Article

Social Science Computer Review
1-25

DOI: 10.1177/0894439319871015
journals.sagepub.com/home/ssc
SAGE

The Market for Heritage: Evidence From eBay Using Natural Language Processing

Mark Altaweel[1]

Abstract

The trade in antiquities and cultural objects has proven difficult to understand and yet is highly dynamic. Currently, there are few computational tools that allow researchers to systematically understand the nature of the legal market, which can also potentially provide insights into the illegal market such as types of objects traded and countries trading antiquities. Online sales in antiquities and cultural objects are often unstructured data; relevant cultural affiliations, types, and materials for objects are important for distinguishing what might sell, but these data are rarely organized in a format that makes the quantification of sales a simple process. Additionally, sale locations and the total value of sales are relevant to understanding the focus and size of the market. These data all provide potentially useful insights into how the market in antiquities and cultural objects is developing. Based on this, this work presents the results of a machine learning approach using natural language processing and dictionary-based searches that investigate relatively low-end but high sales volume objects sold on eBay's U.S. site, where sales are often international, between October 2018 and May 2019. The use of named entity recognition, using a conditional random field approach, classifies objects based on the cultures in which they come from, what type of objects they are, and what the objects are made of. The results indicate that objects from the United Kingdom, affiliated with the Roman period, mostly constituting jewelry, and made of metals sell the most. Metal and jewelry objects, in fact, sold more than other object types. Other important countries for selling ancient and cultural objects include the United States, Thailand, Germany, and Cyprus. Some countries appear to more greatly sellspecific types of objects, such as Egypt being a leader in selling Islamic, terracotta, stone, and wood artifacts and Germany selling Viking/ early Medieval weapons. Overall, the approach and tool used demonstrate that it is possible to monitor the online antiquities and cultural objects market while potentially gaining useful insights into the market. The tool developed is provided as part of this work so that it can be applied for other cases and online sites, where it can be applied in real time or using historical data.

Keywords

natural language processing, machine learning, name entity recognition, conditional random fields, antiquities, heritage, culture, eBay, dictionaries

[1] University College London, United Kingdom

Corresponding Author:

Mark Altaweel, University College London, 31-34 Gordon Square, London WC1H 0PY, United Kingdom.
Email: tcrnma3@ucl.ac.uk

Recent conflict and upheaval in regions such as the Middle East have led to wide-scale looting of archaeological and cultural heritage sites as well as their destruction in places (Bauer, 2015; Brosché, Legnér, Kreutz, & Ijla, 2017). Increasingly, online sites and platforms, including social media, are used to sell antiquities, with the illicit market also taking advantage of these tools (Antiquities Trafficking and Heritage Anthropology Research [ATHAR], 2019; Hardy, 2016). Furthermore, knowing where demand for antiquities and cultural heritage objects, henceforth antiquities, might develop may require the monitoring of online sites that deal with the trade of antiquities. Such sites provide the potential to monitor demand on the types of objects sold and where different cultures are sold. There is also a potential link between the legal and illegal antiquities market. The legal trade of antiquities has been seen as one reason why the illegal trade and looting of heritage sites had developed more substantially, as demand from primarily Western state markets fuel and motivate looting of sites that are often poorly protected (Brodie, Kersel, & Tubb, 2006). The legal trade of antiquities can act as a proxy that provides insights as to what types of objects are in demand in the illegal antiquities market while also giving some clarity into how the antiquities market is unfolding. There simply is a limited number of public sources that allow the understanding of the size and scope of antiquities sold worldwide. Online sources, however, provide one potential avenue to understand this market.

There are challenges to understand the antiquities market; mainly, it is a dynamic one that changes as interests and opportunities for buying and selling develop in given regions (Brodie, 2012). One challenge has been to quantify the market and know if it is substantially changing as different cultures become of greater interest. The measurement of the antiquities market is difficult for the illegal trade, but in the realm of the legal antiquities trade, or at least legal sites used to trade antiquities, it is possible to better monitor the selling of antiquities by focusing on some of the major sites where they are sold. This includes eBay, where thousands of antiquities are bought and sold worldwide each month. Antiquities sold on eBay have been studied before (Brodie, 2015; Fay, 2013; Hardy, 2016), with the site seen as one barometer of at least low-end but high sales volume antiquities because the site makes it easy for almost any type of seller to sell worldwide. Nevertheless, as the descriptive data of objects on eBay are often unstructured, questions such as what cultures are of interest and what object types sell more frequently and even material characterization of objects sold are not always easily evident without a lot of manual data collection. However, these data could be crucial in determining the focus and consumer demand in the market at any given time. Furthermore, there are few tools that apply a systematic analysis of online sites and sales of antiquities. The methods and tools advanced here help address these identified gaps on what objects are sold and allow a systematic analysis of the results that show the broader market in antiquities as demonstrated on eBay.

This article applies a content analysis approach that looks at U.S. eBay sales data in order to determine where antiquities are sold, the dollar value of sales, the type of cultures affiliated with given objects sold, the types of objects sold, and the material characterization of objects sold. The method employs named entity recognition (NER) using a conditional random field approach (CRF; Lafferty, McCallum, & Pereira, 2001; Lee et al., 2006) in order to categorize terms and objects sold to determine evident patterns in sales. The method also deploys a simple dictionary to aid in finding terms and applies spell-checking on descriptions. The outputs demonstrate antiquity sale patterns covering October 21, 2018 to May 5, 2019. The outputs also show how unstructured data within sites such as eBay can be monitored for sales in antiquities. The tool used to conduct this analysis is provided to users to download and apply for their own work. The article begins by giving a brief background on the antiquities market and the use of natural language processing (NLP) as a way to classify text applicable for this type of research. The NER and related methods are then presented, with additional details made available in the code provided. The results are given, which are then followed by a discussion on the implication of the results. Future direction and limitations of this work are also provided.

Background

Online Cultural Heritage

There have been few works that have applied computational content analysis, including the application of NLP, to the trade of antiquities. The most comparable study is by Huffer and Graham (2017) that looked at social media data using hashtag searches in order to track the sale of human remains. The results showed that a community of small-scale traders are increasingly making higher profits using social media platforms as they are able to sell human remains. Greenland, Marrone, Topçuoğlu, and Vorderstrasse (2019) very recently applied a machine learning technique to estimate the value of antiquities sold so that one can better estimate how much in value looted archaeological sites might be yielding. In the realm of studying the past, social media data have been analyzed using NLP in the formation of identity politics and how archaeological and historical references are integrated in the debate revolving around Brexit, which is the process of the UK leaving the European Union (Bonacchi, Altaweel, & Krzyzanska, 2018). Overall, however, there has been very few large-scale, systematic assessments of the broader antiquities market using what is available on the web.

Most work looking at the antiquities market have been focused on applying qualitative or quantitative approaches that often utilize manual data collection to assess sales patterns within sites such as eBay or even well-known auction houses such as Sotheby's (Brodie, 2012, 2014a; Fay, 2013; Hardy, 2016). Scholars have described the legal sale of antiquities on online sites such as eBay as a gray area because provenance and discernible acquisition of objects are often not clear (Mackenzie & Yates, 2017). The use of eBay for selling antiquities is legal, but the initial acquisition of specific objects sold, including when and how objects may have entered a selling country, and rules regarding the transaction may not be. In other words, the platform is legal but the item sold may not always be legal. The market of antiquities is international, with Western states often seen as key markets for buying and selling antiquities (Barker, 2018; Bowman, 2008). With the nature of Internet sales, and often lack of clear provenance of objects, there is difficulty in knowing whether objects sold on sites such as eBay or others are always legal, while many objects are likely to be fake (Fay, 2011). Social media, and auction sites such as eBay, do appear to have made the sale of antiquities generally easier and have likely helped to fuel the growth in overall antiquities trading and likely looting of cultural sites (Barker, 2018). In fact, as many sites such as eBay offer lower end antiquities, one argument is they also broaden the market by making it easier to sell lower end antiquities to a wider audience. Based on this, sites such as eBay potentially provide a barometer of interest in antiquities, in particular low-end but also high sales volume, including types of objects and cultures that are likely to be of greater interest or demand by the broader public. In particular, eBay's reach is global, and its accessibility and brand mean that it enables a much wider community to conduct trade than the traditional marketplace of antiquities (Kotha & Basu, 2011). Furthermore, both smaller and larger scale vendors have been known to use eBay as a way to sell antiquities (Campbell, 2013). For many types of objects sold, even if they are forgeries, the large number of sales can provide insight into the market and wider developments in demand for antiquities, including what cultures sell, the objects that are of greater demand, and the materials these objects are made from.

NLP

One obstacle for studying the antiquities market is data are mostly unstructured or semistructured, where descriptive data provide key information but are not in a format that makes them amenable for direct statistical analyses when dealing with a large data set of sales, such as in the tens of thousands of sales or more. In different fields, NLP has been applied in assessing such cases using a variety of techniques, ranging from methods that deploy different machine learning, rule-based, and statistical procedures (Cambria & White, 2014; Kumar, 2012). Some popular research within NLP has included semantic pattern analyses, understanding semantic networks, and developing and understanding ontologies of language and classification. Often, a combination of such methods are used, although machine learning techniques generally predominate.

One method within NLP is NER, which helps organize terms in text into relevant informative categories (Konkol, Brychcín, & Konopík, 2015; Nadeau & Sekine, 2007). It is a form of text classification, generally employing some form of machine learning, that helps identify entities of interest, that is relevant terms, and then identifies those terms within predefined categories. For instance, a sentence such as "The president called on the UN to lift the embargo" could categorize *president* as a "person," the *UN* as an "organization," and *embargo* as an "action." While the terms are specific, the categories are intended to be more general but useful for understanding larger or more general patterns in text such as how often individuals are discussed or when organizations might be a key focus in the example given. Such methods, therefore, identify where in given text do relevant terms appear. Methods generally apply supervised, semi-supervised, and unsupervised techniques to classify text (Neelakantan & Collins, 2014; Nothman, Ringland, Radford, Murphy, & Curran, 2013; Ritter, Clark, Etzioni, & Etzioni, 2011). Increasingly, semi-supervised techniques and deep learning methods, such as convolutional neural networks (CNNs), are becoming prevalent, mainly because they help minimize time required for training data sets to be developed that are required in supervised classifications (Agerri & Rigau, 2016; Lample, Ballesteros, Subramanian, Kawakami, & Dyer, 2016; Nothman et al., 2013). Such techniques, for instance, attempt to cluster common features or groups of relevant terms that can be categorized, although accuracy may still be low relative to what might be required. Overall, a mixture of methodologies within NER might be needed to obtain a useful set of classifications for analyzed text.

