

# Zero-Shot Information Extraction to Enhance a Knowledge Graph Describing Silk Textiles

Thomas Schleider and Raphael Troncy

EURECOM, France

thomas.schleider@eurecom.fr | raphael.troncy@eurecom.fr

## Abstract

The knowledge of the European silk textile production is a typical case for which the information collected is heterogeneous, spread across many museums and sparse since rarely complete. Knowledge Graphs for this cultural heritage domain, when being developed with appropriate ontologies and vocabularies, enable to integrate and reconcile this diverse information. However, many of these original museum records still have some metadata gaps. In this paper, we present a zero-shot learning approach that leverages the ConceptNet common sense knowledge graph to predict categorical metadata informing about the silk objects production. We compared the performance of our approach with traditional supervised deep learning-based methods that do require training data. We demonstrate promising and competitive performance for similar datasets and circumstances and the ability to predict sometimes more fine-grained information. Our results can be reproduced using the code and datasets published at <https://github.com/silkknow/ZSL-KG-silk>.

## 1 Introduction

The mentioning of European silk textiles often evokes images of clothes and furniture of the old aristocracies and the lavish lifestyles of kings and queens. Nowadays, the knowledge about the occidental way of producing these expensive items is, however, more and more endangered. Many museums and collections around the globe fortunately still have silk objects, or at least public records with metadata and images illustrating them. Such specific museum data, from many different sources about Cultural Heritage objects that are partly centuries old, have naturally some gaps: sometimes, the production year or place is unknown, but the material and technique used is described; sometimes, a rich textual description is provided with many little details about the object production and

what it depicts, but categorical values informing about the exact material or technique used is not provided (Figure 1).

The recent progress in natural language processing and more specifically in information extraction can help to address these problems. In particular, it is common to train text classification models from metadata records in order to predict missing categorical values in other records. However, such models do require a significant amount of annotated data for training, which is expensive to get when the domain is very specific. In this paper, we first briefly introduce the development of a multilingual knowledge graph (KG) about silk textiles production. This knowledge graph, developed with the help of historians and museum experts, uses the CIDOC-CRM ontology and a domain expert designed thesaurus defining the silk textile concepts. Next, we propose a Zero-Shot Learning (ZSL) approach based on the ConceptNet common-sense knowledge graph (Speer et al., 2017), to predict the missing categorical metadata while avoiding to rely on training data. We compare our approach with supervised approaches and we show competitive results and demonstrate the ability to predict more fine-grained concepts despite the specificity of this domain.

The remainder of this paper is structured as follows. We present some relevant related work in Section 2. We summarise the development of the Knowledge Graph and we describe the dataset being used in our experiments in Section 3. We detail our Zero Shot Learning approach as well as two supervised learning baselines in Section 4. We analyze the classification results in Section 5. Finally, we provide conclusions and outline some future work in Section 6.

## 2 Related Work

Integrating Cultural Heritage data into a knowledge graph can be difficult and time consuming, as data



Figure 1: Examples from three different museums with missing categorical properties: a) no subject depiction for the record 37.80.1 from the Metropolitan Museum of Art; b) no material for the record Cl. XXIV n. 1748 from the Musei di Venezia; c) no technique for the record GMMP-733-002 from the French Mobilier National

has to be usually collected from many different sources that generally do not use standard formats. For museum data, there are models and ontologies that make such a process less challenging. We consider the CIDOC Conceptual Reference Model (CRM) to be one of them. CIDOC-CRM is an event-centric ontology through which everything can be represented as an event. A man-made object has, for example, been produced at some point in time. Certain materials might have been used during this production that has taken place at a specific location in a specific century. CIDOC-CRM offers not only classes and properties for such an event-based representation, but is also easily expandable with classes from other ontologies. Hence, CIDOC-CRM comes also with useful extensions such as CRMSci (Scientific Observation Model) and CRMdig (model for provenance metadata). CIDOC-CRM is an official ISO standard since 2006 and this status has been renewed in 2014.

