A word is worth a thousand pictures

Why the use of Iconclass will make Artificial Intelligence smarter

Table of contents

Hans Brandhorst
Wijngaardenlaan 42
2252XP Voorschoten
Netherlands
jpjbrand@xs4all.nl

1. The purpose of systematic iconography

1.a. History's handmaiden: images as sources of historical information

Of all types of historical sources, images speak to us in the most straightforward, direct way. Even if a picture tells a story we do not know, or contains details we do not completely understand, we are usually still capable of finding words for what we see.
Just have a look at the following pictures. Most people will admit that all four of them contain recognisable representations of lions, even though the lions in the first two pictures are far less realistic than those in the other two.

At the same time, we can all see that these images are more than snapshots of lions in their natural habitat. These pictures were produced under particular historical and cultural circumstances, and for particular reasons. It is our job, as historians, to study them in their context and find out what they tell us about those circumstances and about those reasons.
Obviously, studying images in their context can mean a wide variety of things, but by default it is a research process that moves in two directions. On the one hand we zoom in on the object of study. In the case of these examples that will tell us, for instance, that the first picture is found in a Book of Hours, where it accompanies a prayer to saint Jerome. Identifying the lion as an attribute of saint Jerome helps us to understand its gesture. The lion does not simply paw the saint like a pet dog, but lifts its paw to remind the viewer that saint Jerome pulled a thorn from it. It also tells us that the second picture, from

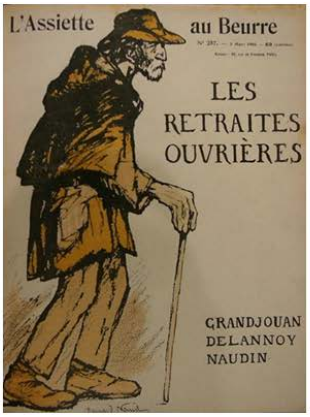
Maerlant's *Der natueren bloeme,* represents the state of zoological knowledge and artistic capabilities in the Low Countries of the 14th century. Zooming in on the third lion we learn from the Rijksmuseum's catalogue that the drawing was probably made by a pupil rather than by Rembrandt himself. More relevant here is the cataloguer's observation that "*the difference between the two eyes is striking – one seems open, the other closed.*[1]" Since the drawing carries the annotation that it was made '*ad vivum*', this raises an interesting question, also in relation to the fourth picture, an engraving by Matthäus Merian depicting a lion in a European landscape with a village and castle in the background. This '*pictura*' was used in several editions of emblem books written by Julius Zincgreff. Zincgreff describes the lion's half-open eyes and its menacing tail as signs of the alertness of a good ruler, who watches over his people even in his sleep[2]. He thus gave a politico-ethical twist to an ancient tradition, as the lion has been described as sleeping with his eyes open since classical times.
The obvious question is whether the lion's vigilance was enough of a commonplace to influence the drawing of a sleeping lion '*ad vivum*' by Rembrandt or one of his pupils[3]. In other words: did this lion really sleep with one eye open or was the artist guided by his awareness of an old tradition?
Only in the rarest of cases do we have primary sources to answer a question like that for a particular image, so the research process moves in the opposite direction, zooming out for a wider view of the field. It needs to assess how likely it is that the artist was familiar with this metaphor, so it needs to gather information about the frequency with which we encounter sleeping lions in visual and textual sources. It should not limit itself, however, to tracing this specific iconographic motif. It should also attempt to collect data about the more abstract themes like Vigilance, or the Virtues of a Ruler, of which the vigilant lion is only one expression. Typically then, the research process moves back and forth between various levels of abstraction. Our observations trigger new ideas and questions, which in turn guide new observations or help us correct older ones.

1.b. Subject, theme and motif: the scope of systematic iconography



[1] by the author of the Rijksmuseum's web catalogue entry Peter Schatborn
[2] the motto '*Parte tamen vigilat*' is quoted from the description of Argus in Ovidius' Metamorphoses Book I.
[3] In his painting manual (1670) Willem Goeree stressed the importance of drawing animals 'ad vivum': "*It is of great importance to seize the chance to see the rarities, such as lions, tigers, bears, elephants, camels and such beasts as one seldom sets eyes upon, and which one nonetheless needs to use in one's inventions from time to time*"

Research in the history of visual culture concerns itself with small, concrete details, such as the half-open eyes of a lion, but also with broad, abstract concepts, such as the “Good Government” it can symbolize. So, pictorial information systems should cater for the same, very wide spectrum, as there is no a priori reason to assume that researchers will only be interested in information at a particular level of detail or abstraction. Social categories like Old Age or Poverty, for example, can be a theme in their own right; they can also be an implied element in images that convey a moral lesson, for instance about Christian charity. In the three medieval images above we see poor Lazarus being refused entry to a rich man’s dinner party, a lame beggar being miraculously healed by saint Peter, and a third beggar receiving half of saint Martin’s mantle. Poverty in all three stories is related to physical handicaps, as is visualized by, among other things, the crutches these poor men are all leaning on.
Poverty and disabilities are also at the core of the next set of images. The barefoot old man warning the boy with the hammer not to do as he did, i.e. spend all his money at a young age; the bright young student held back by lack of funding; the elderly labourer who could never save enough for his pension; the halt and blind working as a team: all are of potential interest to a wide range of historians.



What is important for the mission of systematic iconography is that details, narratives and abstract themes are all part of the same continuum. The rattle of Lazarus, the bare feet and crutches of the poor and the old, the wings and the stone pulling at the student’s arms; the stories of Lazarus, saint Peter and saint Martin; Christian charity, Frugality, a decent Old Age, Mutual Helpfulness: exactly how they play their role in the continuum of visual communication will be evaluated time and time again by new generations of researchers. However, they all fall within the scope of systematic iconography, they can all be objects of study and therefore they all are aspects of historical source material that should be available for retrieval in our online databases.
Depictions of narratives, whether they are biblical, hagiographical, mythological or historical, are a natural target for iconographical explanation. We analyse those images in order to identify the stories they may tell, or to interpret the moral, political or religious messages they aim to convey.


Of the Mouse. 505

Scaliger.

Albertus.

10

ding Mice are sometimes blackish, sometimes white, sometims yellow, sometimes broune and sometimes as the colour. There are White Mice amonge the people of *Sauoy*, and Dauiphin in France called *Alobroges*, which the inhabitants of the country do beleeu that they feede vpon snow. But the white Mouse is aboue all other most lacciuious and leacherous, and therefore it came into a prouerbe, *Mys Leucos, Mys Cacos*, the white Mouse is an ill Mouse, of whose lust *Alcisatus* made this emblem;

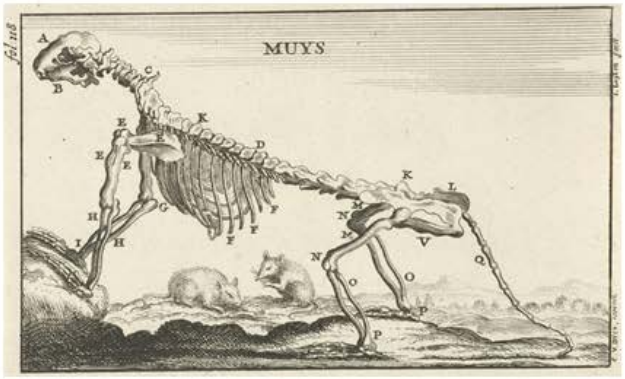
The quantity and seueralty parts of Mice.