Among techniques used, CRFs are one type of method that incorporates a supervised and semi-supervised structure to train classifiers for NER (Zafarian, Rokni, Khadivi, & Ghiasifard, 2015). The use of CRFs enables context of sentence structure and word orderings to inform on the potential classification of given terms (Lafferty et al., 2001). Terms are conditioned with respect to other terms by the fact that given text occur and relate to others in given word orders as undirected graphs with term probabilities based on Markov properties for the graph. Term order effectively provides the context and likelihood a given term may occur. The classifier is trained on term order, and relevant terms can be given a prescribed categorical designation useful for the researcher (e.g., the term *iron* is classified as a "metal"). The classifications are often subjective but can provide the level of generalization useful for given analyses such as whether metal objects are sold more frequently than nonmetal objects. Deploying CRFs can also be in combination with other techniques, such as using word dictionaries that might be developed for a given domain, to aid in task deployment (Wang, Jiang, Liu, He, & Hu, 2017). This also helps to minimize the required number of texts used for classifier training.

Method

Based on the need to better understand the antiquities market, and benefits given by NER-based methods, several steps are developed in this work, which will be discussed below. Figure 1 indicates

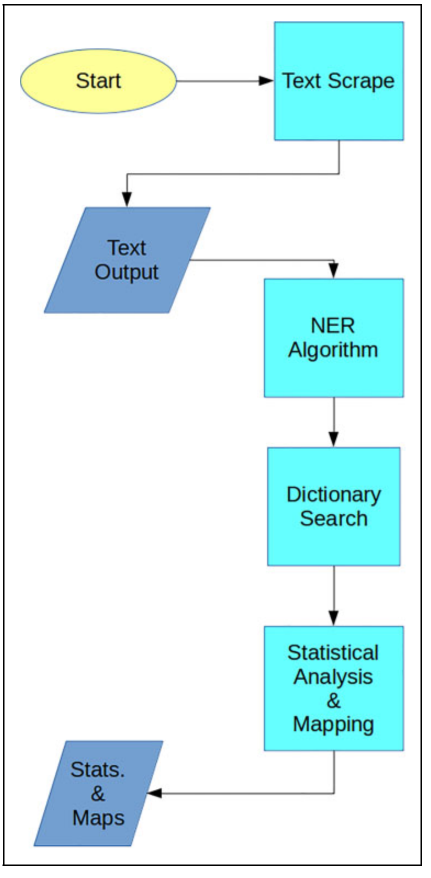

Figure 1. Workflow deployed in the applied methods.

the steps undertaken, which are text scraping, NER analysis, dictionary searches, and statistical analysis, where this also includes mapping of spatial data. The Python (Version 2.7+) tool used for NER, called eBay Scraper, is provided with this article. This tool conducts text scraping, spell-check on descriptions, NER analysis, and dictionary searches, including lemmatization of terms, that is grouping term inflections as part of the search. Additionally, a Java (Version 8+) tool, called NERProject, is incorporated within eBay Scraper, where the NER model creator is provided and the training text as well. Stanford's NLP libraries are used to create the NER model applied here (NLP Stanford, 2019). The scraping output and the raw NER and dictionary analysis outputs are also given along with the code provided. The tool does not apply statistical tests and precision and recall tests (see below), but the outputs from the tool can be applied in other software for such analyses (e.g., the R 3.5 statistical package). The tool does place output in a shapefile, which can be visualized in a geographic information system tool such as QGIS 3.8 or ArcGIS 10+. This provided material also includes a more detailed explanation of the Python and Java used for the applied steps and what the applied classes and modules do in different processes. The description also incorporates a complete flow diagram of the NER training algorithm, scraping, and NER/dictionary analysis described below. Input data utilized, with the exception of the NER model which is too large, are provided. The NER model can be recreated using the training text in NERProject. For simplicity and space limitations, the methods used are summarized below.

Text Scraping

To understand eBay in relation to the antiquities market, the first step is to obtain sales data from the site. The focus here is on the main U.S. eBay site (eBay.com) rather than other countries since this is the biggest market. It also focuses on English-speaking sellers who often sell internationally; this means you also have may non-U.S. listings selling to a global audience. After initially exploring the eBay API (2019) and scraping links within eBay.com, older completed and sold auctions are found and obtained using the following link: https://www.ebay.com/sch/37903/i.html?_sop=13&_sadis=15&LH_Auction=1&LH_Complete=1&LH_Sold=1&_stpos=90278-4805&_from=R40&_nkw=%27+Antiquities. This enabled the analysis to extend the temporal coverage of this study and obtain more data to analyze than what was available on eBay API. Other parts of eBay and non-U.S. sites are used to sell antiquities, but these are not part of searches because these are not the main sites for buying antiquities internationally.

Once relevant antiquities are located, one can search under different antiquities that have been sold, including Celtic, Roman, Viking, Greek, Neolithic, and various other objects. Although objects are listed within their cultures (e.g., Roman), it is found many are mislabeled (e.g., Scythian objects in a Viking category). Therefore, part of the approach taken is to classify the cultures of the objects based on descriptive data, rather than the category they are found in, as that is found to be more accurate. Overall, a total of 54,717 unique sales, covering the period analyzed, are identified, with objects coming from cultures in the Old and New Worlds. Data within eBay are semistructured, with information such as date sold, sale price, and location of the seller given in defined HTML tags, but data within the description of an object sold are unstructured, with descriptions containing information on culture, type of object, and the object's material composition (Figure 2). Once sales data are scraped, they are placed into .csv files, with different subsites within the primary antiquities sales site scraped to produce the total unique sales. Images of the objects are not part of the scraping utilized, although this is a potential option for future work.

NER

After text scraping, the next step included NER analysis. The NER method deployed incorporates a CRF model created using Stanford's Named Entity Recognizer written in Java (Finkel, Grenager, & Manning, 2005; Manning et al., 2014; NLP Stanford, 2019). Training texts are utilized; terms that fit identified categories in relation to cultural affiliation (e.g., "Roman," "Greek,"), object types (e.g., "weapon," "tool," "coin,"), and material composition (e.g., "metal," "glass," "wood") for given sold objects are used in the training texts. Table 1 lists all the cultures, object types, and material composition used to categorize terms, with a total of 45 categories. It should be noted that British spelling (e.g., "jewellery" rather than "jewelry") is used in the applied application, although the methods can search variations in spelling such as British or U.S. spellings. For simplicity, US spellings are used here. Generally, yt is not practical to be too specific in how objects are classified, as resulting patterns might be unclear. However, general category designations (e.g., "jewelery" used to indicate objects such as *rings* or *bracelets*) are used. In addition, common cultures and objects found in the training texts are used to guide the choice of category names. For instance, Roman objects are common in training texts, resulting in "Roman" as a categorical term. However, other cultures and objects, for example Scythian and Bactrian or South American, are categorized in more broad categories such as cultures from "Central Asia" or "America," respectively. This is also true for objects, such as the category "tool" being general to include items such as *knives* or *adzes* as some examples. Table 2 provides some examples of how terms and categories are used for objects, cultures, and material types. In the NER method, if a term fits within the category determined, then that category is designated for the text provided.

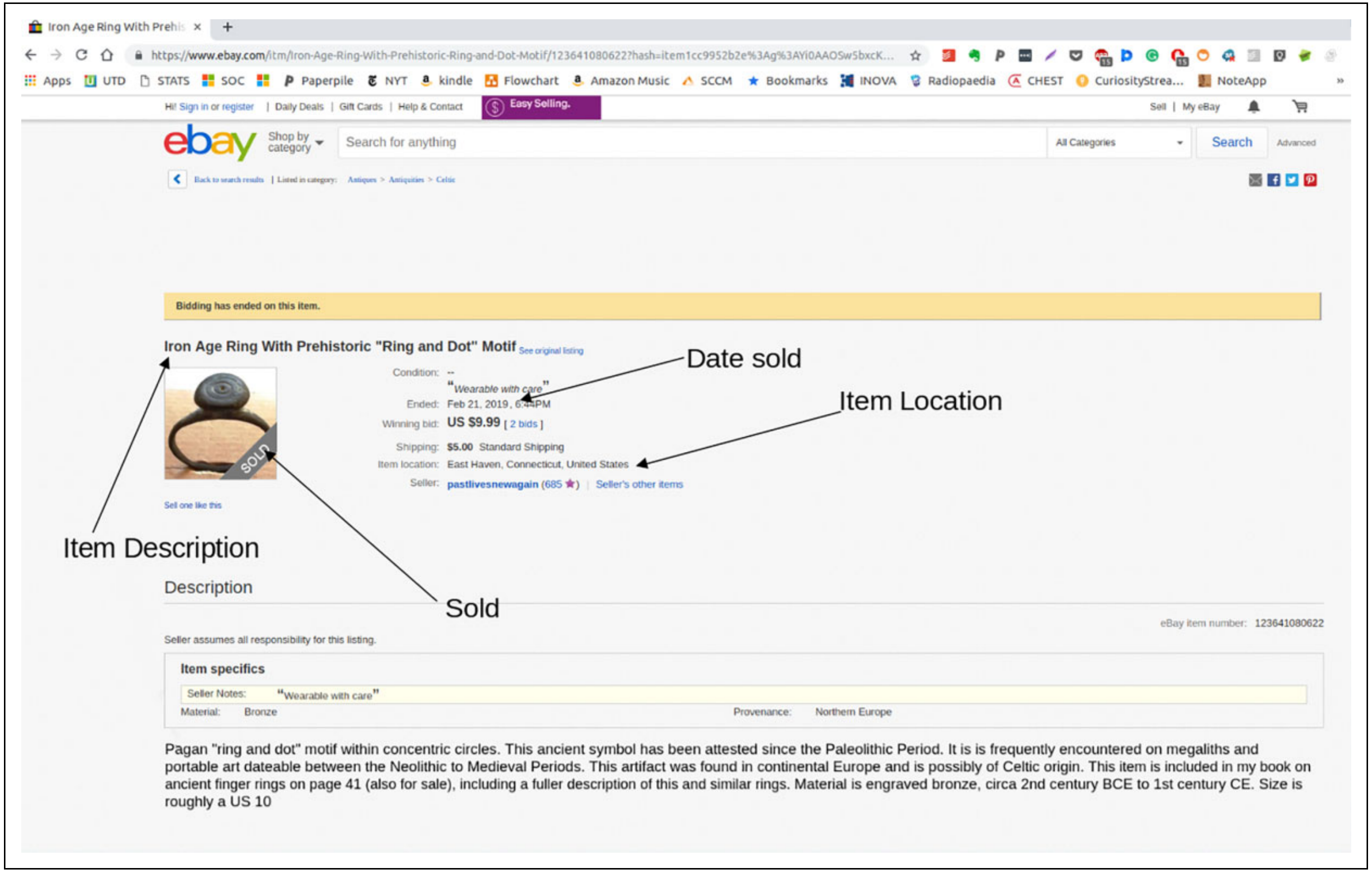

Figure 2. Example of a scraped eBay sale.