There are more and more successful examples of KG development in the Cultural Heritage realm. For example, ARCO (Carriero et al., 2019) is a knowledge graph about the Italian Cultural Heritage that at least indirectly reuses some CIDOC-CRM classes and properties. The Dutch Rijksmuseum collection is available as linked open data (Dijkshoorn et al., 2018) while DOREMUS is the largest knowledge graph available about about classical musical works and interpretations (Achichi et al., 2018).

Information extraction from texts generally rely on supervised machine learning and deep learning classification techniques that require labelled training data. While general purpose transformer-based language models are often used, fine-tuning them often require numerous examples (Dodge et al., 2020). One possible solution to address this lack

of labelled data is to use active learning (Yuan et al., 2020) in conjunction with pre-trained models such as BERT (Devlin et al., 2019) that comes with possible bias problems (Papakyriakopoulos et al., 2020). In Cultural Heritage domains, several works are trying to compensate for the lack of labelled data. One method is to leverage human annotations through crowdsourcing together with extracted visual and textual features and automatic annotation through transfer learning (Shabani et al., 2020). In general Convolutional Neural Networks (CNN) are often used for image, audio and video data (Clermont et al., 2020), whereas Recurrent Neural Networks (RNN) is also used for textual Cultural Heritage data (Kambau et al., 2018). Regular neural network (Belhi et al., 2018) for textual data are also used together with word embeddings.

More recently, zero-shot learning approaches have attracted a lot of attention for their ability to offer text classification without relying on training data. It is a form of automatic classification for which textual documents unseen by the model can be analyzed and classified. Several frameworks have been proposed over the years, based on BERT (Zhang et al., 2019), (Ye et al., 2020) or other large pre-trained models (Weller et al., 2020). Our approach relies on ZeSTE<sup>1</sup> (Zero Shot Topic Extraction) (Harrando and Troncy, 2021) that provides a framework for extracting topics from textual documents using the ConceptNet commonsense knowledge graph. In addition, the framework provides explainability of its classification results using the ConceptNet KG neighborhood.

<sup>1</sup><https://github.com/D2KLab/ZeSTE>

### 3 A Knowledge Graph of Silk Cultural Heritage

#### 3.1 Knowledge Graph Development

The development of the knowledge graph has followed an established extraction, transform and load (ETL) pipeline. First, domain experts have made a relevant selection of museum websites. We have developed a crawling software that directly scrapes the public records and images from their websites or relies on open APIs when available, and saved the metadata information into a common intermediate JSON format. Next, domain experts have provided mapping rules by interpreting the original metadata to a unified target ontology model based on CIDOC-CRM. This enables to address a first heterogeneity problem as how similar information are differently expressed across museums. For example, the production place of a silk textile is represented in a field named “Culture” in the case of the MET museum and in a field named “Object Place” in the case of the Museum of Fine Arts of Boston. Museums come from all over the world and the field names are also multilingual. For example, the production place would be informed in a field named “Lugar de Producción/Ceca” in the case of the Red Digital de Colecciones de Museos de España.

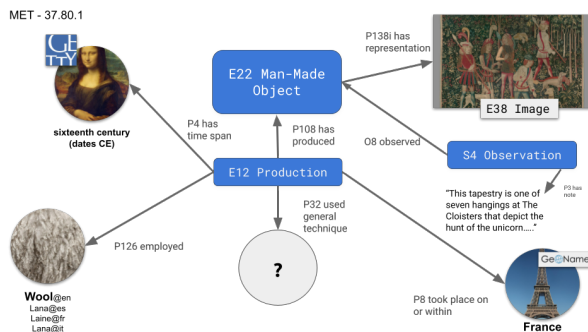


Figure 2: The MET record provided in Figure 1a represented in the knowledge graph using the CIDOC-CRM ontology and controlled vocabularies

The conversion process also addresses a second heterogeneity problem by disambiguating the original metadata field values to common concepts defined in a multilingual thesaurus about the silk production. Hence, specific weaving techniques or materials used that are mentioned in the metadata get linked to these unique concepts (Figure 2). This disambiguation is performed by string matching and simple heuristics. Other controlled vocabular-

ies for normalizing place names (Geonames<sup>2</sup>) and time periods (Getty AAT) are also used. The resulting knowledge graph can be accessed via SPARQL queries or a dedicated RESTful API.