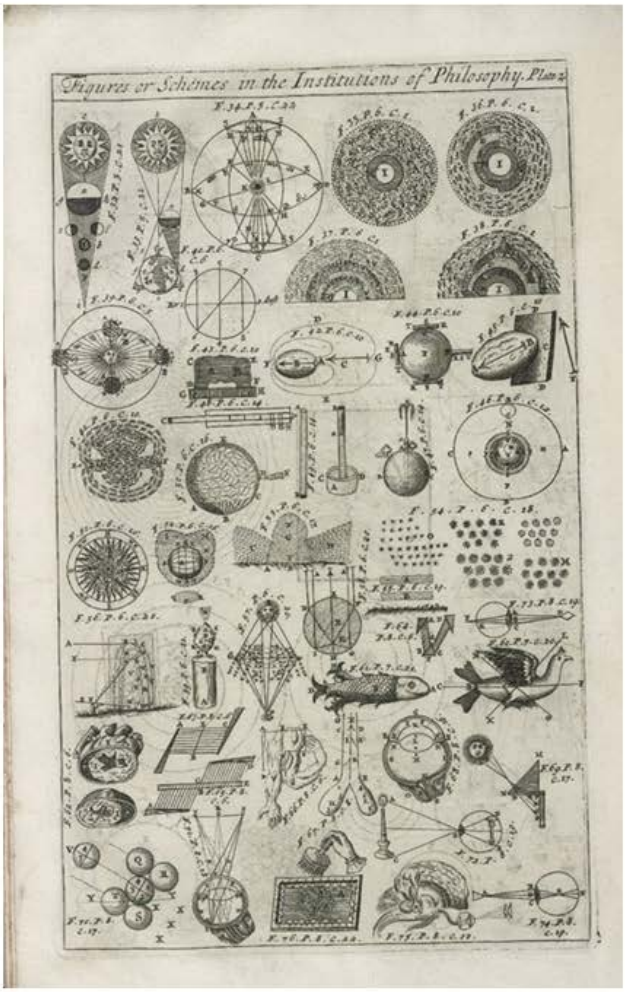
Matharolus.

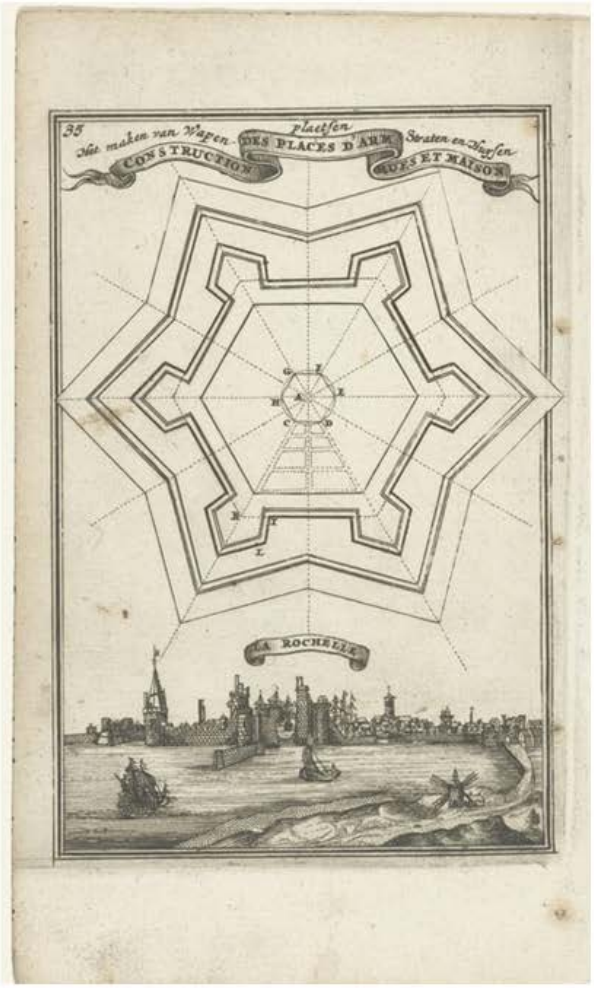
Many images, however, are used for more straightforward scientific or illustrative purposes. These may range from the medieval bestiary to which the 14th century lion quoted above is related, or from the early modern

studies on zoology which mix contemporary observations with classical tradition, to the more ambitious anatomical studies of the skeleton of a mouse, and to the illustration of astronomical and mathematical concepts, or the practical studies of architecture and fortifications.

All of these are potential sources of historical information. But until we systematically describe their iconography, they lay dormant in our libraries, museums and archives – raw data, but not yet information.







2. The lion's trail

Systematic iconography which transforms the raw data of our observations into information, is a labour-intensive task. To fathom the content of pictures is a fascinating intellectual game, but producing words for our observations is a time-consuming process. As with all laborious processes, people always ask themselves whether technology can be used to do the job with less effort and better results. By default the answers to this recurring question have been in sync with the evolution of technology itself. Technical limitations impose natural restrictions on our ambitions. The most recent answers, however, appear to be radically different from the older ones. In our field - the research of visual culture - developments in Computer Vision and Artifical Intelligence seem to point to a very near future where computer software will retrieve pictures based on an automatic analysis of their content. The comparison of visual characteristics which is the basis of modern pattern recognition software suggests that content searches with the help of *automatically assigned* words, or even *without* words, is just around the corner.
Facial and object recognition software unleashes many ideas on how this technology could alleviate the task of the cataloguer and benefit the researcher of visual culture. But before we, as students of visual culture, leave the field of systematic iconography to software and computer scientists, we should put some of those ideas to the test – so let us return to the lion.