Table 1. Categories for Cultures, Object Types, and Materials for Antiquities as Determined by the Named Entity Recognition Analysis.

Cultures	Object Types	Materials
Africa	Clothing	Bone
America	Coin	Glass
Bronze Age	Decoration	Ivory
Buddhist	Household	Leather
Cambodian	Jewelry	Metal
Celtic	Mask	Papyrus
Central Asia	Religious	Terracotta
China	Statue	Stone
Egyptian	Text	Wood
European	Tool	
Greek	Painting	
India	Portrait	
Iron Age	Vessel	
Islamic	Weapon	
Japan		
Medieval		
Mongol		
Near East		
Prehistoric		
Roman		
Russian		
Viking		

Table 2. Example Showing Types of Categories (i.e., Culture, Object, or Material) and Example Terms Used for These Categories Based on the Named Entity Recognition Applied.

Type	Category	Term Example
Culture	Roman	Roman
Culture	Celtic	Celtic
“ “”	“ ”	Seltic
Culture	America	Maya
“ ”	“ ”	Aztec
“ ”		Native American
Object	Jewelry	Necklace
“ ”	“ ”	Ring
“ ”	“ ”	Pendant
Object	Household	Mirror
“ ”	“ ”	Box
Object	Tool	Needle
“ ”	“ ”	Wheel
Object	Statue	Statue
“ ”	“ ”	Figurine
Material	Glass	Glass
Material	Terracotta	Clay
		Ceramic

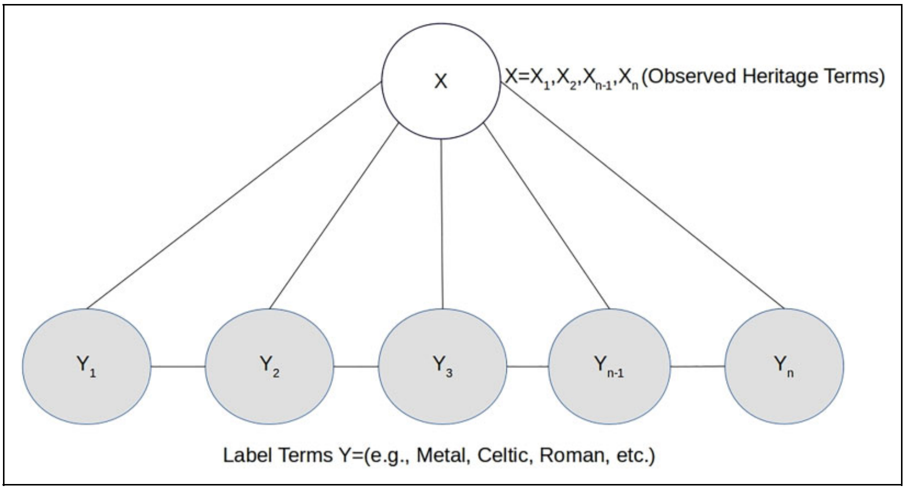

Figure 3. Summary graphic showing a conditional random field model where the shaded *Y* labels are generated by the model.

Multiple categories are possible for terms, including multiple cultures, object types, and material composition. In part, this reflected the sometimes ambiguous and overlapping nature of terms in archaeology. For instance, *axe* can be a "tool" or "weapon." In some cases, the description of the object is used to determine which is a better categorization, such as a description involving war (i.e., "weapon") or for likely common use (i.e., "tool"), helping to distinguish whether one category is a better fit.

Ambiguity, based on subjective choices for categorical designations and their meaning, is possible, and the categories are prone to confusion or disagreement among scholars. Therefore, definitions of categories used are given in the data and code provided for download; the category names are shared with the dictionary used and discussed below. Users could change the category names and how categories are designated for texts used for training, which is encouraged if better designations are found, enabling modification of the categorization in the analysis once it is executed. The CRF model creator using Stanford's NER tools is provided as part of the eBay Scraper tool and used here in the NER classification. The CRF model enables a probability of a given label to be determined based on the summarized function:

$$p(y \vee \mathrm{x},\ \lambda) = \frac{1}{Z}(x)\exp\left(\sum\nolimits_j \lambda_j F_j(\mathrm{y},\ \mathrm{x})\right), \tag{1}$$

where the y label sequence, that is identified entities, is based on the observation sequence x (i.e., term sequence such as input terms in object descriptions), with F a feature function that is based on an empirical distribution from training data, Z is the normalization factor, and λ is an estimated parameter from training (Finkel et al., 2005). The training data consist of more than 5,000 lines of sales data text, which are run using 99 or more training iterations to build the model. One benefit of the NER is it assists in differentiating terms that could be ambiguous. For instance, *iron* would normally mean a type of "metal"; however, *Iron Age* would denote a period. The CRF approach attempted to use the context, such as *Age* affiliated with *Iron*, to help denote that this is a period rather than a material. One can think of a CRF as an undirected graph consisting of vertices and edges for terms based on observed and output term sequences. This is summarized in Figure 3. After completing the training classification, the model used in the NER analysis is created and deployed to be called by the application with each line of scraped text analyzed.

Dictionary Searches

After attempting to categorize texts using the NER approach, the next step is to apply regular expression searches using keywords from a predefined corpus dictionary in eBay Scraper, similar to other approaches (Manning, Raghavan, & Schütze, 2008; Wang et al., 2017). This also lemmatized words, meaning that word inflections are searched. Python's Natural Language Toolkit library is used for processing text, including lemmatization and tokenization. The dictionary approach is applied because relevant terms could possibly appear but may not be easily identified in the NER approach. For instance, a less common term could be *diadem*, which is classified as a type of "jewelry" (i.e., a type of crown). Less common terms may not have been sufficiently applied in training the CRF model created. Using the same types of categorization as the NER approach, dictionary terms relate to cultures, type of objects, and material composition of objects. A few terms, such as *seals*, are placed in multiple categories (e.g., in the case of *seals* in "tool" and "jewelry"), as they can be utilized in multiple ways. The dictionary also includes common alternative spellings of terms (e.g., *arrow head* instead of *arrowhead*); the use of a spellchecker is applied to supplement this approach. As stated, due to possible confusion of terminology and subjectiveness in categorization, definitions are provided as well for categorization (see previous section). The benefit of an applied dictionary is it allowed direct terms to be searched within the text. However, ideally there would be no need for a dictionary, but the dictionary is useful when the NER model is not sufficient for categorizing terms encountered within the searched data. In effect, this indicates a larger training data set would need to be used to completely remove the need for a dictionary. Applying a dictionary potentially enables an increase in accuracy and retrieval of relevant information without having to have such a large training set that would require substantial development time. However, it also potentially introduces errors, as dictionaries do not take incorporate context of terms. Testing of the effectiveness of the approach is, therefore, needed.

Statistical Analysis and Mapping

The NER/dictionary results are analyzed using different statistical methods including distribution and multiple correlation analyses. As the location of sales are indicated, these are used to map where sales took place for the different categories (i.e., culture, object type, and material) assessed. This is done by outputting results to a shapefile.

For validation of the NER and dictionary methods applied, the text classification results are evaluated using precession and recall tests. This included scoring the evaluated text to demonstrate the accuracy and sensitivity of the approach; the F-score (or F_1 score) represents a harmonic mean value that represents both precision and recall (Goutte & Gaussier, 2005; Powers, 2011). Precision measures information given based on the categorization designated for the text, while recall reflects the fraction of relevant documents retrieved successfully. Precision reflects the accuracy of providing correct information, while recall allows the sensitivity of the approach to be measured in retrieving relevant information. In this case, the automated classification is checked against an independent evaluation that determined precision and recall. A random selection of 500 sales and descriptions are retrieved and checked for their F_1 score. The basic step for precision (P) is:

$$\mathrm{P} = \frac{T_p}{T_p + F_p}, \tag{2}$$

where T_p is true-positive results and F_p is false positives. For recall (R), it is:

$$\mathrm{R} = \frac{T_p}{T_p + F_n}, \tag{3}$$

Table 3. Precision and Recall Results.

P	R	F_1
.97	.92	.94

where F_n represents false negatives. Calculating the F_1 score, that is the harmonic mean of precision and recall, yields:

$$F_1 = 2\frac{\mathrm{P} \times \mathrm{R}}{\mathrm{P} + \mathrm{R}}. \tag{4}$$

This evaluation is applied to the total of the relevant categories created, which are the cultural, object type, and material composition categories for items sold. This results in 1,500 tested categories for the 500 randomly selected sale items.

Results

Precision–Recall Test

The first output is testing the precision and recall of the data from the NER/dictionary analysis. Table 3 provides the P, R, and F_1 scores that are based on 500 randomly selected sales data from the overall analyzed data, with each of the three categories (culture, object type, and material) analyzed. The results are deemed to be sufficiently precise and sensitive in capturing relevant information for the intended purpose of this work. The file used for the test is provided with the data and code given. Overall, over 90% of the total categorization is accomplished by the NER, while the remainder is accomplished using the dictionary.

Summary Results

Summary results are provided to give an overview of the key results from the content analysis approach. Table 4 shows some general statistics for sales data for cultures, object types, and materials. For the entire corpus, total sales for the period covered are US$2,556,092. Figure 4A, showing a log–log plot, indicates sales had truncated power law properties; the distribution cannot be called Gaussian based on a Cullen and Frey (1999) test, with skewness near -1 and kurtosis at more than 7 (Figure 4B). Roman (US$873,809) antiquities are by far the highest total value for cultures sold, with Egyptian (US$357,256), and unknown (or undetermined, US$299,434) cultures the next highest monetary value in sales. The next highest are Viking (US$273,632) and Near East (US$232,599) cultures. Islamic (US$156) cultures have the highest mean sale price, while cultures from India (US$46) have the highest median. Jewelry (US$869,316), religious (US$379,606), statues (US$344,780), which included figurines, and vessels (US$318,361) are the first through fourth, respectively, in the object type categories for total sales. Masks (US$150; US$59) though have the highest mean and median values, respectively. Metals (US$1,359,629), unknown (or undetermined, US$613,424), stone (US$356,993), and terracotta (US$226,339) are the highest selling materials, respectively. Nevertheless, it is more perishable materials that have higher median and mean sale prices, with wood (US$146) the highest mean and papyrus (US$88) the highest median. As is evident in the standard deviations in Table 4, many categories have a wide spread in the value of sales.