#### 3.2 Dataset and Preprocessing

The multilingual Knowledge Graph of silk textile productions consists of descriptions of 36210 unique objects illustrated by 74527 images in four languages: English, Spanish, French and Italian. While the information integration process has been effective, one general problem of the KG is that many properties have missing values. In this paper, we focus on three important properties describing the silk production namely: the material used, the weaving technique employed and a the subject depicted. Consequently, we extract from the knowledge graph three subsets corresponding to the set of objects having values for those properties.

The silk thesaurus which is being used to normalize the values of those properties contains a very exhaustive inventory of possible materials and techniques, organized in a hierarchy. While some of those materials or techniques are widely used in the data, others are niche and the knowledge graph includes only a very limited number of objects with some of them. One solution is to walk up the thesaurus hierarchy and only consider more general concepts. Ultimately, we need to find a trade off between using fine-grained concepts with the risk of having too sparse data, or too broad concepts with the risk of being non informative. This is a manual process informed by both the thesaurus hierarchy and the available data. Table 1 provides the list of the thesaurus concepts that we finally aim to predict for the three properties (material, technique, depiction) as well as the number of unique objects.

As a general preprocessing step, we also removed all records that have multi-valued properties. Some records have, for example, both “Gold Thread” and “Vegetable Fiber” set as material properties. Including such a record would make the training of a model capable of distinguishing these two concepts harder. We also create language specific subsets. We observe significant differences between the English and Spanish subsets, which highlight the heterogeneous nature of our sources. In particular, subject depiction sticks out as we only have objects from Spanish records having this property informed. The language specific subsets

<sup>2</sup><https://www.geonames.org/>

will be used by our Zero Shot Learning approach (Section 4.2) while supervised learning methods will make use of the complete multilingual dataset in the 4 languages (Section 4.1).

Property	Concept	Nb of Objects	
		ES	EN
Material	Vegetal Fibre	225	-
Material	Gold Thread	34	125
Material	Silver Thread	10	176
Material	<b>SUBTOTAL</b>	<b>269</b>	<b>301</b>
Technique	Damask	512	25
Technique	Brocaded	-	190
Technique	Tabby	-	37
Technique	┆ Muslin	2	37
Technique	┆ Louisine	11	-
Technique	Satin	228	93
Technique	Embroidery	66	343
Technique	Velvet	111	236
Technique	┆ Voiled Velvet	-	1
Technique	┆ Façonne cut velvet	102	-
Technique	┆ Ciselé velvet	26	-
Technique	┆ Plain cut velvet	23	-
Technique	┆ Cut velvet	8	-
Technique	<b>SUBTOTAL</b>	<b>1089</b>	<b>955</b>
Depiction	Floral Motif	965	-
Depiction	┆ Bunch	39	-
Depiction	┆ Rose	2	-
Depiction	┆ Fleur de lis	8	-
Depiction	Vegetal Motif	205	-
Depiction	┆ Leaf	36	-
Depiction	┆ Thistle	35	-
Depiction	┆ Vine	4	-
Depiction	Geometrical Motif	125	-
Depiction	┆ Rhombus	6	-
Depiction	<b>SUBTOTAL</b>	<b>1425</b>	<b>-</b>
<b>All</b>	<b>TOTAL</b>	<b>3425</b>	

Table 1: Number of objects exported for each property (material, technique, depiction)

## 4 Approach

### 4.1 Supervised Approaches

In order to be able to evaluate our approach, we propose to compare it with two supervised classification methods that we use as baselines. For both of them, we will use the three sub-sets described in the Table 1 and perform a 80%-20% split in order to have training data. We will perform a five-fold cross validation for testing the models.

#### Classification based on textual descriptions.

The goal of this approach is to predict missing categorical values of museum records based on lengthy textual descriptions such as the note of the `S4_Observation` in Figure 2. This approach consists of a Convolutional Neural Network (CNN) built over cross-lingual pre-trained word embeddings which are the aligned fastText vectors trained

on Wikipedia (Joulin et al., 2018). More precisely, a series of convolutional blocks with varying kernel sizes (2,3,4), each consisting of 100 filters, are applied to a sequence of such word-embeddings that got mapped from input description texts from the Knowledge Graph. These filters create an output for which a Gaussian Error Unit (GELU) non-linearity is used and a max-pooling operation is applied for each block. The idea is to, hopefully, select the best features of each block. Afterwards, they get concatenated into one single vector, regularised by a dropout layer and finally sent to a softmax classification layer to come up with the final predictions per input.