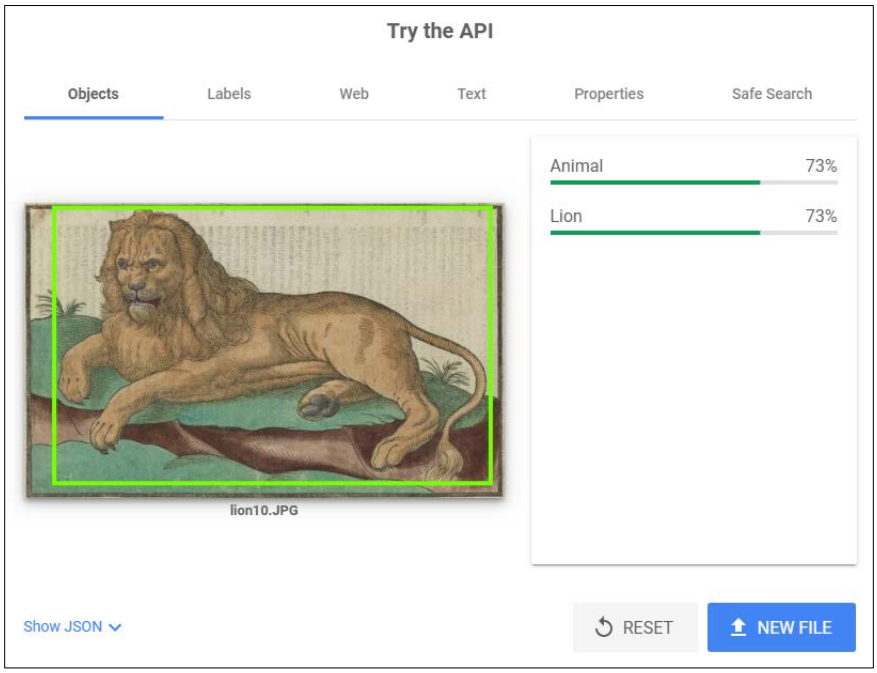
2.a. The shape of the lion

I started with the assumption that most humans would recognize the lion in all four of these pictures without much effort. At the same time, however, I also assume that identifying the areas in these bitmaps that represent the shape of a lion, will not be a trivial task for computer software. The different postures of the animals, the varying amounts of picture space they take up, the varying levels of realism of the draughtmanship, and the 'unnatural' elements surrounding the lions, are factors the algorithm[4] will have to take into account when attempting to isolate the lion's shape. When putting this assumption to the test it seems to make sense to use a slightly larger sample. Hence the 12 pictures on the next page. These pictures of lions were created between the early 13th and the 20th century. They are randomly selected from the circa 15,000 objects in the Arkyves database retrieved with the keyword '*lion*'. The meaning and the context of most of these pictures will not be self-evident, and we shall come to that aspect later. For now, let us assume that the shapes of the lions in these additional samples will also be fairly easily spotted by human observers. Possible exceptions: the pictures in the second row, since they represent a lion's skeleton, a lion with a piece of cloth over its head and a lion drawn rather awkwardly by the early Gothic miniaturist. You can test my assumption about the ease of identification for yourself: just have a look at the individual pictures and ask yourself whether you would indeed have difficulties spotting the lions.

[4] I am using the singular, aware of the fact that many algorithms are at work.


Omnia vincit amor
Ambitiosa Temeritas.
MAIDS
PVGNO PRO

This private test will of course have a private answer only relevant to you. But whether computer technology would also be capable of identifying those shapes is a question relevant to every student of visual resources. Fortunately this can be tested in a less private fashion. We can offer these pictures to a state-of-the-art piece of software, i.e. to Google's Cloud Vision API[5], which is freely available for online testing. The Cloud Vision interface is simple: you just drag and drop a picture on a window and the software will attempt to interpret its subject matter using several strategies. This first screenshot shows the result of the *object detection* part of the algorithm when we feed it the coloured engraving of a lion from Edward Topsell's *The history of four-footed beasts*[6]. As you can see in the same screenshot, Cloud Vision also creates various other types of metadata for this picture.[7] The most relevant ones are **Labels** and **Web Entities** as listed below.

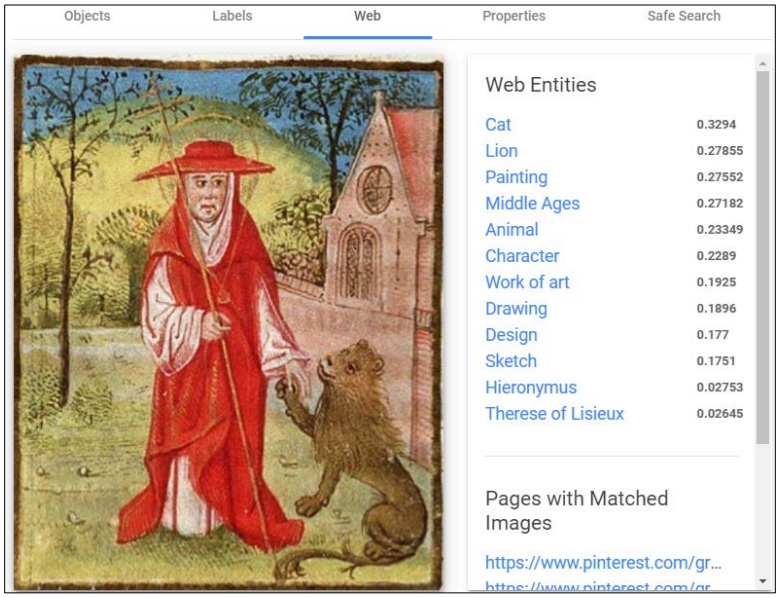


Labels	Web Entities
Lion 97%	Lion 1.73076
Felidae 91%	Dog 0.86641
Big Cats 85%	Cat 0.3711
Wildlife 84%	Mammal 0.2955
Masai Lion 84%	Terrestrial animal 0.28389
Carnivore 84%	Art 0.27939
Art 74%	Canidae 0.25896
Organism 72%	Big cat 0.2483
Illustration 62%	Fauna 0.24037
Ancient Dog	Animal 0.1823
Breeds 62%	Meter 0.1797
Cougar 57%	
Drawing 54%	
Painting 50%	

Labels and **Web Entities** are categories the algorithm detects in the image. They are linked to Google searches so clicking on a **Label** or a **Web Entity** will perform a search triggered by that particular word. The Entity search uses the term for a search in Google's Images database.
While the algorithm that detects these features will do its job in a more objective[8] way than a human individual, its output will still be influenced by regional, local, and individual variables that may affect the search result[9]. Those with more insight in statistics than I have– and there is a 90% chance that you do – will have a better understanding of the percentages and the numbers shown next to the labels and entities, but in this first example the implication is clear enough: Cloud Vision's best

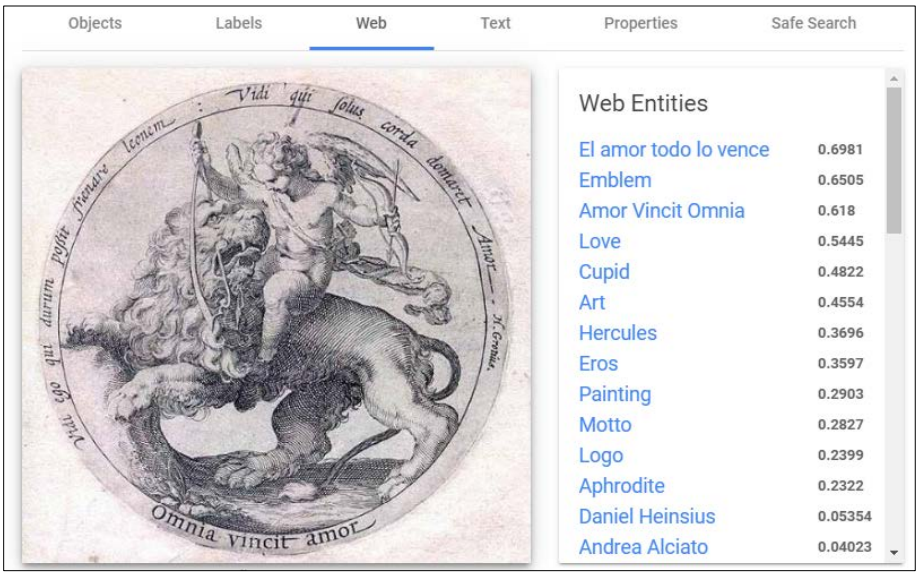
[5] I have also tested Amazon's *Rekognition* software - https://aws.amazon.com/rekognition/ - but with our material the responses were too generic to be of use.
[6] Published in London in 1607, p. 457, from the copy at the Folger Shakespeare Library.
[7] You may also note that I used 'lion10.JPG' as the filename for this photograph. I have tested the same picture with different filenames – with the same results.
[8] Perhaps 'a more mathematical way' would be a better expression …
[9] Repeating the analysis for the same picture over time or on various locations can thus seriously affect the suggestions the software returns.

guess by far is that this is the picture of a lion. Furthermore, the software also relates the 'lion object' to concepts of a higher level of abstraction, such as *carnivore, mammal, animal* and *fauna*. The picture is furthermore recognized as a work of art, and the concepts *illustration, drawing,* and *painting* are assigned to it.



