Proceedings of the 1st International Workshop on Artificial Intelligence for Historical Image Enrichment and Access (AI4HI-2020), pages 11–15
Language Resources and Evaluation Conference (LREC 2020), Marseille, 11–16 May 2020

Interlinking Iconclass Data with Concepts of Art & Architecture Thesaurus

Anna Breit[1]
[1]Semantic Web Company, Austria; [1]anna.breit@semantic-web.com

Abstract

Iconclass, being a well established classification system, could benefit from interconnections with other ontologies in order to semantically enrich its content. This work presents a disambiguating and interlinking approach which is used to match Iconclass Subjects and concepts of the Art and Architecture Thesaurus. In a preliminary evaluation, the system is able to produce promising predictions, though the task is highly challenging due to conceptual and schema heterogeneity. Several algorithmic improvements for this specific interlinking task, as well as and future research directions are suggested. The produced matches, as well as the source code and additional information can be found at https://github.com/annabreit/vocabulary-interlinking.

Keywords: Ontology Matching, Word Sense Disambiguation, Semantic Enrichment, Iconclass, AAT

1. Introduction

Iconclass (IC) (Van De Waal et al., 1973) is a widely used resource to annotate and describe iconographic content of artworks. Its entities are highly specific, where one IC entry often represents an entire scene. However, due to its narrative and descriptive focus, it is difficult to semantically exploit these entities. An art recommender for example would benefit from a semantic enrichment of IC entities, as it could better understand the content of artworks favoured by the user and thus his preferences. The interlinkage with a more general ontology would increase the semantic interpretability, both for machines and humans.

This work aims to semantically enrich IC data by interlinking IC Subjects — a controlled vocabulary used to annotate IC concepts — with concepts of the Art and Architecture Thesaurus (AAT) (Peterson, 1990). A novel interlinking algorithm is introduced and preliminarily evaluated on a non-expert annotated dataset, yielding promising results. Still existing weaknesses of the proposed system are addressed in an extensive discussion on suggestions for improvement.

2. Data Sources

2.1. Iconclass (IC)

Iconclass is a well established taxonomy-like classification system published between 1973 and 1985 by the Royal Netherlands Academy of Arts and Sciences. It contains iconographic entities which are widely used by museums and art institutions around the world to describe the content of artwork images. The entities are mainly hierarchically organised, where the hierarchy is reflected in the identification code: Iconclass data is subdivided into 10 root nodes — so-called "main-divisions" — with corresponding ID-codes of the digits 0-9. For each level of depth added, the identification code is expanded by either (1) an alphanumeric digit, to introduce a subdivision (a child node with increased specificity), (2) bracketed text, to introduce a specific entity (like a person) as child node, or (3) bracketed text starting with a plus-sign, to "add a 'shade of meaning' to the definition or meaning"[1] via the child node.

IC entities are quite heterogeneous: While the main divisions 1 to 5 describe general topics to represent principal elements of art — such as *44D211 tax payment* — entities whose ID starts with a digit ranging from 6 to 9 are more narrative, describing specific religious or mythological scenes and elements, like *94L3221 the Hydra is killed by Hercules assisted by Iolaus, who sears the roots of the severed heads with burning brands; an enormous crab nips Hercules' foot*. The main division 0 is used for abstract art. Moreover, even entities describing general topics can be both concept-centric like *44D211 tax payment* or action-centric like *34C11 feeding wild animals in winter*.

To further describe IC entities, *Subjects* were introduced. Subjects are tag-like, elements based on a controlled case sensitive vocabulary. IC entities can have Subjects in multiple languages, however, the tags are not interlinked across languages. Moreover, a 1-to-1 matching between languages is not possible, as there is a different amount of tags per language for some Subjects. Even though Subjects are disambiguated, the specificity varies which sometimes makes it hard to understand the range of the intended meaning: The subject *plate* is for example used for *41B2133 hearth-plate*, *41D11 fashion plates*, and *48C6143 plate, film ∼ photography*. Subjects are inherited which means that an IC concept is annotated both with its individual Subject tags as well as with all tags from all of its broader concepts.