Looking at the object types and material categories based on some of the top cultures that sold, specifically Roman, Egyptian, Viking, and Near East, it is possible to demonstrate how different types of objects and material types sold within these cultures. Figure 5 uses the natural log of sales for listed object, which facilitated visualization, and material types for given cultures, where it

Table 4. Summary Results of eBay Antiquity Sales for All Objects, Including Different Cultures, Object Types, and Materials.[a]

Type	Total (US$)	Mean	Median	*SD*	Min	Max
Total	US$2,556,092.00	US$46.7	US$21.62	US$119.91	US$0.01	US$6,000
Roman	US$873,809.24	US$46.94	US$21.50	US$98.52	US$0.01	US$2,941.85
Egyptian	US$357,256.92	US$47.93	US$31.27	US$77.31	US$0.01	US$3,562.50
Unknown culture	US$299,434.10	US$48.54	US$19.00	US$135.62	US$0.01	US$6,000.00
Viking	US$273,632.65	US$41.30	US$20.00	US$94.94	US$0.01	US$3,051.55
Near East	US$232,599.42	US$45.94	US$18.92	US$150.02	US$0.95	US$5,249.00
Greek	US$178,332.60	US$49.28	US$26.49	US$83.47	US$0.95	US$1,300.00
Islamic	US$152,316.15	US$156.06	US$28.92	US$481.30	US$0.01	US$5,249.00
Medieval	US$139,970.33	US$30.09	US$17.58	US$67.15	US$0.01	US$2,332.17
Prehistoric	US$92,990.02	US$39.30	US$16.55	US$104.76	US$0.95	US$2,738.00
Celtic	US$56,610.23	US$38.93	US$19.00	US$85.05	US$0.74	US$1,800.00
Russian	US$40,549.33	US$52.94	US$21.99	US$129.78	US$0.99	US$2,570.00
Central Asia	US$33,027.28	US$86.69	US$17.00	US$285.92	US$0.75	US$2,425.00
India	US$32,960.39	US$53.08	US$46.17	US$40.97	US$0.99	US$419.81
Bronze Age	US$27,516.18	US$57.81	US$34.00	US$77.90	US$0.99	US$595.00
European	US$26,229.06	US$48.75	US$21.02	US$158.18	US$0.75	US$3,300.00
America	US$25,325.77	US$102.53	US$21.50	US$320.05	US$0.99	US$3,300.00
China	US$14,915.66	US$35.18	US$20.25	US$69.85	US$1.28	US$828.88
Africa	US$6,758.26	US$22.30	US$16.40	US$24.39	US$0.99	US$224.72
Iron Age	US$2,976.76	US$41.93	US$23.47	US$49.77	US$0.95	US$276.00
Japan	US$2,498.83	US$59.50	US$41.50	US$61.12	US$5.50	US$330.00
Cambodian	US$1,317.62	US$48.80	US$26.05	US$65.06	US$5.50	US$295.00
Buddhist	US$7.95	US$7.95	US$7.95	US$0.00	US$7.95	US$7.95
Jewelry	US$869,315.72	US$36.00	US$18.50	US$79.05	US$0.01	US$3,300.00
Religious	US$379,606.37	US$41.27	US$24.52	US$82.48	US$0.01	US$4,750.00
Statue	US$344,780.61	US$62.57	US$41.86	US$87.99	US$0.01	US$2,497.55
Vessel	US$318,361.25	US$72.57	US$38.00	US$208.91	US$0.01	US$5,249.00
Unknown object	US$309,929.87	US$67.20	US$22.50	US$194.37	US$0.01	US$6,000.00
Tool	US$282,983.99	US$39.76	US$20.00	US$88.94	US$0.01	US$3,051.55
Weapon	US$273,143.53	US$52.40	US$20.49	US$140.35	US$0.01	US$2,738.00
Text	US$137,261.74	US$59.04	US$33.11	US$137.83	US$0.01	US$4,750.00
Clothing	US$95,277.19	US$27.81	US$15.50	US$77.67	US$0.01	US$3,150.00
Household	US$75,254.49	US$39.94	US$19.20	US$100.17	US$0.01	US$1,775.00
Decoration	US$56,327.67	US$47.69	US$21.19	US$83.78	US$0.01	US$1,115.10
Coin	US$53,094.11	US$34.32	US$16.00	US$65.08	US$0.01	US$950.00
Mask	US$14,935.68	US$150.87	US$59.00	US$297.64	US$0.99	US$2,332.17
Metal	US$1,359,629.35	US$49.30	US$21.19	US$133.33	US$0.01	US$5,249.00
Unknown material	US$613,424.55	US$46.44	US$20.99	US$129.49	US$0.01	US$6,000.00
Stone	US$356,993.44	US$44.29	US$19.50	US$97.35	US$0.01	US$3,300.00
Terracotta	US$226,339.25	US$49.81	US$33.56	US$64.64	US$0.01	US$1,115.10
Glass	US$130,068.50	US$33.98	US$21.11	US$44.64	US$0.01	US$662.18
Wood	US$19,467.51	US$146.37	US$68.87	US$302.75	US$0.01	US$2,500.00
Bone	US$11,168.94	US$38.78	US$24.56	US$50.54	US$0.01	US$410.00
Papyrus	US$3,524.47	US$95.26	US$88.73	US$56.86	US$1.78	US$300.00
Leather	US$1,653.20	US$19.92	US$0.99	US$66.91	US$0.01	US$461.00
Ivory	US$54.00	US$54.00	US$54.00	US$0.00	US$54.00	US$54.00

[a]The column *SD* is standard deviation. Highlighted rows indicate different sets of measures (i.e., total sales, cultural categories, object types, and material composition of objects).

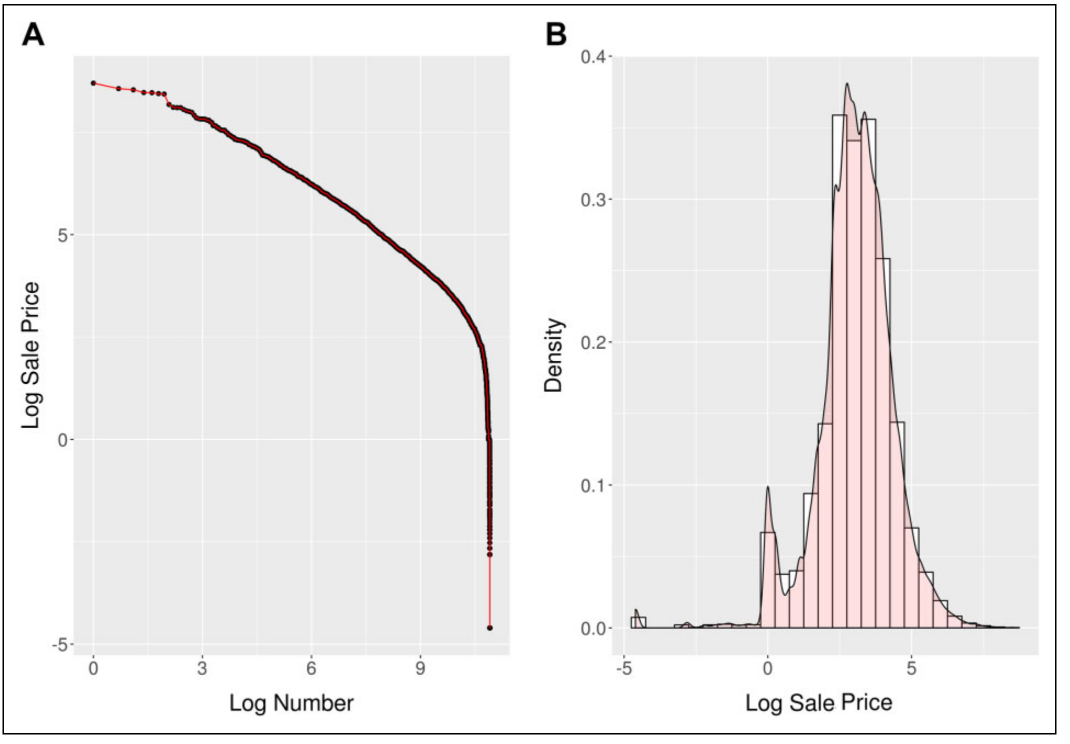

Figure 4. Natural log values of sales (A) and sales distribution (B).

shows the distribution of sales. Looking at Roman culture, the top selling culture, jewelry (US$392,055), vessels (US$134,399), statues (US$92,528), and tools (US$87,301) are the top selling objects. Top materials sold include metals (US$625,572), unknown (?; US$97,512), stone (US$93,756), and glass (US$82,858). A metal and vessel object is the top selling (US$2,941) object. Leather has the lowest mean and median sale price (US$9). Interestingly, weapons are also relatively low in median sale price (US$15), although it has a higher mean sales price (US$30). Wood objects have the highest mean and median price (US$112 and US$48, respectively) but are relatively rare. Egyptian objects indicate that statues (US$160,455), religious (US$105,721), text (US$75,262), and jewelry (US$72,144) types have the highest sales, respectively. For materials, unknown (?; US$171,942), terracotta (US$99,040), stone (US$51,266), and metals (US$31,285) are the top materials. Perhaps unsurprisingly, papyrus materials have relatively high median and mean sales prices (US$99 and US$88, respectively) but are relatively rare. Masks (US$160) have the highest mean sale price, although the highest median is decorative objects (US$95). An unknown (or undetermined) material is the highest selling object (US$3,562). For Viking (i.e., Norse and Dane cultures), top selling objects include jewelry (US$104,743), religious (US$79,653), weapons (US$46,465), and tools (US$39,970). Metal (US$183,094), unknown (US$76,398), stone (US$7,382), and glass (US$6,074) are the top materials. Vessels have the highest mean (US$104) and median (US$73) prices, although statues are not far behind (US$103 and US$67 mean and median prices, respectively). A tool object with an unknown material has the highest sales price (US$3,051). In Near East objects, jewelry (US$87,210), vessels (US$42,742), and weapons (US$30,586) are top selling objects, while metal (US$98,136), unknown material (US$78,618), glass (US$28,034), and stone (US$27,343) are the highest selling material types, respectively. The most expensive item is a vessel metal object (US$5,249). The cheapest are US$0.01 for different