The Object Detection in this miniature of saint Jerome returns '*person*' and '*hat*'. The Labels it assigns - *painting, art, prophet and miniature* – are very generic, with the exception of '*prophet*'. The Web Entities are also a list of broad topics but the list contains the very specific – and correct - '**Hieronymus**'. This is not because the algorithm 'magically' interprets the visual properties of the miniature, but because the similarity search algorithm was able to find precisely this miniature[10] on "*Pages with Matched Images*" and then read back to us some of the accompanying metadata it also found on those web pages. Among those are pages that identify the miniature as "medieval" and "saint Jerome", also naming the lion as an attribute. Thus we find Hieronymus[11] among the identified Web Entities, but also *Cat*, *Lion* and *Middle Ages* – and the 19th century saint *Thérèse of Lisieux* …

The success rate of this hybrid, two-step strategy – first locate copies of the same or a very similar image, and then make the most of the surrounding words – depends on the size of the database and the density and quality of the metadata and additional information exposed to Google. The next example illustrates the same principle. In this emblem – *Omnia vincit Amor* - by Daniel Heinsius the tamed lion is used to demonstrate the power of love - quite literally, as Love (Eros, Cupid) rides a lion with the help of reins and a bit. The pattern matching part of the software retrieves multiple web pages that include this particular picture. Then Web Entities are inferred from the metadata describing and interpreting the emblem on those pages. That is why the Entity list starts with the Spanish translation of the motto, which the image does *not* contain, and why the word order of the Latin motto reads 'Amor Vincit Omnia' although in the



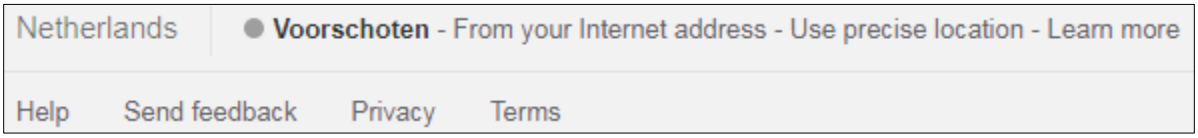
[10] The Hague, Koninklijke Bibliotheek, MS 135 G 19, fol. 129v

[11] "Saint Jerome" is apparently normalized to "Hieronymus"...

engraving it reads ‘Omnia vincit Amor’. It is also why *Hercules, Aphrodite, Daniel Heinsius* and *Andrea Alciato* are found in the list, even though they are not properties of this particular emblem, let alone *visual* properties. Obviously, these entities are associated with the domain of emblems in one way or another, but it is crucial to understand that they were not *created* as metadata by an algorithm. On the contrary. First, words were written by human authors, and only in a second step those words were analysed, ranked and linked to the image by algorithms.

As a consequence of this procedure the lists produced by the Cloud Vision software are hybrid lists that sometimes look quite haphazard. They mix words from different domains and of different levels of abstraction like *Cupid, Painting* and *Art*; they slip in ‘commercially’ relevant words like ‘*Logo*’; and they do *not* include words that would have been important keys for the retrieval of the image for students of visual culture.

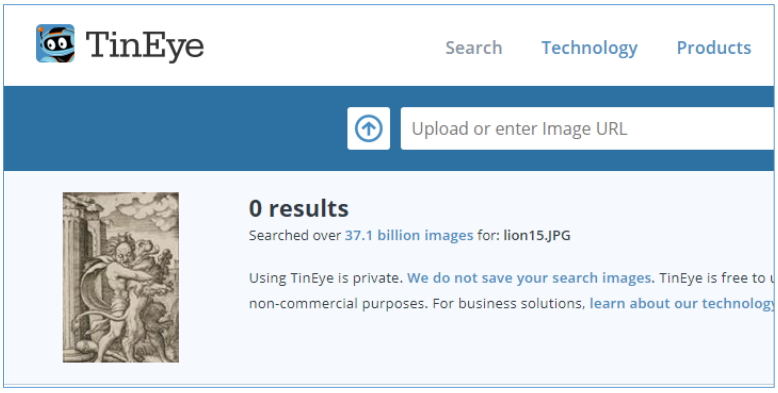
The role of contextual metadata, mostly assigned by humans, is particularly striking when the same picture is offered to the Cloud Vision algorithm at different locations. This emergency coin, made of pressed layers of paper, showing a heraldic lion at its centre, was tested twice against the Vision algorithm, once in Voorschoten, when I was working at home, and then later on the same day while at the library of Erasmus university in Rotterdam. The results were quite different. None of the specific properties attributed to the coin shown in the screenshot – e.g. *Siege of Leiden, 3 oktober 1574, War of Independence* – appeared when testing the software in Rotterdam. Apparently the algorithm’s interpretation of my first location – Voorschoten is close to my internet provider’s Leiden network node, as established by my Chrome browser – added some information to the ensuing search and made the result much more specific.



Google’s algorithm is not ‘open source’, so we have to infer how it operates from circumstantial evidence. But there can be little doubt that here too, the initial match is a visual one – the medal is included in the Rijksmuseum’s webcatalogue and shown on several webpages that mention the 1574 Siege of Leiden. Verbal metadata found on various webpages were then used to compose the list of Web Entities. This explains, for example why the name of a pub in present-day Leiden appears on the list. The “*Stadscafé van der Werff*” has little to do with the 16th century medal, except for the fact that its name “Van der Werff” is also the name of the famous burgomaster of Leiden who led the resistance during the siege. The “Coin” entity may have been inferred from verbal metadata in combination with the circular shape of the image.

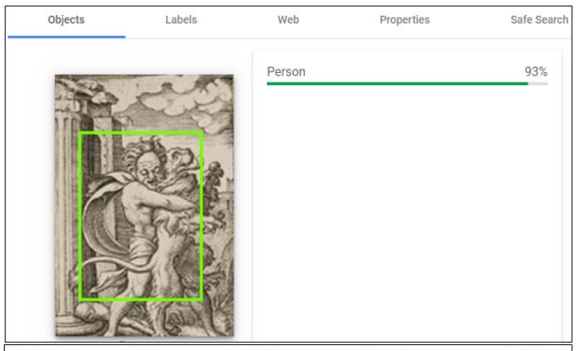
But even though we have to infer the finer details of the algorithm from circumstantial evidence, the general principle is clear enough. As I have argued above, it follows a hybrid procedure, first trying to match an image's visual pattern against Google's image database, and then working with the verbal metadata surrounding the matched images.
This leads us to the question what happens if the first step fails, i.e. if no matching image with helpful metadata is found. The logical way to test this is to use a picture in a Google Images reverse image search. I did this with these two engravings of the story of *Hercules and the Nemean lion*, the one on the left by Virgil Solis (dated between 1534 and 1562) and the one on the right by Heinrich Aldegrever, dated 1550.

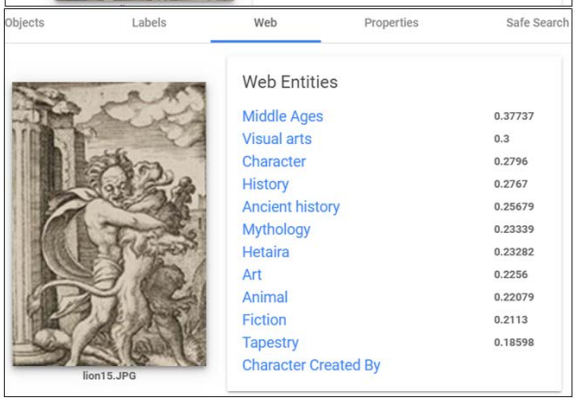
If the reverse image search in Google's database does not find anything, that can mean two things. First and most obvious: the image simply is *not* in the database. Secondly, the reverse image search algorithm has missed a possible match, either because it does not do its job to perfection, or because additional factors – "pictorial noise" - complicate the matching process.
To double-check the outcome of Google's reverse image search algorithm I have also used a well-known alternative reverse image search engine, called **TinEye**. In the case of these engravings the result for both algorithms[12] is identical. I started with the engraving by Virgil Solis, which did not produce a match in either Google Images or in TinEye.