#### Classification based on images and metadata.

The goal of this approach is also to predict missing categorical values but this time, using the images illustrating the objects with the other metadata values. The underlying assumption is, that it is to some degree visible on images what materials have been used or which technique was employed to produce a silk textile. For this model, a CNN was also used. More precisely, a pre-trained ResNet backbone network served as a generic feature extraction network. The output is then processed by several fully connected network layers that deliver a joint representation of the images and a final classification layer offers afterwards a probabilistic class score per variable and concept. The model is trained based on multi-task learning to perform predictions for the three properties simultaneously: material, technique and depiction. The training is based on stochastic minibatch gradient descent and using focal loss. For ResNet only, the last convolutional layers are fine-tuned.

### 4.2 Zero-Shot Prediction

The benefit of our approach is to perform a similar prediction without relying on any training data. The underlying assumption is that a textual document about a topic such as “Embroidery” will probably mention other words that are similar to this concept such as “fabric” or “stitch”.<sup>3</sup> Our approach relies on the ConceptNet common sense knowledge graph. More precisely, we need to feed our approach with mappings between the targeted values we wish to predict and the ConceptNet network. These mappings have been manually established (Table 3).

<sup>3</sup>See for example the object described at <https://ada.silkknow.org/object/c57358a7-c908-3110-b65d-70b09f5f4c4b>

Approach	Language	Class	Precision	Recall	F1-score	Support
ZSL	ES	Vegetal Fiber	2.7%	100%	5.2%	6
ZSL	ES	Gold	70.6%	10.3%	17.9%	234
ZSL	ES	Silver	50.0%	17.2%	25.6%	29
ZSL	EN	Gold	96%	57.1%	71.6%	210
ZSL	EN	Silver	49.9%	95.5%	64.4%	91
CNN - Image	EN/ES/FR/IT	Vegetal Fiber	36.8%	24.1%	29.2%	
CNN - Image	EN/ES/FR/IT	Metal	32.1%	36.4%	34.1%	
CNN - Text	EN/ES/FR/IT	Vegetal Fiber	79.0%	81.0%	80.0%	229
CNN - Text	EN/ES/FR/IT	Metal	76.0%	61.0%	68.0%	125

Table 2: Results for the material property across approaches

Property	KG Concept	ConceptNet
Material	Vegetal Fibre	/c/es/vegetal
Material	Gold Thread	/c/es/oro, /c/en/gold
Material	Silver Thread	/c/es/plata, /c/en/silver
Technique	Damask	/c/es/damasco, /c/en/damask
Technique	Embroidery	/c/es/bordado, /c/en/embroidery
Technique	Velvet	/c/es/terciopelo, /c/en/velvet
Technique	Voided Velvet	/c/en/velvet
Technique	Cut Velvet	/c/es/terciopelo
Technique	Plain Cut Velvet	/c/es/terciopelo
Technique	Façonne Cut Velvet	/c/es/terciopelo
Technique	Ciselé Velvet	/c/es/terciopelo
Technique	Tabby (silk weave)	/c/en/tabby
Technique	Louisine	/c/es/tafetán
Technique	Muslin	/c/es/tafetán
Technique	Satin (Fabric)	/c/es/raso, /c/en/satin
Technique	Brocaded	/c/en/brocaded
Depiction	Vegetal Motif	/c/es/planta
Depiction	Vine	/c/es/planta
Depiction	Thistle	/c/es/planta
Depiction	Leaf	/c/es/planta
Depiction	Floral Motif	/c/es/flor
Depiction	Fleur-de-lis	/c/es/flor
Depiction	Rose	/c/es/flor
Depiction	Bunch	/c/es/flor
Depiction	Geometrical Motif	/c/es/geometría
Depiction	Rhombus	/c/es/geometría