2.2. Art & Architecture Thesaurus (AAT)

The Art & Architecture Thesaurus is a ontology describing art, architecture, conservation, archaeology, and other cultural heritage, covering a broad temporal and geographic spectrum. Included are not only entries for objects, but also those describing colours, materials, art-styles and -periods. This wide range of concepts is divided into 7 main facets: Associated Concepts, Physical Attributes, Styles and Periods, Activities, Agents, Materials, and Objects. Facets are further divided into one or multiple hierarchies with a systematic focus. Entries within the AAT hierarchy can have different record types, i.e. *facet*, *concept*, *hierarchy name* and *guilde term*. For most of the facets, only general concepts but no instances exist. For example, while *300417304 sun gods* is in the AAT, the Egyptian god *Ra* will not be found. However, there are exceptions to this rule where named entities are required to describe a concept, e.g. art-styles and -periods or a specific type of furniture.

3. Problem Statement

As described above, AAT and IC, though both being ontologies in the same domain, have very different foci which

[1]http://www.iconclass.nl/contents-of-iconclass

makes matching quite challenging. Not only is the entity overlap of these two resources limited, but also the hierarchical structure of the entries that do have a matching equivalent in the other ontology are fundamentally different. Matching AAT and IC data on concept level would therefore not add much information to the entities. However, Subjects used in IC can be seen as a more general description of the concept which better resembles the nature of AAT concepts.
Therefore, in this work, the two resources are not interlinked by matching their concepts, but IC Subjects are interlinked with AAT concepts to add a layer of semantics. To be more precise, the task of ontology matching of a source and a target ontology is reformulated as the alignment between a controlled vocabulary, which annotates a source ontology, and a target ontology. Due to this restatement, classical ontology matching algorithms — especially based on structure-level matchers — can not be applied intuitively, as IC Subjects are not structured.

4. Related Work

The idea of matching Iconclass Subjects and AAT concepts was first explored by Weda in 2017 (Weda, 2017). He used two different data management and alignment tools, i.e., OpenRefine[2] and Cultuurlink[3], to align the two taxonomies based on lexical features. Weda provided a comprehensive qualitative report on the matching results, allowing insight on difficulties arising with the alignment of the two ontologies. Unfortunately, the resulting matches were not made publicly available.
The first work to explicitly exploit IC Subjects in order to add semantic richness to IC concepts within a real world use case was provided by (David and Kamerling, 2019). The authors presented a recommendation system for artworks based on relevancy scores of the interconnected IC and AAT concepts.In the proposed algorithm, the concepts of the two resources were aligned via IC Subjects, meaning that the IC concept is linked to all AAT conceps, to which at least one of its IC Subjects matches. This means, that the same Subject can be matched with different AAT concepts, depending on the IC concept it was attached to. To create the matches between IC Subjects and AAT concepts, first candidate matches were created using a simple string matcher. Then disambiguation was performed using a proprietary algorithm, where a match is considered correct if at least one of the other IC Subjects assigned to the Iconclass concept could be found in the hierarchical AAT parent path, or, if only one candidate match was found. Unfortunately, this algorithm produced many false positive matches.
In order to improve the interlinking of IC and AAT data this work presents an algorithm for matching IC Subjects and AAT concepts directly. This gives the possibility to better understand the meaning of an IC Subject, by being able to analyse the IC concept it analyses. Herefore, a sophisticated disambiguation algorithm is introduced, which classifies each candidate match independently and therefore is capable of producing an arbitrary number of matches for each instance.

[2]http://openrefine.org/
[3]http://cultuurlink.beeldengeluid.nl/app

5. Matching strategy

As mentioned in 2., the specificity of the meaning of IC Subjects varies among the tags, resulting in IC Subjects aggregating the meaning of several AAT concepts. Therefore, the matching algorithm must be able to take one, multiple and no correct matches into account. Also, Word Sense Disambiguation has to be performed, as various terms appear more than once with different meanings within the resources. For example, the term “craft” exists as an English IC Subject, while AAT contains four concepts having “craft” as prefLabel or altLabel: *300212527 aircraft, 300212528 spacecraft, 300042940 watercraft* and *300054704 crafts (art genres)*. To determine which of these AAT concepts actually fit to the IC Subject *craft*, some kind of context for both the IC Subject and AAT concepts is necessary. By comparing these contexts, disambiguation can be performed and a decision on which are correct matches can be made.