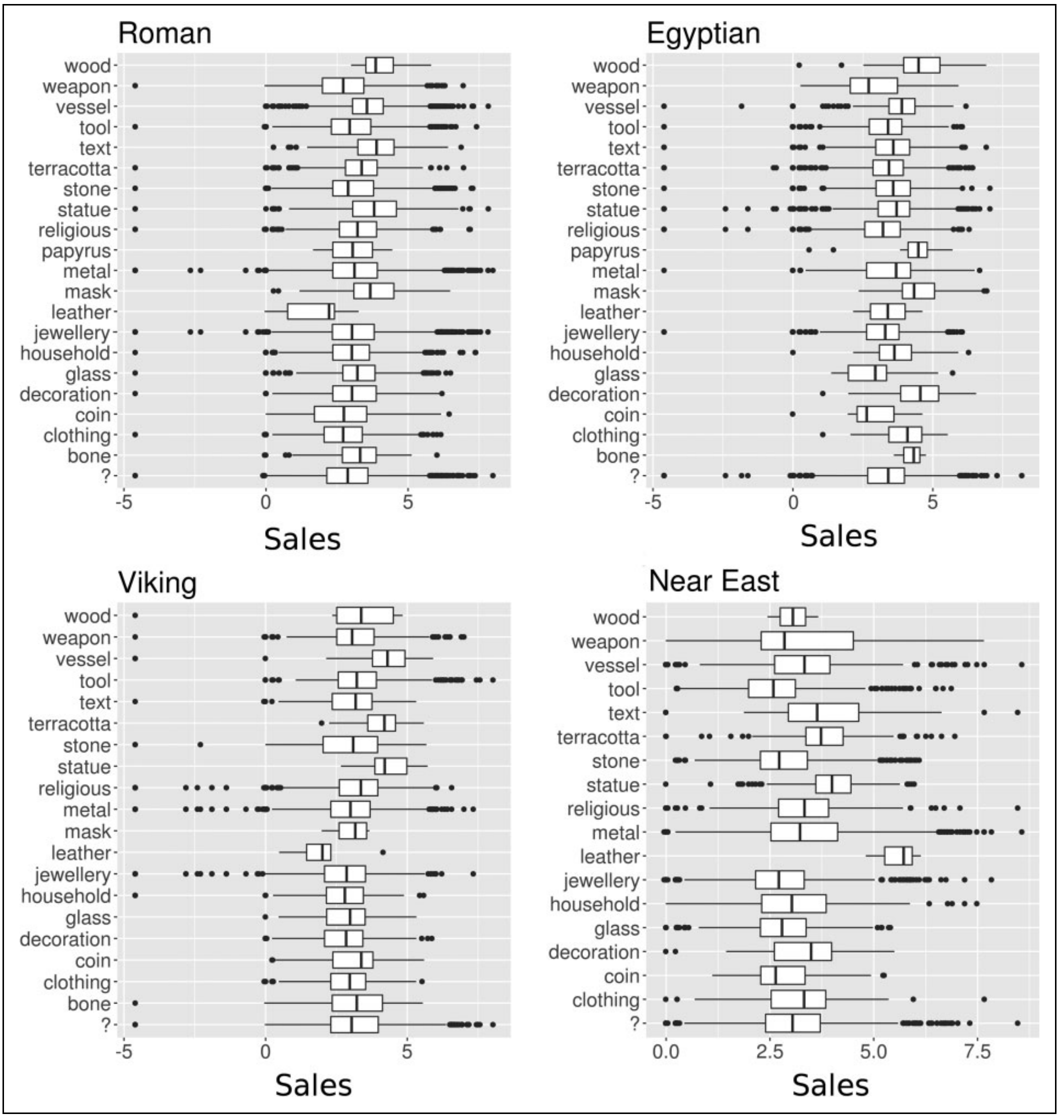

Figure 5. Box whiskers plot of natural log sale values for combined object types and materials for different top selling cultures.

categories. Leather objects, while few in number, have the highest mean and median for sales (US$296 and US$305, respectively). Text objects have a relatively high mean sale price (US$179), although the median is US$38, with a standard deviation of US$534, as this category has generally few sales. Tools and coins have the lowest median at US$13, with means at US$24 and US$34, respectively. Glass objects have the lowest mean value (US$22; US$16 for median).

Country-Level Results

One way to organize the sales data is by the country in which objects are sold from. Sales did not always have clear information on where they are sold, but of the total +US$2.5 million in sales,

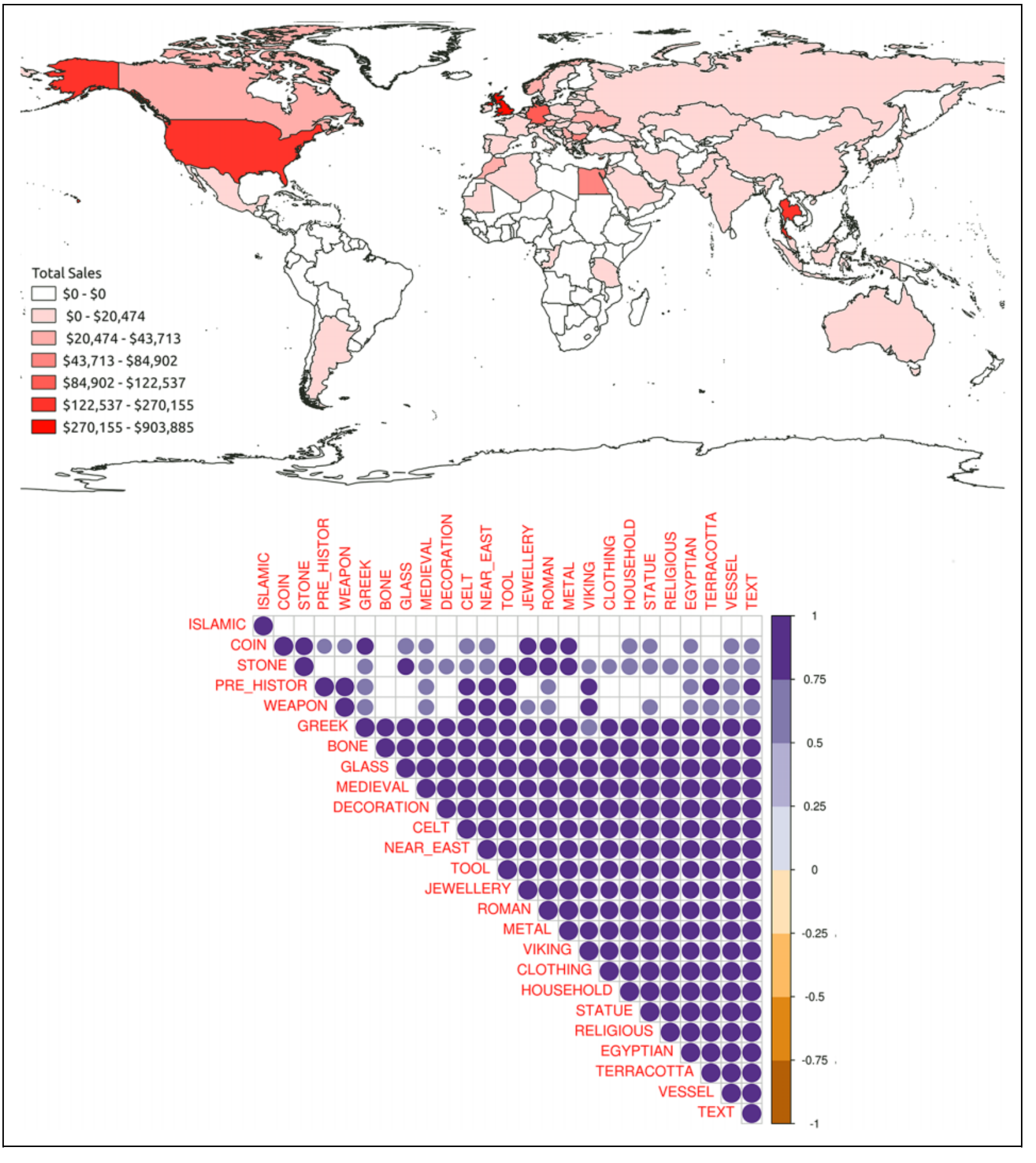

Figure 6. Countries selling antiquities and a correlation analysis for the top 30 selling countries looking at cultures, object types, and materials sold.

US$2,162,983 are affiliated with a given country. This gives us a good sample of where objects are being sold. Figure 6 shows countries in terms of total antiquity sales. The figure also shows a multiple correlation analysis of cultures, objects, and material types among the top 30 countries in sales. This shows that for the top countries, several cultures, object types, and materials have a strong positive correlation, where the correlation matrix is adjusted using the Holm–Bonferroni method (1979) method, indicating a somewhat broad coverage in these three categories for countries that sell relatively many objects. For instance, Greek, Near East, Viking, Roman, Celt, and Egyptian objects have strong positive correlations with jewelry, decoration, household, clothing, tool, religious, and statue

Table 5. Top Ten Selling Countries of Antiquities, Including Their Top Selling Cultures, Items, and Materials.

Country	Total Sales	Top Culture	Top Object	Top Material
United Kingdom	US$903,885.21	Roman	Jewelry	Metal
United States	US$270,154.55	Roman	Jewelry	Metal
Thailand	US$204,886.11	Roman	Jewelry	Metal
Germany	US$122,537.04	Viking	Weapon	Stone
Cyprus	US$113,470.18	Roman	Jewelry	Metal
Egypt	US$84,902.14	Egyptian	Vessels	Metal
Bulgaria	US$65,320.02	Roman	Jewelry	Metal
Ukraine	US$43,713.14	Viking	Weapon	Metal
Morocco	US$37,464.23	Viking	Jewelry	Metal
Austria	US$36,650.43	Roman	Jewelry	Metal

objects. They also have strong correlations with metal, glass, and terracotta objects. On the other hand, Islamic cultures, coins, prehistoric, and weapons show generally weaker broad correlations among cultures, object types, and materials. Overall, the strong correlations suggest that many top selling countries sell a variety of different cultures, object types, and materials.

There are, however, noticeable patterns specific to countries that allow one to identify some key differences. Table 5 depicts the top 10 countries selling antiquities including the top cultures, object types, and materials sold in those countries. For most of these countries, Roman, jewelry and metal objects are the most common objects sold. Germany, Ukraine, and Morocco sell Viking-related cultures as their top cultures, while Egypt sells more Egyptian objects in its sales. In two cases (Germany and Ukraine), weapons are the top items and in one case (Egypt) vessels. Only one country (Germany) sells more stone objects than metals. Only two of the countries (United States and Thailand) are not in Europe or North Africa. Total sales in Europe and North Africa are nearly US$1.6 million of the US$2.16 million value, with North America (US$300,885) the other major region for sales. While the United Kingdom has, by far, the highest sales at over US$900,000, Cyprus stands out as a relatively small country that has the fifth highest overall sales in antiquities. Figure 7 shows top selling cultures in European and North African countries and their total sales. Roman objects did dominate sales, globally and in Europe, but there is also a pattern of northern European countries selling prehistoric and Viking-related cultures, while in North Africa and the Middle East, Islamic cultures are prevalent in sales. This suggests some spatial pattern in sales is evident where countries that have hosted given cultures of interest did have relatively higher sales for those cultures. In other words, many objects sold likely originate from the country selling them, although clearly this is also not always the case such as Roman objects selling in North America.