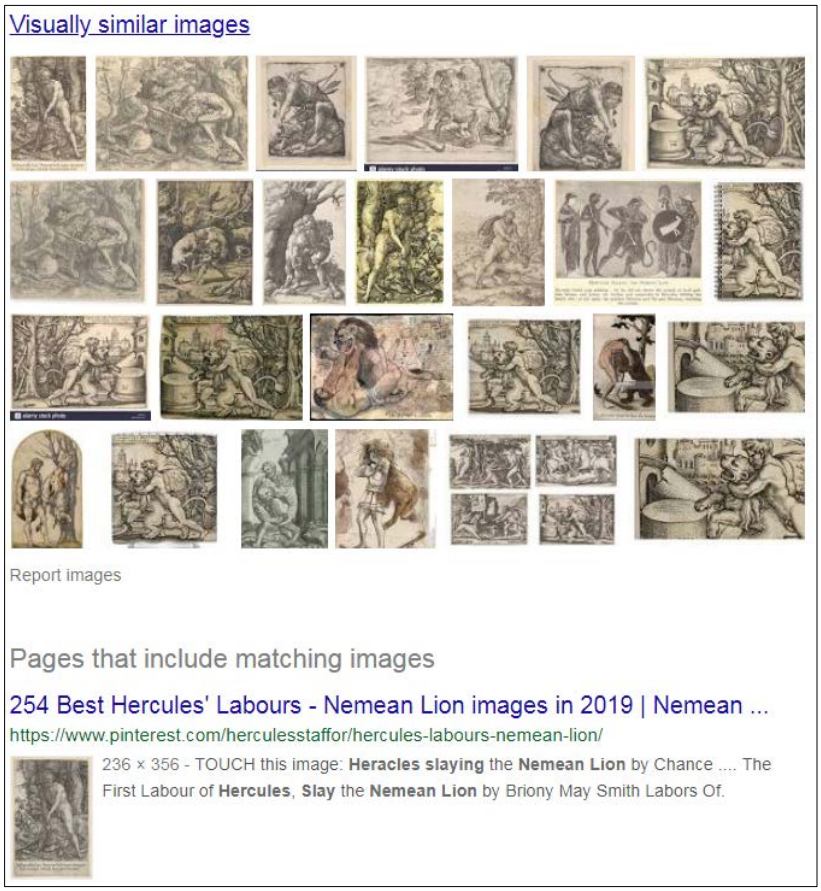
[12] which may actually be only one algorithm used by both companies ...

When no other copy of the image is found, the **Objects** or **Entities** listing cannot use hints from existing subject metadata, so the algorithm has to take a different route. In all probability it relies on a partial comparison and uses the part of the picture of which the outlines are most easily 'recognized'. In this case a "Person" Object is indicated, which is correct but also too abstract to be of much use. The Web Entities are also too generic to be very useful. The only exception is '*hetaira*', a word used for 'prostitute' in classical Greece. Possibly the visual pattern of Hercules' posture – embracing the lion with both arms - has triggered this association.
However that may be, we can safely conclude that without a matching image as a source of subject metadata, the entities suggested by the algorithm are quite unspecific – and the *lion* remains undetected.





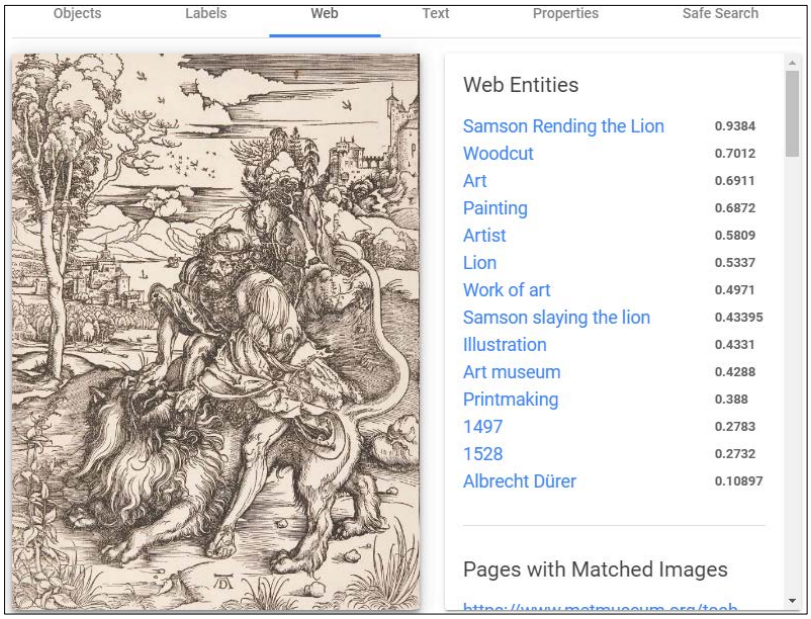
Unlike the engraving by Virgil Solis, the print by Aldegrever does produce matches in the databases of both Google and TinEye. The TinEye interface simply states that the matching algorithm produces "10 results". The numbers mentioned on the Google result page are confusing[13] and best ignored, but the other information is highly relevant to our present topic.
First of all, Google suggests that we use ***hercules slaying the lion of nemea*** as a "*possibly related search*". The suggested query is composed from the verbal metadata found on the pages with matching images. It is no trivial matter to turn the information from those pages into such a neat query. Its pertinence, however, should not surprise us that much. In a sense Google simply reads back to us what we, collectively, have already said about those images.
Secondly, and crucially, the set of "**Visually similar images**" actually demonstrates the power of the underlying *verbal* information. This may not be apparent at first sight as these images do appear *visually* similar and all of them illustrate the story of Hercules and the Nemean lion. That,



[13] at one point the tally read "About 25.270.000.000 results (0,67 seconds) ..."

however, is precisely why it is unlikely that these images were filtered from the database merely on the basis of their *visual similarity*.
If *visual* similarity had been the single or even the main criterium, the selection should have included at least one or two images like the following.