5.1. The algorithm

Let $\{s_0, s_1, s_2, ...s_n\}$ be the elements of the source resource S to be matched to the target resource T. Each s_i has zero, one or multiple matching candidates $\{t^i_0, t^i_1, ..., t^i_m\} \in T$. The aim is to disambiguate the matching candidates in order to identify correct matches.
For each $s_i \in S$, three different levels of context C^1_i,C^2_i and C^3_i can be defined, where C^1_i corresponds to the narrowest context which best describes s_i, while C^3_i corresponds to the broadest one having only a distant relation to s_i. Each element $t^i_j \in T$ is associated with a context Z_j. These contexts are dependent on the resources to be matched and must be defined by the user. For deciding whether the candidate t^i_j should actually match with s_i the overlap between their contexts is calculated as follows:

$$D(s_i, t^i_j) = \frac{\sum_{k=1}^{3} \alpha_k \cdot \phi(C^k_i, Z_j)}{\sum_{k=1}^{3} \alpha_k \cdot \sigma(C^k_i)}$$

with

$$\phi(X,Y) = \begin{cases} \sqrt[\lambda]{\frac{|X \cap Y|}{|X|}}, & if|X| > 0 \\ 0, & otherwise \end{cases}$$

and

$$\sigma(X) = \begin{cases} 1, & if|X| > 0 \\ 0, & otherwise \end{cases}$$

$\alpha_k \in [0, 1]$ and $\lambda \in N$ are hyperparameters. The resulting $D(s_i, t^i_j)$ corresponds to the disambiguation value. If it is higher than a certain threshold h, the tuple (s_i,t^i_j) is considered a match.

6. Experiment

A preliminary experiment was performed on matching IC Subjects and AAT concepts in order to achieve a first insight on the suitability and remaining challenges of the proposed matching procedure. IC and AAT data was extracted in October 2019. The produced matches of all English IC Subjects, as well as the source code and additional information can be found at https://github.com/annabreit/vocabulary-interlinking.

6.1. Setup

For matching IC Subjects S to AAT concepts T , the following contexts were defined:

- C_i^1 **: Content in label parentheses.**
  Subject labels may contain additional content in parentheses which adds additional meaning to the label. This content helps to disambiguate the label either directly e.g. *square (shape)*, or by further describing it (*Deadly Sins (Seven)*, *left (opposite to right)*).

- C_i^2**: Direct sibling Subjects.**
  This context contains Subjects that were co-assigned to the same IC concept as the Subject of interest s_i. However, only the individual Subjects of the concept are taken into account, inherited Subjects are filtered out.

- C_i^3**: Inherited sibling Subjects**
  This context contains Subjects that were co-assigned to the same IC concept as the Subject of interest s_i. However, only inherited Subjects are taken into account.

- Z_j**: Broader concepts**
  For all "preferred broaders" of the concept of interest t_j^i, prefLabels and altLabels are added to this context.

Matching candidates were collected by performing language-aware case-insensitive exact string matching for lightly pre-processed subject labels of IC and prefLabels and altLabels from AAT concepts. Pre-processing of Subject labels consists of the removal of content in parenthesis. The hyperparameters were experimentally set to, $\alpha_1 = 0.9$, $\alpha_2 = 0.8$ and $\alpha_3 = 0.7$, while λ was set to 5. The threshold parameter h was chosen to be 0.2. The decision of using such a low threshold is based on the known different structure of the two ontologies. The overlap between the contexts will be small, especially for Subjects that are also used in IC concept describing narrative content. For example, *94L3221 the Hydra is killed by Hercules assisted by Iolaus, who sears the roots of the severed heads with burning brands; an enormous crab nips Hercules' foot* has 16 inherited Subjects and 5 individual ones, including *crab*. Here, *crab* will collect a lot of context Subjects which will most likely not help to disambiguate, such as *foot*, *history* or *twelve*. Therefore, already a small amount of matching elements in the different contexts can be seen as a strong indicator of an actual match.