Although Roman, jewelry, and metal objects, primarily from Europe, dominate sales, more rare items can be categorized in the global sales data. Figure 8 shows cultures that include those from Central Asia, China, Celtic, and Islamic regions and the countries where they are sold. Germany is the top selling country for Central Asia cultures (US$19,881), the United Kingdom is the top selling Chinese (US$7,471) and Celtic cultures (US$20,032), while Egypt is the top selling for Islamic cultures (US$29,376). Results show that less common cultures sell mainly in North America and Europe, but, similar to before, a few countries where some cultures were/are located also sell those cultures to a greater extent (e.g., Islamic in Egypt).

Figures 9 and 10 display different common objects, outside of jewelry, and material types, outside of metals, respectively. Vessels sell most in the United Kingdom (US$143,718) and the United States (US$39,715). For tools, the United Kingdom (US$83,716) and Germany (US$37,735) are the two leading sellers. Statues, which include figurines, and religious objects sell the most in the United Kingdom (US$178,742 and US$159,746, respectively) and the United States (US$31,127

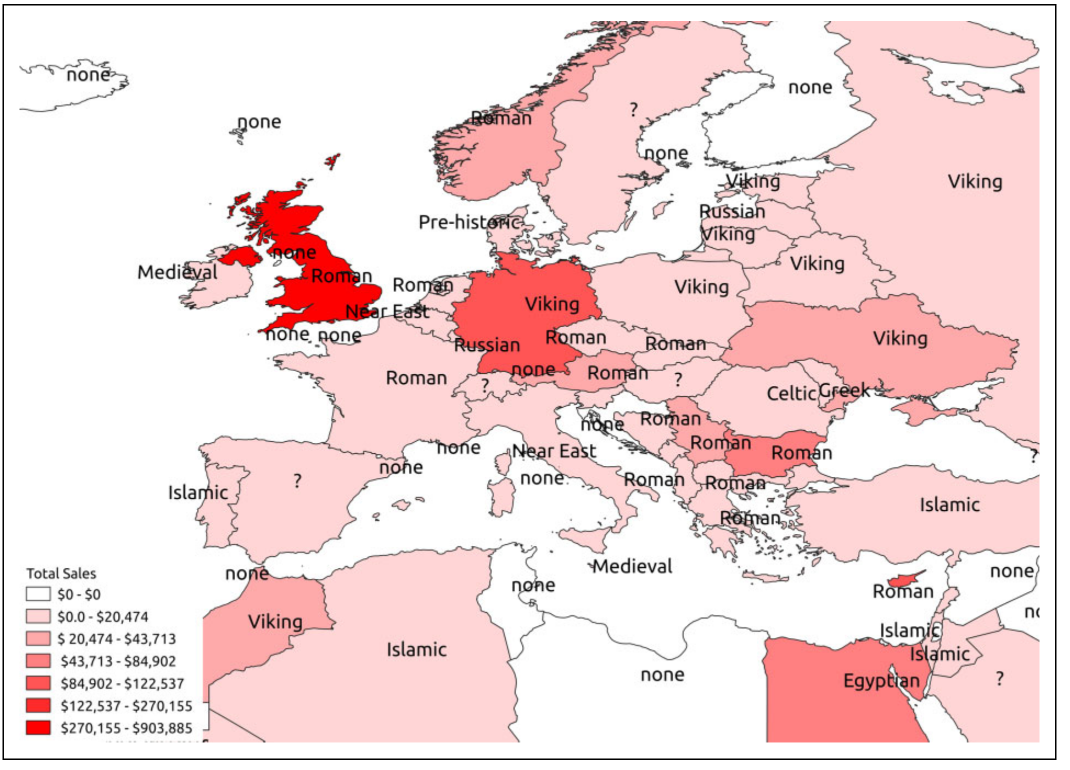

Figure 7. European and African states depicted along with total sales and top selling cultures.

and US$30,105, respectively). It is also clear that Egypt is a prominent seller of antiquities that included vessels, statues, and religious objects, where it is the third highest seller for those three object types. For glass, terracotta, and wood objects, the United Kingdom and the United States are again the top two countries, respectively; Thailand is the third most common seller of glass (US$13,532) objects and top seller of stone objects (US$100,972). Cyprus is the third most common seller of terracotta (US$15,821). Egypt is the third highest selling country for wood objects (US$4,021).

Discussion

Insight From Results

This article presented a content analysis approach using NER that is assisted by dictionary searches identifying relevant culture, object, and material types for antiquities sold on eBay. The work demonstrates an automated and machine learning method to monitor the sale of antiquities so that sale patterns could be more informative. Previous work has looked at the sale of antiquities in different countries (e.g., Brodie, 2014b; Bowman, 2008); this work represents a novel, systematic approach to studying low-end but high sales volume antiquities on eBay, with the tool potentially extensible to other sites, while also determining where such objects are sold. Cultures such as Roman and Egyptian objects are among the top cultures sold on eBay, while Europe, some countries in North Africa, and the United States are among the top sellers. There is considerable variation in object sales and values, although most objects that sell are relatively low-value objects (i.e., less US$100). Some cultures have relatively higher values for their sales (e.g., those from India), with

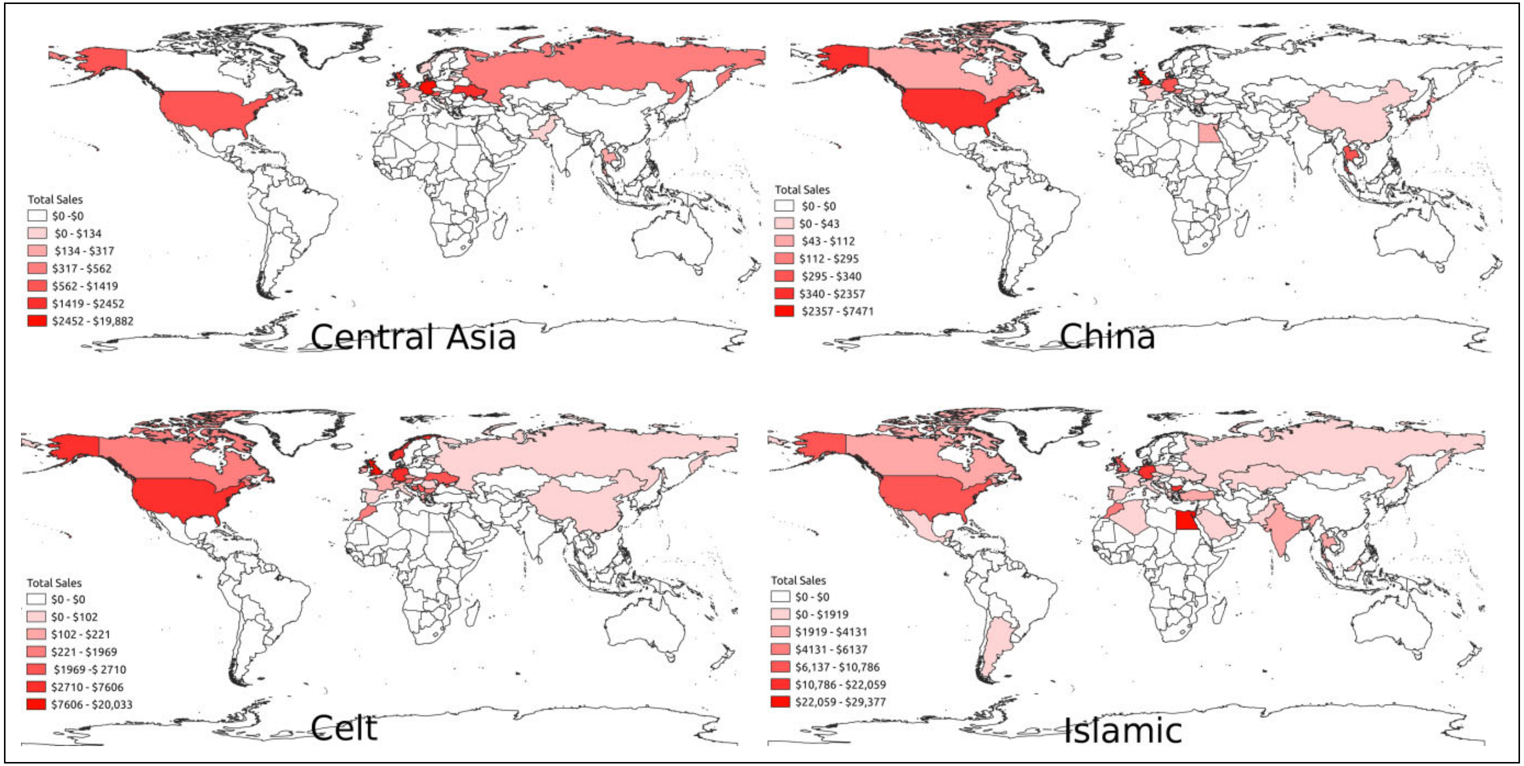

Figure 8. Four cultures and cultural regions (Central Asia, China, Celtic, and Islamic) and where they sold.