Table 3: Mapping between the concepts used in our knowledge graph and ConceptNet

ConceptNet is then used to produce a list of candidate words related to the concept of interest, which can be called “topic neighbourhood”. Each topic neighbourhood is created by querying every node that is N steps away from the label node. In our experiments, we set N=2. A score for each label is computed based on the content of the text document that we give as input. This score is calculated based on cosine similarity via ConceptNet Numberbatch, the graph embeddings of the network. Even if a word has several meanings, only one neighbourhood per spelling is generated. The score is then supposed to represent the relevance of any other term to the main label inside a neighbourhood. Based on these scores, the whole document (museum record) gets also a score and, therefore, a document label too. This is done by quantifying the overlap between the document content as a list of tokens and the label neighbourhood nodes. Finally, as mentioned before, all predicted document labels can be explained by the model through showing the path between nodes or highlighting the words or n-grams that contributed to the final classification.

This approach has a number of limitations: the concepts that should be predicted must exist in ConceptNet. Furthermore, while ConceptNet is multilingual, the embeddings are language specific. Therefore, our zero shot learning approach will make use of the language specific subsets described in Section 3.

## 5 Evaluation

In this section, we compare the results of our Zero-Shot Learning (ZSL) approach with the supervised methods described in the Section 4.1. We present the results alongside each of the three properties of interest: material, technique and depiction. It is worth to note that the precision, recall, F1 scores

are obtained on 20% of the dataset following a 5-fold cross validation while the figures reported for the ZSL method concerns the entire language specific datasets described in Section 3.

Table 2 shows the results for predicting material concepts. The two baselines approaches can only predict whether the material used is “Metal” or “Vegetable Fiber” while our ZSL approach can predict more fine grained concepts than just “Metal”, such as “Gold” or “Silver”. On the English subset, the ZSL method shows promising results with F1-score of 71.6% and 64.4% respectively. On the Spanish subset, the prediction results are lower for the ZSL method, in comparison to the supervised approach. The topic neighborhood in this language is also less elaborated. The Text CNN method benefits clearly from the multilingual embeddings.

The results for the prediction of technique concepts are presented in Table 4. On the Spanish subset, we again observe lower scores for the ZSL approach. The problem relies on both the quality of the Spanish textual descriptions and on the Spanish concept neighbourhoods in ConceptNet. On the English subset, the ZSL approach performs in par with respect to the Image CNN baseline but less well than the Text CNN one. We observe that “Tabby”, a very domain-specific concept, is particularly difficult to predict and it was discarded by the Text CNN approach since too little usable textual descriptions were available. The ZSL method performs reasonably well with a similar input and better than the Image CNN baseline.

Figure 3 provides the confusion matrix of the predictions made for the technique property with the ZSL method on the English subset. We can see that the true label is usually predicted the most per class, especially in the case of “embroidery” and “velvet”. We observe that “Brocaded” is often confused with “embroidery” or “velvet” while “Damask” and “tabby” are rarely predicted. In the case of “damask”, this just reflects a low amount of samples, whereas in the case of “tabby”, the predictions almost do not work at all (F1-score of 2.9% ).

Figure 4 shows an example where “Embroidery” is correctly predicted, while Figure 5 depicts a counter-example (the correct technique should have been “Embroidery”). The graph highlights the most relevant words used for the predictions. The first example is based on the text: "*Spot samplers feature motifs that are scattered in a seemingly ran-*

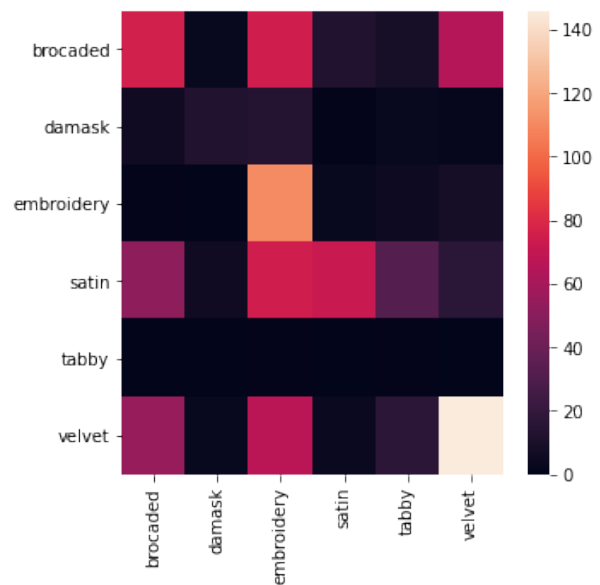


Figure 3: Confusion matrix for the property technique on the English subset for the ZSL method. The Y-axis represents the true labels and the X-axis the predicted ones.