6.2. Evaluation Set

To estimate the matching quality and to get an impression of remaining problems, an evaluation set was created and manually evaluated by two non-experts. As the focus lays on the disambiguation capability of the matching algorithm, the evaluation set consists of 100 randomly selected English IC Subjects from those that had multiple matching candidates. Candidate matches were created for these 100 Subjects, resulting in 242 total potential matches to be evaluated.

To create the evaluation set, the annotators were asked to first develop an understanding of the meaning of the Subject by inspecting up to 10 IC concepts which they were assigned to as individual tag (not by inheritance). After disambiguating the Subject label, the annotators looked up the matching candidates from AAT where they were told to primarily use the altLabels and the concepts hierarchy to disambiguate. When in doubt, they were instructed to further use the *Note*s which hold an explanation and in some cases a usage recommendation of the AAT concept. When both a broader and a narrower concept seemed fitting, both matches should be marked as correct.

Comparing the annotations of the two non-experts shows only a very low inter-rater agreement of 37%, which is an indicator of the difficulty of this task. Though the annotators marked about the same amount of connections as correct (45% and 49% of the matching candidates), their annotations still were very different.

For evaluating the performance of the matching system, both the intersection and the union of the connections marked as correct by the annotators were created, resulting in two ground truths, *union_truth* and *intersection_truth*, consisting of 90 and 138 correct links, respectively.

		Precision	Recall	F1
union	**this**	0.76	0.46	0.57
	all	0.57	1.00	0.73
intersection	**this**	0.59	0.54	0.57
	all	0.37	1.00	0.54

Table 1: Results of the presented algorithm (*this*) compared to an all-true baseline (*all*). As ground truth, *union_truth*(*union*) and *intersection_truth* (*intersection*) were used, respectively.

6.3. Results

The results of the matching evaluation can be found in Table 1. Precision, recall and F1 measure were calculated for both configurations, using *union_truth* and *intersection_truth* as ground truth. As all 1-to-0, 1-to-1 and 1-to-n matches must be taken into account, each connection was treated independently in the evaluation step. This means, that a matching candidate for the subject s_i which was discarded, though it was marked as correct in the ground truth, will be counted as a false negative, regardless of how other candidates for s_i were treated.

A baseline (*all*) was provided which does not consider disambiguation and marks all candidate matches as correct.

6.4. Discussion

The evaluation provided some fruitful insights on the difficulties of the task introduced. First, creating the evaluation set is a particularly challenging task, especially for non-experts. As there is no explicit definition, the meaning of IC Subjects has to be extracted via their assignment to different IC concepts. AAT concepts on the other hand tend to be very precise in meaning with only nuanced differences. In combination with the aforementioned varying specificity of IC Subjects, there is a lot of room for interpretation, which leads to the little agreement of the two annotators over the evaluation set.

The quantitative matching metrics in Table 1 show severe differences between the *union_truth* setting and the *intersection_truth* setting. Naturally, the precision of the baseline is higher in the *union_truth* setting, as more candidates are considered as correct. This together with the recall of 100% that arises with all-true classifiers creates a very high F1 score, which cannot be topped by the system presented in this paper, even though, its precision is at an acceptable level of 76%. For the *intersection_truth* setup, the presented algorithm can slightly beat the *all* baseline regarding the F1 measure, however both precision and recall are below 60%. This shows, that the cases that were more obvious to the human annotators not necessarily were as easy to distinguish for the system.

Taking a closer look at the produced matches on a qualitative level, different kinds off errors can be distinguished. A false positive match is categorised as "hard error" when the two entities that are aligned have completely different meaning, e.g., when the IC Subject *opening* which is for example assigned to the IC concept *31G332 opening of the book of life* is matched to the AAT *300002765 concept openings (architectural elements)* which describes *"apertures or breaks in the surface of a wall"*. "Soft errors" on the other side appear, when the (falsely) predicted matching entity is semantically close to the true entity or where the predicted matching entity and the true entity could be aggregated to a concept. For example, for the IC Subject *white poplar* (the tree) a proposed match to *white popular* (the wood) would be considered as soft error. Another example is the IC Subject *redingote* connected to the IC concept *41D211(REDINGOTE) dress, gown: redingote*, which is both matched with the AAT concept of the dress (300254632) and the concept of the overcoat (300209851). The evaluation against the *union_truth* resulted in 20 false-positive predictions, where only 8 were "hard error" and 12 were "soft errors".