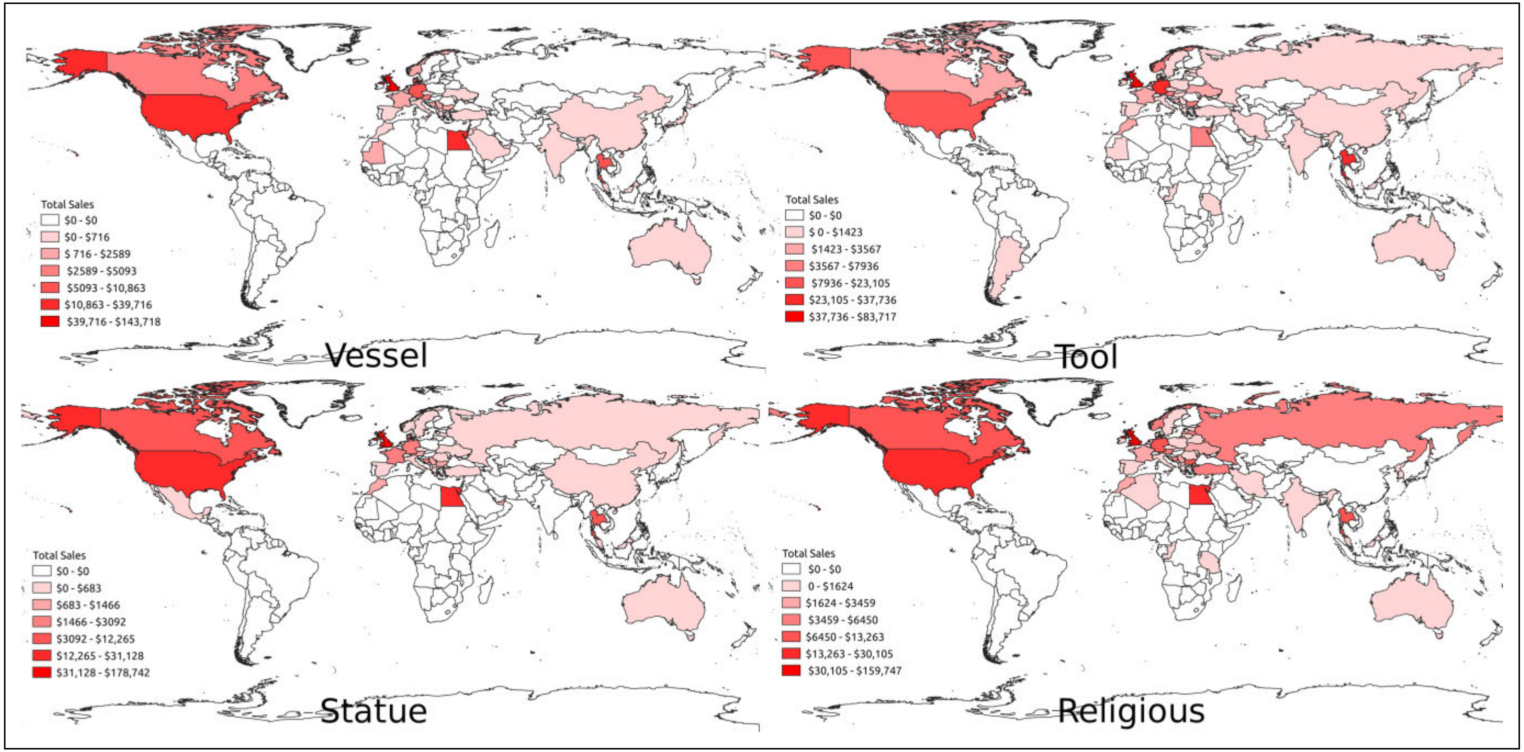

Figure 9. Type of objects sold in different countries.

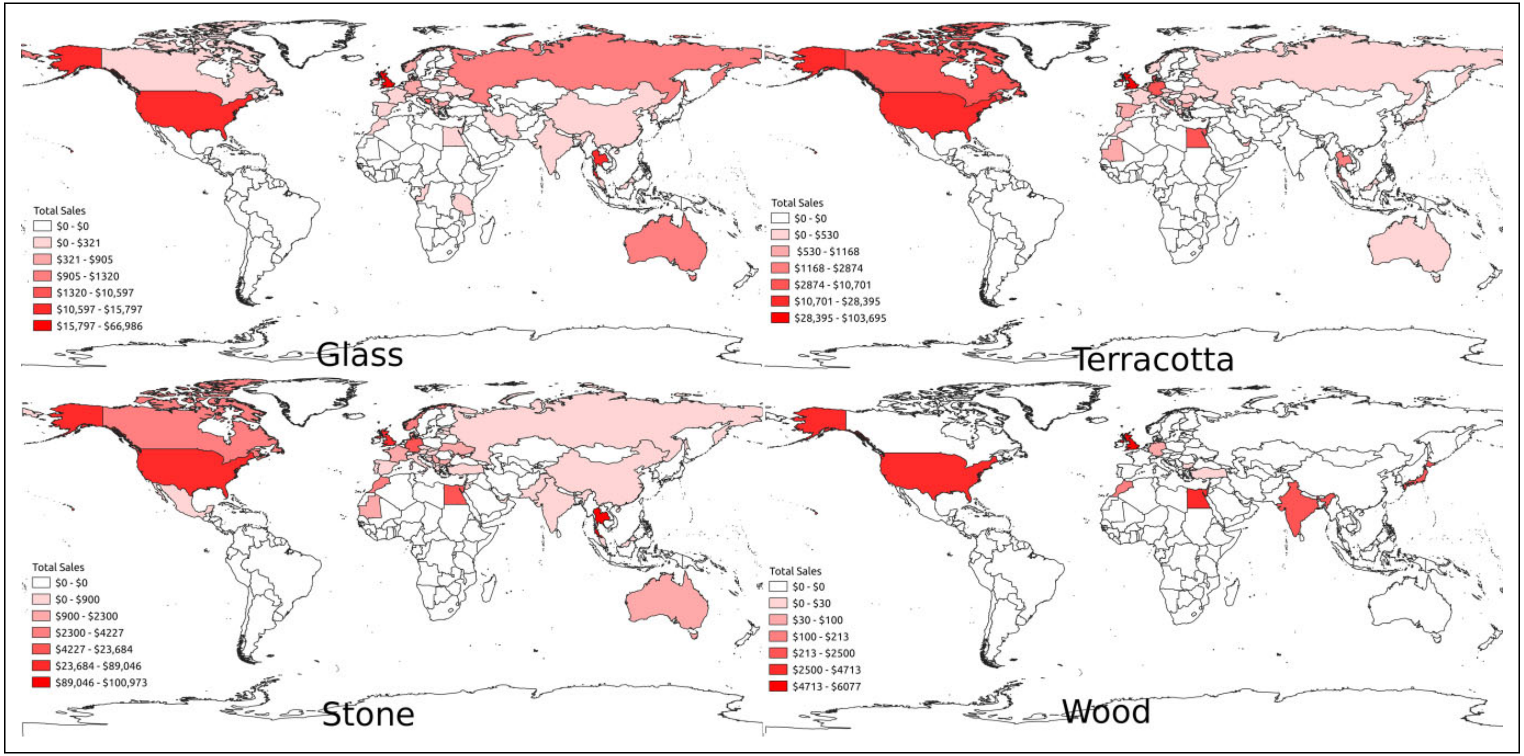

Figure 10. Material types sold in different countries.

masks, papyrus, and wood objects having commanded relatively higher sale prices. Jewelry and metal artifacts are top selling items for object types and materials, respectively. Unlike the other top selling cultures, Egyptian artifacts are mainly statues, religious, and text objects, while stone and terracotta are common materials. Some countries did specialize in certain goods, such as Germany selling Central Asian artifacts and Thailand selling stone artifacts. The United Kingdom is often the top selling country for different artifact types that include religious, tools, vessels, glass, terracotta, and wood objects. The United States is the second highest antiquities seller and similarly has a broad coverage of artifact types sold, while Thailand is the top selling country in Asia. In addition to the United Kingdom, Germany is a leading seller of weapon and tool artifacts; Cyprus is the third highest seller of terracotta. Egypt is among the leading countries in selling terracotta, stone, and wood artifacts. Looking at how cultures, object types, and materials correlate among top selling countries, many of these categories show a strong correlation, which suggested that the top selling countries in antiquities have a relatively broad coverage of artifacts sold, including the cultures, types, and materials these artifacts relate to. Viking, prehistoric, Egyptian, and Islamic cultures are commonly sold in countries where these cultures have been found in Europe, North Africa, and the Middle East; Roman objects are sold in a wide area.

It is worth noting that South America, parts of Asia, and much of sub-Saharan Africa are not among the top sellers of artifacts. Cultures from these areas are also not among the top selling. It is possible eBay is not commonly used for sales from some of these countries, although cultures from those countries are sold in different countries. For instance, although sales of South America antiquities have been covered by the 1983 Cultural Property Implementation Act, artifacts have been smuggled and sold from different selling countries such as the United States (Brodie, 2014b). Cultures from different regions, including South America (classified as part of "America" cultures), are found to be sold on eBay, but sales generally occur in regions outside of where they are found. On the other hand, among top selling countries, Cyprus (2019), Egypt (2019), and Morocco (2019) have laws against selling antiquities; however, these countries are among the top sellers in eBay. It is not always clear whether these cases are illegally sold antiquities as the sellers could have been selling objects coming from other countries; however, illegal sales of antiquities are a clear possibility in many cases. Furthermore, many objects are likely to be forgeries, given the low value for many objects sold, although one cannot determine which objects and how many are fake. Nevertheless, even if sales incorporate fake objects, knowledge of what fakes are sold and details about them help to demonstrate broader demand, a key goal of this research.

Wider Benefits

As for broader benefits, the approach developed is relevant for cultural heritage experts and those interested in the antiquities market because the wider cultural heritage market has been difficult to understand and few systematic opportunities are available to know how such a market continues to develop. The tool presents a dynamic way to monitor part of the antiquities market, including in real time. The analysis of eBay presents only one type of platform, but it does represent one of the larger sellers of at least lower valued antiquities at a global scale (Brodie, 2015; Fay, 2013). Many countries that appear to sell antiquities in eBay, including the top selling countries, are also countries that have been seen as some of the top sellers of antiquities by experts (Anderson, 2017). This suggests eBay could potentially act as a type of wider barometer of the antiquities market, at least for lower end antiquities, although how closely is unclear given empirical data are difficult to retrieve and the fact regions such as South America do not appear to be well represented in sales. Additionally, eBay results may also inform on the illegal market and where it is emerging, given that some of countries that sold antiquities on eBay have clear laws against selling antiquities and there may also be underground markets in some of these countries. Overall, the tool enables insight into the antiquities market, particularly the cultures, object types, and material composition that are selling. This makes it among the first such tools to be able to provide such knowledge on Internet sale patterns.

Conclusion

Studying the market of antiquities has been difficult for researchers for a variety of reasons. The application of NLP is a nascent development in this area, with this study among the first applying NER. The outputs demonstrate the utility of the approach, with the relevant methods and data made publicly available via a repository. Although the precision and recall test showed the approach as having a high F_1 score, improvements can be made by increasing the diversity of texts, such as different sites selling antiquities and covering a longer period of study, that are analyzed so that in the long term, a dictionary would be less needed to assist in text classification. More information, including from other auction sites, could improve the quality of data studied and improve our understanding of objects sold and where the antiquities market is developing. The approach could be reoriented toward other platforms used to sell antiquities, with the NER and dictionary method similarly applied there. Recent work (ATHAR, 2019; Paul, 2018) has indicated social media and the dark web as areas where antiquities are sold, in particular the illegal antiquities market. Another limitation is the ambiguous nature of antiquities themselves. A broad consensus on the categories objects belong to, and their definitions, could help, which is possible through consultations among experts, in the domain of cultural heritage, archaeology, and antiquities. What this work has done is create an initial engine to enable a clearer understanding of the antiquities market at a given time, or at least some part of it that may serve as a wider proxy, and make it potentially more quantifiable, with the approach extensible to other sites.