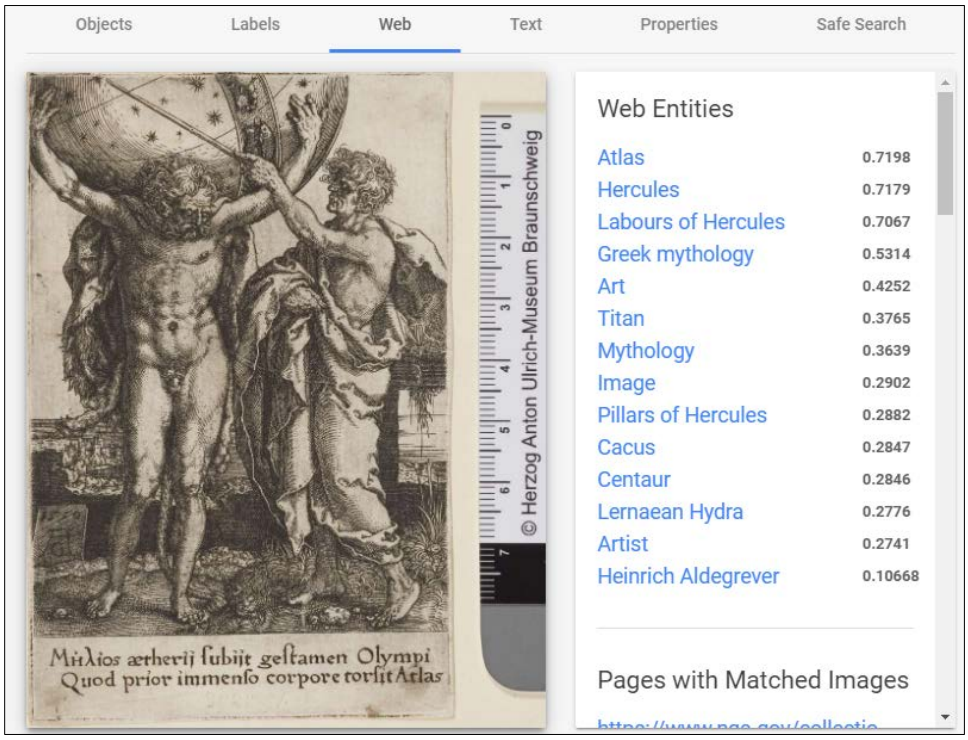
Visually, most of these – and many more like them – are closer to Aldegrever's design than most of the pictures in Google's selection. The reason none of them was included is *not* that they are not in Google's image database, but that their accompanying verbal metadata identify them as representations of a different story, namely that of *Samson tearing apart a lion with his bare hands*. Let us be clear about it: there is no doubt that Aldegrever did intend to show Hercules killing the Nemean lion. In the background the lion is flayed and the print is normally accompanied by a verse identifying Hercules. Yet, he clearly derived his composition – Hercules straddling the lion and gripping its jaws - from a different iconographic tradition, i.e. the one that is based on Judges 14:6 "*The Spirit of the Lord came powerfully upon him so that he* ***tore the lion apart with his bare hands*** *as he might have torn a young goat.*"



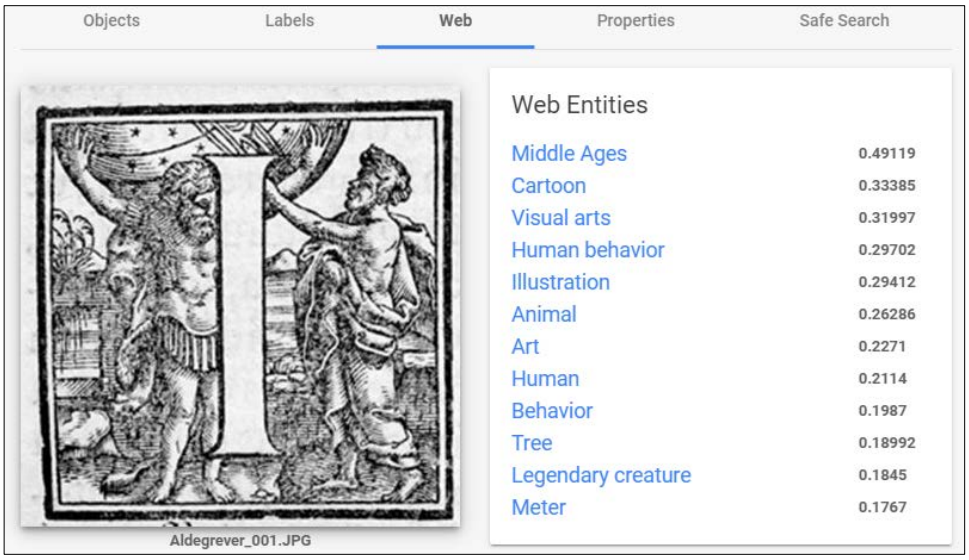
If we do a reverse search for the image at the right – this woodcut by Dürer – Google gathers a comparable amount of "visually similar images". However, they only include representations of the Samson story and none of Hercules and the Nemean lion, a clear indication that "visually similar" is a misleading label. This is confirmed again by the list of Web Entities produced by the Cloud Vision API.

When looking at Aldegrever's Hercules, a human observer probably would need a second look. He would need to notice that the protagonist is naked and that in the background he is pulling the skin off the lion, before realizing that he is actually looking at Hercules and not Samson... At the same time, it seems that the algorithm does not need such a second look, neatly linking the Aldegrever print to Hercules and the Dürer one to Samson.
Now, does this mean that the Cloud Vision API is capable of tagging these images more efficiently than a human observer? Hardly. First of all, its accuracy depends on the assignment of correct subject terms by humans to begin with. Secondly, a simple 'noise' test demonstrates that pattern recognition still has

some way to go before an algorithm will ‘see’ an analogy or a similarity with the same ease as a human observer. Have a look at this picture - again an engraving by Aldegrever, dated 1550, from the same series of the Labours of Hercules. This one shows a scene from the story of the Apples of the Hesperides where Hercules takes the burden of the heavens over from Atlas who is pointing at one of the stars on the celestial globe – probably before fetching the apples. As with the scene of Hercules and the Nemean lion, this engraving is found on many webpages, many of them in German to judge by Google’s “possibly related search”, which reads as **herkules und atlas.**

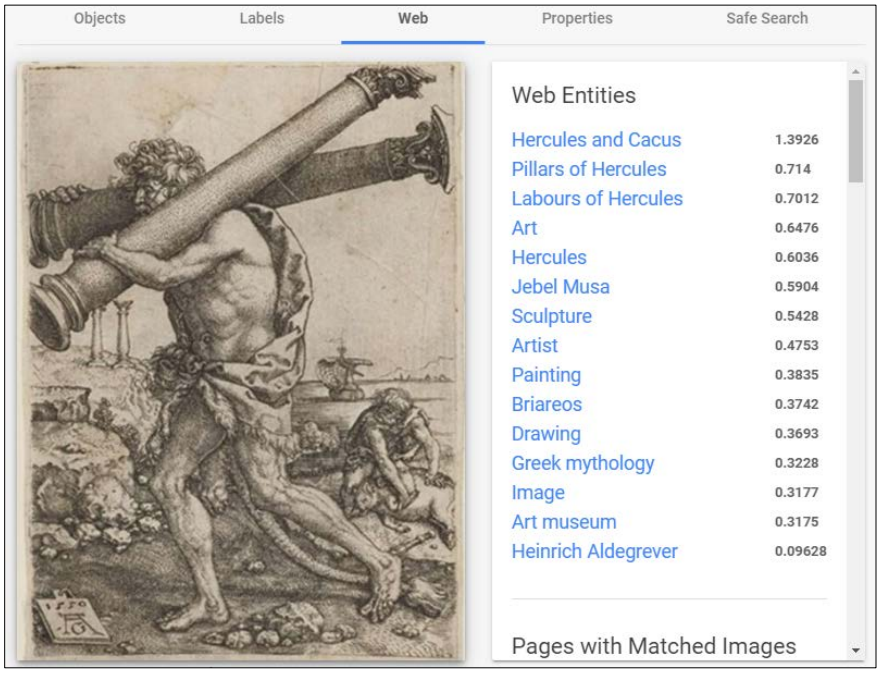


Again, the “visually similar images” are actually images on webpages where first of all the *words* **hercules** and **atlas** can be found, although some are indeed visually similar. The Web Entities suggested by the Cloud Vision API confirm this again. The algorithm may ‘see’ *Atlas* and *Hercules* in this picture, but clearly *Pillars of Hercules*, *Cacus* and *Lernaean Hydra* are inserted from contextual metadata as they are all tags belonging to other Labours of Hercules. Particularly fascinating is what the Cloud Vision API does with this initial, which any human observer will immediately recognize as a copy after the Aldegrever design.