When evaluating against the *union_truth*, 75 false-negative matches were produced , where over 70% (53) can be backtraced to 36 IC Subjects for which the algorithm did not predict any matches. For the vast majority of these IC Subjects, the created contexts did not overlap at all with the context of any of the matching candidates. This can either be an indicator for poor choice of context, or for the heterogeneity of the two resources resulting in completely different viewpoints and thus unmatchable contexts of the same concept.

7. Future Work

Many different approaches could be taken into account when trying to improve the interlinking of Iconclass Subjects and AAT concepts. First of all, the presented system could be adapted to achieve better results. For example, the defined context could be improved for both taxonomies. IC subjects' contexts could benefit from the removal of siblings and inherited siblings from Iconclass concepts from the main divisions 6 to 9, as these narrative Iconclass concepts add a lot of noise for the disambiguation algorithm. For AAT contexts, also *related* concepts could be added. Furthermore, the matching quality could benefit from adding post matching rules based on the resource knowledge, like Iconclass Subjects representing individuals or gods will not have a match in AAT, or IC Subjects describing trees should not match to wood concepts in AAT. Last but not least, a hyperparameter optimisation could be performed to find more suitable parameters than the experimental choice presented in this work. However, to find the most suitable and impactful actions, further analysis should be performed, starting by investigating the performance of the presented system on those IC Subjects with only one candidate match as well as the matching performance for other languages. Also, the evaluation process could be reconsidered, as false negatives could have too much weight in the current setting due to non-matching contexts. Another approach would be to rethink the matching strategy entirely, for example by adding external knowledge sources, which could potentially overcome the problem of low recall values.

Though IC Subjects exist in different languages, they are not interlinked. These can potentially be exploited in two of the following ways: Either, they can be taken into account during the disambiguation process, as ambiguities often do not persist across languages, or, the created matches can be used to interlink the IC Subjects, by leveraging the multilingual labels in AAT.

A platform and comprehensive interface for collaboratively suggesting, evaluating and correcting potential matches between Iconclass and AAT would offer great added value for the process of enriching Iconclass data. The availability of information that helps understanding the meaning of IC Subjects (e.g., IC concepts they are attached to) would accelerate this task. To facilitate the disambiguation of IC Subjects and to add another layer of semantic richness to Iconclass, IC Subjects could further be connected to a general purpose ontology, such as WordNet (Fellbaum, 1998).

Finally, the suitability for the matching algorithm proposed in this work for interlinking other resources could be explored. Though the approach was developed with the focus on disambiguating and interlinking tag-like objects associated with concepts in a source ontology with concepts of a target ontology, its generalistic definition makes it easily applicable to other data structures.

8. Conclusion

In this work, a matching and disambiguation algorithm to interlink IC Subjects with AAT concepts to improve semantic richness was introduced and its performance was investigated in a preliminary analysis. As ground truth served an evaluation set annotated non-expert. A quantitative analysis of the results shows rather moderate outcomes, though precision is always significantly above the baseline. This highlights the difficulty of the task (both for the algorithm and non-experts). A qualitative analysis provided important insights in remaining weaknesses of the system — especially in terms of recall — while showing that the system is able to produce promising predictions, as only a very limited number of false-positives are considered as "hard errors". Finally, several potential algorithmic improvements and research directions were suggested which are yet to be investigated.

9. References

David, R. and Kamerling, T. (2019). Relevancy Scoring for Knowledge-based Recommender Systems. In *IC3K 2019 - Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, volume 2, pages 233–239.

Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. Bradford Book.

Peterson, T. (1990). Art & Architecture Thesaurus. *Oxford University Press*.

Van De Waal, H., Couprie, L. D., Tholen, E., and Vellekoop, G. (1973). *Iconclass : an iconographic classification system*. North-Holland Pub. Co., Amsterdam.

Weda, R. (2017). Bringing Two LOD Vocabularies Together. In *CIDOC Annual Conference*, pages 1–11, Tbilisi.