if referring to the Figures 1 and 2.

[5] Anjan Chatterjee, *The Aesthetic Brain: How We Evolved to Desire Beauty and Enjoy Art* (New York: Oxford University Press, 2013).

[6] Although traditional scientometrics focuses heavily on citations for recording the impact of academic research, the rise of social media has opened up several alternative new channels for tracking the impact. These measures are called altmetrics, and described as scholarly impact measures based on activity in online environments. Erdt Mojisola, Aarthy Nagarajan, Joanna Sei-Ching Sin, Theng Yin-Leng, "Altmetrics: An Analysis of the State-of-the-art in Measuring Research Impact on Social Media," *Scientometrics*, 109(2), 2016, 1117–1166, https://doi.org/10.1007/s11192-016-2077-0 (10 April 2018).

[7] Table sorted by coeff score, which calculated as follows: coeff = views + 835*comments + 109*likes - 1950*dislikes. This coefficient was developed as an easy way to compare videos or topics. However, there are more views per video than comments, and there are more comments than "likes" and "dislikes," on average. Therefore, coefficients for each of the variables (i.e. 835, 109 and 1950) were needed to be determined, as otherwise only views would determine positions of topics.

[8] Table sorted by views per document.

[9] Ossi Naukkarinen and Johanna Bragge (2016).

[10] Megan R. Brett, "Topic Modeling: A Basic Introduction,"*Journal of Digital Humanities*, 2(1), at http://journalofdigitalhumanities.org/2-1/topic-modeling-a-basic-introduction-by-megan-r-brett/ (10 April 2018).

[11] Xiaobing Sun, Xiangyue Liu, Jiajun Hu, Junwu Zhu, "Empirical Studies on the NLP Techniques for Source Code Data Preprocessing" in *Proceedings of the 2014 3rd International Workshop on Evidential Assessment of Software Technologies - EAST 2014* (New York, New York, USA: ACM Press), pp. 32–39. https://doi.org/10.1145/2627508.2627514 (10 April 2018). Software that was used for Topic Modelling is described here: Radim Řehůřek and Petr Sojka, "Gensim - Statistical Semantics in Python" in *EuroScipy* (Paris, 8/2011), pp. 25.–28. http://www.fi.muni.cz/usr/sojka/posters/rehurek-sojka-scipy2011.pdf (10 April 2018). Pre-processing was done with the Gensim python library and lemmatization with the Pattern python library.

[12] We took Wikipedia articles as a set of documents: one article equals one document. We excluded non-articles that relate to organization of content not providing the Wikipedia-related content.

[13] We took all words presented in English Wikipedia and removed most frequent words, which can be called as stop words and are too general to be used for separating documents, such as 'and,' 'but,' and anyway. Then small words, words that have three or less characters, were removed. We also removed words that are too rare and cannot be used to generalize the area. If a word occurs in less than twenty different Wikipedia articles, it was removed. Finally, we left the 100,000 most frequent words in our dictionary.

[14] All these steps were implemented within Gensim open source software package using Python programming language.

[15] Our aim was to compare different datasets. We finally discovered that the amount of 200 topics and a moderate alpha hyperparameter (0.05 in Gensim) give reliable results. We repeated the procedure until the most obvious topics and minimal ambiguity was found.

[16] We would like to thank the reviewers of*Contemporary Aesthetics* for their excellent suggestions that greatly improved the final version of the essay.