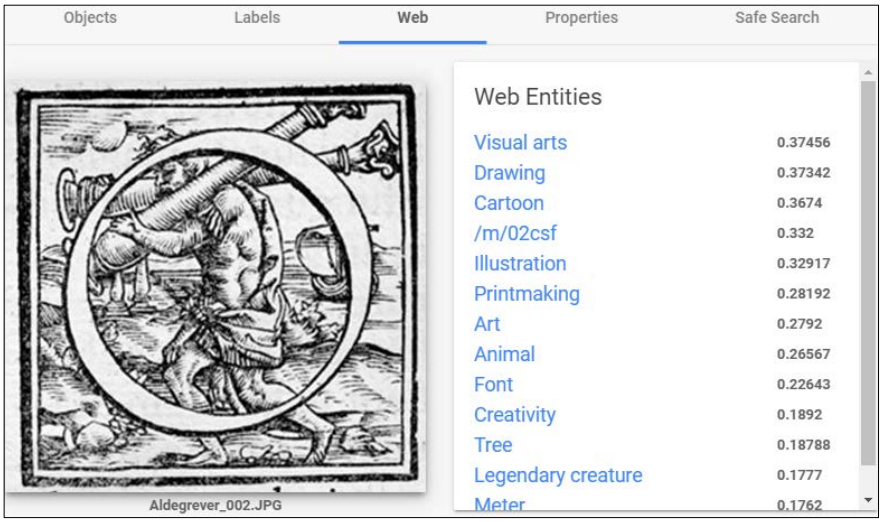


In this case, however, Google’s reverse image search fails to find a match, so the algorithm is in the dark about the iconography and the suggested Web Entities are reduced to very abstract categories. It also means – in all probability - that the letter ‘I’ covering part of Hercules produces enough visual ‘noise’ to prevent the pattern of the historiated initial to be matched with the Aldegrever design. A human observer will easily ‘see through’ the noise.

This is confirmed by a final example, of Hercules carrying the pillars that were named after him. Here too we have a design by Aldegrever and an initial which copies the design. When we 'feed' the images to the Cloud Vision API, the results are almost identical. Several quite specific concepts are linked to the engraving because its pattern can be matched against the Image database. But where the algorithm fails to match the visual pattern – probably becaused of the 'noise' produced by the letter 'O' – it has to fall back on very general terms. Here too no human observer will miss the match, in spite of the minor simplifications of the design or the letter covering Hercules.



All of the above suggests that there is a strong correlation between the verbal metadata Google manages to extract from (webpages with) matching images and the information it offers back to us through its Cloud Vision API. If my conclusion holds water, it implies that the verbal metadata we create are critical for the success of Artificial Intelligence and Machine Learning algorithms. That would also imply that there is a correlation – perhaps even a causal relation - between the *quality of the textual metadata* and the *progress* that can be made in Artificial Intelligence and Machine Learning in the field of content-based image retrieval. Furthermore, if the results of AI and ML in the field of iconography indeed improve ***because** of better subject metadata*, we should perhaps re-assess how we invest our resources or divide our grant money.



2.b. The context and meaning of the lion

In part 3 I shall discuss whether investing in the use of a controlled vocabulary could indeed lead to better metadata. But before I do I would like to return to the issue of the *meaning* of these images of lions, which it is the historian's task to investigate. Can Artificial Intelligence help us there and if so, how? More specifically, can it initiate or stimulate the process of zooming in and out that we go through as we try to understand the meaning and the context of an image?

To answer that question we need a better understanding of how the Google algorithm digests the information it finds on pages with matching images when it composes a list of Entities. If – as I assume - the algorithm has to make the best of what it finds on web pages with matching images, it will have to cope with an element of chance and randomness in sources over which it has no editorial control. We do not have to study a large sample or do a very deep analysis to find confirmation. Just have a look at the Entities list triggered by the Samson scenes in these woodcuts[14]. The lists include captions or titles identifying the subject, one of which is in Dutch:

- *Samson Rending the Lion*
- *Samson slaying the lion*
- *Samson and the Lion*[15]
- *Simson overwint the leeuw*

In addition they contain words that can be regarded as *broader [b] and related [r] terms* for the content of the pictures:

- *Mammal [b]*
- *Human behaviour [b]*
- *Mythology [b]*
- *Legendary creature [b]*
- *Pet [b]*
- *Fiction [b]*

- *Human [r]*
- *Lion [r]*
- *Dog [r]*
- *Horse [r]*
- *Cat [r]*
- *Big cat [r]*[16]

It is unlikely that these words were actually used to describe the pictures on "*Pages with Matched Images*". Although we cannot exclude that words like ***mythology, legendary creature*** or ***fiction*** were found on those web pages, or that the pages did contain references to ***dog*** or ***horse***, it is far more likely that these words demonstrate a core Cloud Vision functionality, i.e. that the algorithm will "*assign labels to images and quickly* ***classify them into millions of predefined categories***". In other words, the intelligence at work here is basically a form of authority

[14] the first one by Albrecht Dürer, c. 1497-98; the second one by the Master of the Amsterdam Cabinet, c. 1470-75

[15] This string is included twice. Also notice that this list of Entities is slightly different from the earlier version I made, at a different location, on a different computer. The earlier listing had only two variants of the title: *Samson rending the lion* and *Samson slaying the lion*

[16] The remaining concepts are linked to or derived from the type of object the algorithm has identified: *Painting, Art, Woodcut, Artist, Work of Art, Illustration, Art Museum, Printmaking, Visual arts,* and *Albrecht Dürer.*

control: the algorithm links actual descriptors like *Samson* and *lion* to the concepts of the classification system that Google uses in the background.
The details of this classification are not public, but it is not very difficult to reconstruct the connections: a *lion* is a *mammal*; it is also a *big cat* and *cats* are *pets*, as are most *dogs*. *Cats* and *dogs* are animals most often related to *humans*, probably closely followed by *horses*. Similar links probably connect *Samson*, a *human slaying a lion*, via *legendary creature* to *mythology* and *fiction*.
Even if this is not in every detail how these concepts are connected, it shows in essence what the algorithm does. It links descriptive words (captions) to predefined concepts and then transforms both into Google queries.
The extent to which this is useful obviously depends on the details. The broader and related terms in the Entities listed for the Samson scene, for example, are too generic to be of much use. Searches for *mammal, pets, legendary creature* or *mythology* will not add to our understanding of the iconography. In comparison, the entities generated for Aldegrever's Hercules scenes are more helpful. Searches for *Cacus*, the *Lernaean hydra*, *Jebel Musa* or the *Labours of Hercules*[17]do point to web pages that could add to our understanding of the meaning of these prints. Still, we should realize that they are based on an algorithmic, not a human evaluation of any of the matching images. If that had been the case – i.e. if human intelligence instead of artificial intelligence had been at work – the background scenes of the *skinning of the Nemean lion* and the *killing of the Erymanthian boar* would probably have been included as "related terms", thus pointing at a recurring feature of Aldegrever's series of the Labours.