*dom fashion over the surface of the foundation fabric, usually linen. These samplers are rarely signed or dated, and often include motifs that are only partially worked, leading to the conclusion that this type of sampler was made as a personal stitch reference for its maker, and not for display, as band samplers were signed by student embroiderers. The sampler features flowers, obelisks on pedestals, and an "S" motif, in addition to geometric designs that are of the type that would have been used to decorate small purses, cushions, and other accessories.*", taken from a record from the MET Museum.<sup>4</sup>

The second text is "*An example of the kind of work [Catherine de Medici] appreciated is the Museum's panel of yellow satin embroidered with silk threads. One of a set of three (the others are in the Musée Historique des Tissus, Lyon), it hung as a valence around the top of a four-poster bed. Various print sources were culled for the airy design of grotesques, while its five vignettes derived from Ovid's Metamorphoses- based on the myths of Europa, Actaeon, Semele, Pyramus, and Salmacis- are adapted from woodcut illustrations published by Bernard Salomon in Lyon in 1557. Its brilliant colors, exquisite design, and sumptuous material would have suited the queen's taste perfectly.*", also

<sup>4</sup><https://ada.silknow.org/object/c57358a7-c908-3110-b65d-70b09f5f4c4b>

Approach	Language	Class	Precision	Recall	F1-score	Support
ZSL	ES	Damask	6.8%	63.6%	11.2%	55
ZSL	ES	Embroidery	89.4%	6%	12.3%	987
ZSL	ES	Tabby	0%	0%	0%	6
ZSL	ES	Velvet	10.4%	100%	6.6%	28
ZSL	ES	Satin	3.5%	61.5%	18.8%	13
ZSL	EN	Damask	52.0%	33.3%	40.6%	39
ZSL	EN	Embroidery	32.4%	86.7%	47.1%	128
ZSL	EN	Velvet	61.6%	50.0%	55.2%	292
ZSL	EN	Tabby	1.5%	50.0%	2.9%	2
ZSL	EN	Satin	77.4%	28.5%	41.6%	253
ZSL	EN	Brocaded	40%	31.5%	35.3%	241
CNN - Image	EN/ES/FR/IT	Damask	70.3%	68.9%	68.9%	
CNN - Image	EN/ES/FR/IT	Embroidery	83.9%	83.2%	83.6%	
CNN - Image	EN/ES/FR/IT	Tabby	17.4%	30.8%	22.3%	
CNN - Image	EN/ES/FR/IT	Velvet	70.1%	67.8%	68.9%	
CNN - Text	EN/ES/FR/IT	Damask	98%	90%	94%	135
CNN - Text	EN/ES/FR/IT	Embroidery	96%	98%	97%	230
CNN - Text	EN/ES/FR/IT	Velvet	95%	84%	89%	62

Table 4: Results for the technique property across approaches

Approach	Language	Class	Precision	Recall	F1-score	Support
ZSL	ES	Flower	99.3%	72.1%	83.5%	1397
ZSL	ES	Plant	2.2%	26.1%	4.0%	23
ZSL	ES	Geometry	3.1%	100%	5.9%	4
CNN - Image	ES/FR	Flower	89.9%	88.8%	89.3%	
CNN - Image	ES/FR	Plant	45.1%	38.1%	41.3%	
CNN - Image	ES/FR	Geometry	35.8%	50.0%	41.3%	