Author's Note

The author would like to acknowledge the Ancient Identities project, which is funded by the UK Arts and Humanities Research Council (AH/N006151/1), for inspiring this research and influencing its approach, although this work is not directly sponsored by this project.

Data Availability

The version of the tool used for this article can be found at http://discovery.ucl.ac.uk/10079023/ and an active GitHub for the project can be found at https://github.com/maltaweel/eBayScraper

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author received no financial support for the research, authorship, and/or publication of this article.

Software Information

Python 2.7+, Java 8+, and R 3.3, which include various libraries within these languages. To download eBay Scraper, see Data Availability above.

References

Agerri, R., & Rigau, G. (2016). Robust multilingual named entity recognition with shallow semi-supervised features. *Artificial Intelligence*, *238*, 63–82. doi:10.1016/j.artint.2016.05.003

Anderson, M. L. (2017). *Antiquities: What everyone needs to know*. New York, NY: Oxford University Press.

Antiquities Trafficking and Heritage Anthropology Research. (2019). Retrieved July 29, 2019, from http://atharproject.org/report2019/

Barker, A. W. (2018). Looting, the antiquities trade, and competing valuations of the past. *Annual Review of Anthropology*, *47*, 455–474. Retrieved from https://doi.org/10.1146/annurev-anthro-102116-041320

Bauer, A. A. (2015). Editorial: The destruction of heritage in Syria and Iraq and its implications. *International Journal of Cultural Property*, *22*, 1–6. doi:10.1017/S0940739115000090

Bonacchi, C., Altaweel, M., & Krzyzanska, M. (2018). The heritage of Brexit: Roles of the past in the construction of political identities through social media. *Journal of Social Archaeology*, *18*, 174–192. doi:10.1177/1469605318759713

Bowman, B. A. (2008). Transnational crimes against culture: Looting at archaeological sites and the "grey" market in antiquities. *Journal of Contemporary Criminal Justice*, *24*, 225–242. doi:10.1177/1043986208318210

Brodie, N. (2012). *Uncovering the antiquities market*. Oxford, England: Oxford University Press. Retrieved from https://doi.org/10.1093/oxfordhb/9780199237821.013.0013

Brodie, N. (2014a). Provenance and price: Autoregulation of the antiquities market? *European Journal on Criminal Policy and Research*, *20*, 427–444. doi:10.1007/s10610-014-9235-9

Brodie, N. (2014b). The Internet market in Precolumbian antiquities. In J. Kila & M. Balcells (Eds.), *Cultural property crime: An overview and analysis on contemporary perspectives and trends* (pp. 237–262). Leiden, the Netherlands: Brill.

Brodie, N. (2015). The internet market in antiquities. In F. Desmarais (Ed.), *Countering illicit traffic in cultural goods: The global challenge of protecting the world's heritage* (pp. 11–20). Paris, France: ICOM.

Brodie, N., Kersel, M. M., & Tubb, K. W. (2006). Introduction. In N. Brodie, M. M. Kersel, C. Luke, & K. W. Tubb (Eds.), *Archaeology, cultural heritage, and the antiquities trade* (pp. 1–24). Gainesville: University Press of Florida. Retrieved from https://doi.org/10.5744/florida/9780813029726.003.0001

Brosché, J., Legnér, M., Kreutz, J., & Ijla, A. (2017). Heritage under attack: Motives for targeting cultural property during armed conflict. *International Journal of Heritage Studies*, *23*, 248–260. doi:10.1080/13527258.2016.1261918

Cambria, E., & White, B. (2014). Jumping NLP curves: A review of natural language processing research. *IEEE Computational Intelligence Magazine*, *9*, 48–57. doi:10.1109/MCI.2014.2307227

Campbell, P. B. (2013). The illicit antiquities trade as a transnational criminal network: Characterizing and anticipating trafficking of cultural heritage. *International Journal of Cultural Property*, *20*, 113–153.

Cullen, A. C., & Frey, H. C. (1999). *Probabilistic techniques in exposure assessment: A handbook for dealing with variability and uncertainty in models and inputs*. New York, NY: Plenum Press.

Cyprus. (2019). *Laws governing antiquities*. Retrieved May 21, 2019, from http://www.mcw.gov.cy/mcw/da/da.nsf/All/A2ABFCFE258EFD71C22571A2003A2B9D/$file/law-en-1.pdf

eBay API. (2019). Retrieved July 27, 2019, from https://developer.ebay.com/

Egypt. (2019). Supreme Council of Antiquities. Retrieved May 21, 2019, from http://www.antiquities.gov.eg/DefaultEn/Documents/LAW/LAW%20ENGLISH6.pdf

Fay, E. (2011). Virtual artifacts: eBay, antiquities, and authenticity. *Journal of Contemporary Criminal Justice*, *27*, 449–464. doi:10.1177/1043986211418887

Fay, E. (2013). *Trading in antiquities on eBay: The changing face of the illicit trade in antiquities* (Unpublished PhD thesis). University of Keele, Keele, England.

Finkel, J. R., Grenager, T., & Manning, C. (2005). Incorporating non-local information into information extraction systems by Gibbs sampling. In *Proceedings of the 43nd Annual Meeting of the Association for Computational Linguistics* (pp. 363–370). Ann Arbor, MI: Association for Computational Linguistics.

Goutte, C., & Gaussier, E. (2005). A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. In D. E. Losada & J. M. Fernández-Luna (Eds.), *Advances in information retrieval* (pp. 345–359). Santiago de Compostela, Spain: European Conference on Information Retrieval. doi:10.1007/978-3-540-31865-1_25

Greenland, F., Marrone, J. V., Topçuoğlu, O., & Vorderstrasse, T. (2019). A site-level market model of the antiquities trade. *International Journal of Cultural Property*, *26*, 21–47. doi:10.1017/S0940739119000018

Hardy, S. A. (2016). *Illicit trafficking, provenance research and due diligence: The state of the art*. Report for the United Nations Educational, Scientific and Cultural Organization, Paris, France.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65–70.

Huffer, D., & Graham, S. (2017). The Insta-dead: The rhetoric of the human remains trade on Instagram. *Internet Archaeology*, *45*. doi:10.11141/ia.45.5

Konkol, M., Brychcín, T., & Konopík, M. (2015). Latent semantics in named entity recognition. *Expert Systems with Applications*, *42*, 3470–3479. doi:10.1016/j.eswa.2014.12.015

Kotha, S., & Basu, S. (2011). Amazon and eBay: Online retailers as market makers. In G. G. Hamilton, M. Petrovic, & B. Senauer (Eds.), *The market makers: How retailers are reshaping the global economy* (pp. 155–180). Oxford, England: Oxford University Press.

Kumar, E. (2012). *Natural language processing*. New Delhi, India: I. K. International Publishing House.

Lafferty, J., McCallum, A., & Pereira, F. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the 18th International Conference on Machine Learning* (pp. 282–289). San Francisco, CA, USA: Morgan Kaufmann.

Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., & Dyer, C. (2016). Neural architectures for named entity recognition. In K. Knight, A. Nenkova, & O. Rambow (Eds.), *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 260–270). San Diego, CA: Association for Computational Linguistics. doi: 10.18653/v1/N16-1030

Lee, C., Hwang, Y.-G., Oh, H.-J., Lim, S., Heo, J., Lee, C.-H., ... Jang, M.-G. (2006). Fine-grained named entity recognition using conditional random fields for question answering. In H. T. Ng, M.-K. Leong, M.-Y. Kan, & D. Ji (Eds.), *Information retrieval technology* (pp. 581–587). Berlin, Germany: Springer. doi:10.1007/11880592_49

Mackenzie, S., & Yates, D. (2017). What is grey about the "grey market" in antiquities? In J. Beckert & M. Dewey (Eds.), *The architecture of illegal markets: Towards an economic sociology of illegality in the economy* (pp. 70–86). Oxford, England: Oxford University Press.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. New York, NY: Cambridge University Press.

Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., & McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. In K. Bontcheva & J. Zhu (Eds.), *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (pp. 55–60). Baltimore, MD: Association for Computational Linguistics. doi:10.3115/v1/P14-5010

Morocco. (2019). *Law on antiquities*. Retrieved May 21, 2019, from https://whc.unesco.org/en/statesparties/ma/laws/

Nadeau, D., & Sekine, S. (2007). A survey of named entity recognition and classification. *Lingvisticae Investigationes*, *30*, 3–26. doi:10.1075/li.30.1.03nad

Neelakantan, A., & Collins, M. (2014). Learning dictionaries for named entity recognition using minimal supervision. In S. Wintner, S. Goldwater, & S. Riezler (Eds.), *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 452–461). Gothenburg, Sweden: Association for Computational Linguistics. doi:10.3115/v1/E14-1048

NLP Stanford. (2019). *The Stanford natural language processing group*. Retrieved March 3, 2019, from https://nlp.stanford.edu/software/CRF-NER.html

Nothman, J., Ringland, N., Radford, W., Murphy, T., & Curran, J. R. (2013). Learning multilingual named entity recognition from Wikipedia. *Artificial Intelligence*, *194*, 151–175. doi:10.1016/j.artint.2012.03.006

Paul, K. (2018). Ancient artifacts vs. digital artifacts: New tools for unmasking the sale of illicit antiquities on the dark web. *Arts*, *7*, 12. doi:10.3390/arts7020012

Powers, D. M. W. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. *Journal of Machine Learning Technologies*, *2*, 37–63.

Ritter, A., Clark, S., Etzioni, M., & Etzioni, O. (2011). Named entity recognition in tweets: An experimental study. In *Proceedings of the conference on empirical methods in natural language processing (EMNLP '11)* (pp. 1524–1534). Stroudsburg, PA: Association for Computational Linguistics.

Wang, X., Jiang, X., Liu, M., He, T., & Hu, X. (2017). Bacterial named entity recognition based on dictionary and conditional random field. In *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 439–444). Kansas City, MO: IEEE. doi:10.1109/BIBM.2017.8217688

Zafarian, A., Rokni, A., Khadivi, S., & Ghiasifard, S. (2015). Semi-supervised learning for named entity recognition using weakly labeled training data. In *2015 the International Symposium on Artificial Intelligence and Signal Processing (AISP)* (pp. 129–135). Mashhad, Iran: IEEE. doi:10.1109/AISP.2015.7123504

Author Biography

Mark Altaweel is a reader at the Institute of Archaeology, where he teaches data science in archaeology and works as a computational social scientist and archaeologist.