Table 5: Results for subject depiction property across approaches

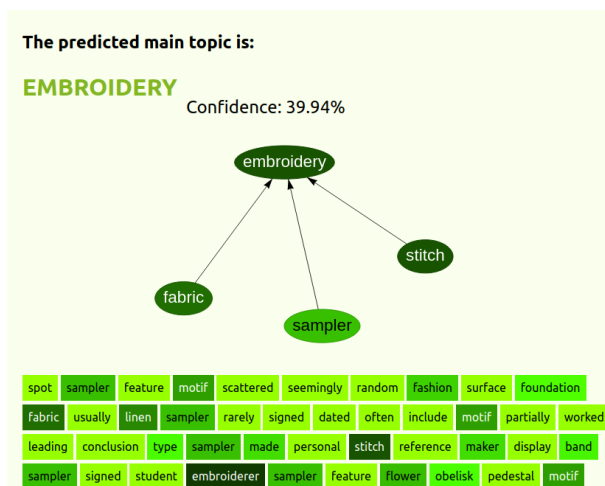


Figure 4: "Embroidery" was correctly predicted by our ZSL approach (English, Technique) in this case. Relevant words in the ConceptNet topic neighborhood are highlighted.

taken from the MET Museum.<sup>5</sup>

The results for the prediction of subject depiction concepts are presented in Table 5. No results are reported for the Text CNN baseline due to the lack of data. We observe that our ZSL approach performs well for predicting the "Flower" concept as does the visual approach. The "Plant" and "Geometry" concepts are however more complicated to predict by the ZSL method. These concepts are general in ConceptNet and the topic neighborhood too broad for the narrower interpretation expected in the silk domain.

## 6 Conclusion and Future Work

In this paper, we hypothesize that a common sense knowledge graph such as ConceptNet can feed a Zero Shot Learning method for enriching a domain

<sup>5</sup><https://ada.silknow.org/object/e2f34144-9ce4-3bc4-b3e0-c67854cd994f>

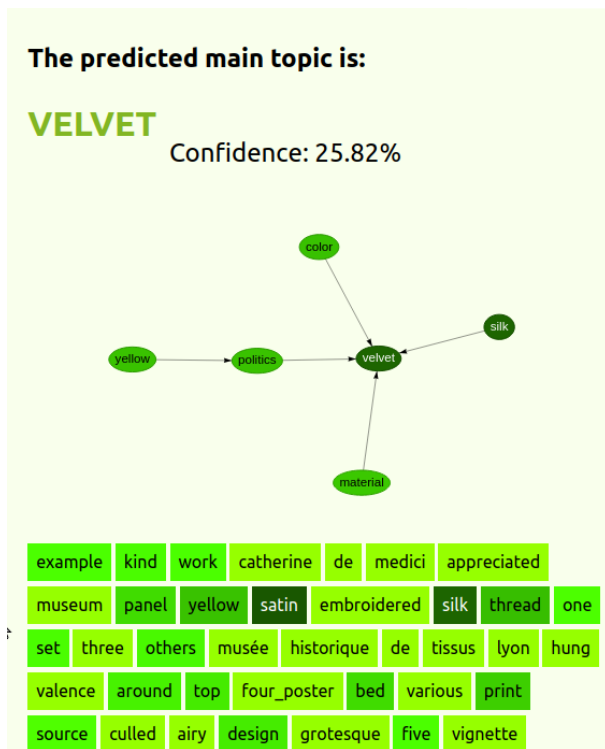


Figure 5: "Velvet" was predicted instead of "Embroidery" by our ZSL approach (English, Technique) in this case. Relevant words in the ConceptNet topic neighborhood are highlighted.

specific knowledge graph such as one describing the silk textile production. Through extensive experiments, we have demonstrated promising results for such an approach in its ability to sometimes reliably predict fine-grained concepts without requiring any training data as supervised classification techniques do. Nevertheless, we observe several limitations: the concepts that should be predicted must exist in ConceptNet with an appropriate topic neighborhood. Our results can be reproduced using the code and datasets published at <https://github.com/silknow/ZSL-KG-silk>.

Even if supervised methods for metadata predictions perform generally better, ZSL remains an interesting method to get accurate predictions even in specific domains that often suffer from data sparsity. We observe that it is also possible to bootstrap the predictions in using first the ZSL method and then applying supervised classification models to further increase performance. We aim to experiment with such hybrid approaches in the future.

## Acknowledgment

This work has been partially supported by the European Union's Horizon 2020 research and innova-

tion program within the SILKNOW (grant agreement No. 769504).

## References

- Manel Achichi, Pasquale Lisena, Konstantin Todorov, Raphael Troncy, and J. Delahousse. 2018. DOREMUS: A Graph of Linked Musical Works. In *International Semantic Web Conference (ISWC)*.
- Abdelhak Belhi, Abdelaziz Bouras, and Sebti Foufou. 2018. Leveraging Known Data for Missing Label Prediction in Cultural Heritage Context. *Applied Sciences*, 8(10).
- Valentina Anita Carriero, Aldo Gangemi, M. Mancinelli, L. Marinucci, Andrea Giovanni Nuzzolese, V. Presutti, and Chiara Veninata. 2019. ArCo: the Italian Cultural Heritage Knowledge Graph. In *International Semantic Web Conference (ISWC)*.
- D. Clermont, M. Dorozynski, D. Wittich, and F. Rotensteiner. 2020. Assessing the Semantic Similarity of Images of Silk Fabrics Using Convolutional Neural Networks. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-2-2020:641–648.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4171–4186, Minneapolis, Minnesota.
- Chris Dijkshoorn, Lizzy Jongma, Lora Aroyo, J. V. Ossenbruggen, G. Schreiber, Wesley ter Weele, and J. Wielemaker. 2018. The Rijksmuseum collection as Linked Data. *Semantic Web*, 9:221–230.
- Jesse Dodge, Gabriel Ilharco, Roy Schwartz, A. Farhadi, Hannaneh Hajishirzi, and Noah A. Smith. 2020. [Fine-Tuning Pretrained Language Models: Weight Initializations, Data Orders, and Early Stopping](#).
- Ismail Harrando and Raphael Troncy. 2021. Explainable Zero-Shot Topic Extraction Using a Common-Sense Knowledge Graph. In *3<sup>rd</sup> Conference on Language, Data and Knowledge (LDK)*, Zaragoza, Spain.
- Armand Joulin, Piotr Bojanowski, Tomas Mikolov, Hervé Jégou, and Edouard Grave. 2018. Loss in Translation: Learning Bilingual Word Mapping with a Retrieval Criterion. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2979–2984.
- Ridwan Andi Kambau, Zainal Arifin Hasibuan, and M. Octaviano Pratama. 2018. Classification for Multiformat Object of Cultural Heritage using Deep



- Learning. In *3<sup>rd</sup> International Conference on Informatics and Computing (ICIC)*.
- Orestis Papakyriakopoulos, Simon Hegelich, Juan Carlos Medina Serrano, and Fabienne Marco. 2020. Bias in Word Embeddings. In *Conference on Fairness, Accountability, and Transparency (FAT)*, pages 446—457.
- Shaban Shabani, Maria Sokhn, and Heiko Schuldt. 2020. Hybrid Human-Machine Classification System for Cultural Heritage Data. In *2<sup>nd</sup> Workshop on Structuring and Understanding of Multimedia Heritage Contents (SUMAC)*, pages 49—56.
- R. Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In *31<sup>st</sup> AAAI Conference on Artificial Intelligence*, pages 4444—4451.
- Orion Weller, Nicholas Lourie, Matt Gardner, and Matthew E. Peters. 2020. Learning from Task Descriptions. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1361–1375.
- Zhiquan Ye, Yuxia Geng, Jiaoyan Chen, Jingmin Chen, Xiaoxiao Xu, Suhang Zheng, F. Wang, J. Zhang, and Huajun Chen. 2020. Zero-shot Text Classification via Reinforced Self-training. In *58<sup>th</sup> Annual Meeting of the Association for Computational Linguistics (ACL)*.
- Michelle Yuan, Hsuan-Tien Lin, and Jordan L. Boyd-Graber. 2020. Cold-start active learning through self-supervised language modeling. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*.
- Jingqing Zhang, Piyawat Lertvittayakumjorn, and Yike Guo. 2019. Integrating semantic knowledge to tackle zero-shot text classification. In *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1031–1040, Minneapolis, Minnesota.