This is not the place to analyse the results of this linking process at length or digress into information theory. Suffice it to notice that when the algorithm links a concept like *Pillars of Hercules* to *Hercules and Cacus*, *Labours of Hercules* and *Greek Mythology* it basically uses standard mechanisms to interlink narrower, broader and related terms. We shall return to this mechanism in the next chapter, when we shall be looking more closely at the predefined categories of the Iconclass classification system.
Assigning images to the Iconclass categories so far has never been done by an algorithm. The main question in the background of the following chapters will be whether we can expect algorithms to assist cataloguers to do precisely that: assign Iconclass's concepts to images. A feature of systems that have already used Iconclass seems to hold the promise that pictorial information systems can both help us to

[17] For some (commercial?) reason the Cloud Vision algorithm by default adds the wors 'logos' to the query...

interpret images and speed up the process of subject cataloguing. The feature I am referring to is that of *co-occurrences*. A preliminary example of Samson's lion scene should clarify what I mean.

This illustration of Samson grippng the jaws of the lion, about to tear it apart, is from a German emblem book[18]. Among the texts accompanying the image we find 1 Peter 5 v. 8 "*Be alert and of sober mind. Your enemy the devil prowls around like a roaring lion looking for someone to devour."*

The association of the lion Samson is tearing apart and the devil is thus made explicit and provides us with an example of co-occurring concepts: the Samson story and the lion representing the devil. It is not hard to imagine how this could lead to additional research questions about Christian stories featuring the devil in the shape of a lion and other predators. A vocabulary – like Iconclass - that can help us to discover without much effort images like these, where the devil is pictured as a lion and a wild bear, would then be a helpful instrument .

[18] Theodorus Brunoviensis, Königliches Seelen-Panget, München 1666, page 353

3. The role of the controlled vocabulary of Iconclass

3.a. Less is more: from data to information through organization

To extract historical information from images, we have to describe them in some way. We have to add a layer of language to our objects of study for the practical reason that language is our primary medium for scholarly reflection and communication. The purpose of systematic iconographic description, however, is not that of the classic *ekphrasis*, i.e. to evoke a single image in the mind of a reader. The purpose is to add some order to our observations of multiple images, so they can be used for retrieval and comparison. It is through organization that we transform data into information: "*Organization per se is the intellectual effort that manufactures information out of such raw material as observation … the more the organization, the higher the level of information ...*"[19]
The simple fact that we use language does not guarantee a high level of organization, of course. Languages are extremely flexible, rich and full of ambiguities. But when we reduce the variety of our vocabulary and increase the structure of the language we use for descriptions, we can turn it into an efficient instrument of organization. This is what we do when we use the artificial language of a *systematic classification*. We use a smaller set of words than natural language puts at our disposal, and by organizing them in pre-coordinated hierarchies we limit their semantic richness, turning words into concepts. Paradoxically enough, we produce a higher level of information by reducing the variety of our vocabulary. When we apply these hierarchically organized concepts to images it is easy to see why. Have a look at the next group of details of pictures.

[19] *R.W. Lucky, Silicon dreams. Information, Man, and Machine. New York, 1989, p…*

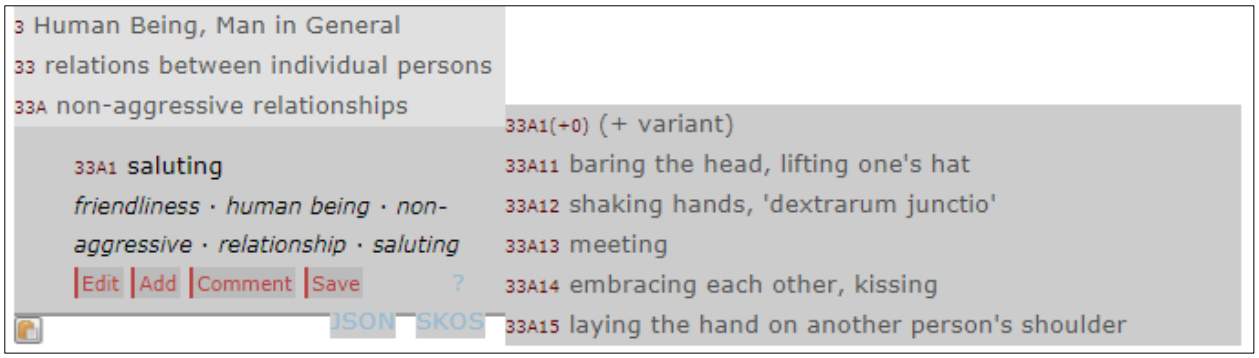
These images were produced at very different times, in different corners of Europe and for very different purposes. Still, they do share a feature distinctive enough to group them: every picture has someone baring his head, whether by removing a hat, holding a hat in his hand, lifting his cap, taking off his helmet, or laying down his crown. If these images would have been described in natural language – even if all descriptions would have been in English(!) - it would probably have been quite difficult to retrieve them as a group on the basis of this common feature.

The reason it was easy to retrieve them as a group is that all these pictures were catalogued using

Iconclass. All it took was the alphanumeric code: **33A11** with which they all had been tagged**.** As shown in this screenshot from the online Iconclass classification system, the concept definition eliminates most of the nuances I have just mentioned, reducing it to '*baring the head, lifting one's hat*'. Eliminating nuances, however, is exactly the point. The purpose of using a classification is to create groups by identifying analogies. Iconclass concepts are like the labels on the drawers of a huge filing cabinet. Tagging an image with a concept is like storing the image in a drawer, using the label as helpful information for our fellow researchers looking for a theme or a motif. By tagging an image with as many concepts as we think are relevant, we store them in many different drawers at the same time. As every concept is also part of an hierarchical chain, tagging an image with a few concepts embeds it in a rich web of analogies and quickly leads to a combinatory explosion of retrieval options.

3.b. Iconclass as serendipity fuel

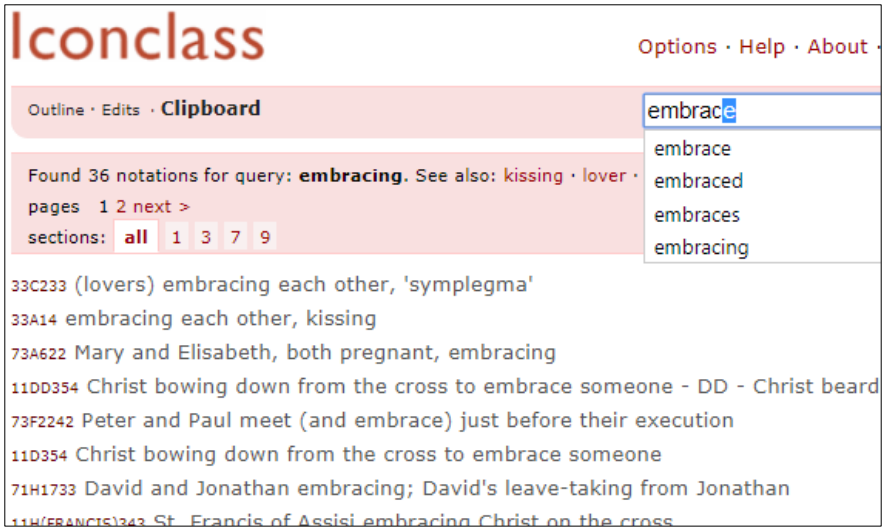
One thing leads to another. As the screenshot from the Iconclass Browser shows, '*baring the head*' is defined as an instance of '*saluting*', which in its turn is an example of '*non-aggressive relationships*', which, again, is a subset of '*relations between individual persons*'. As the next Browser screenshot shows, Iconclass presently contains four more instances of '*saluting*'.



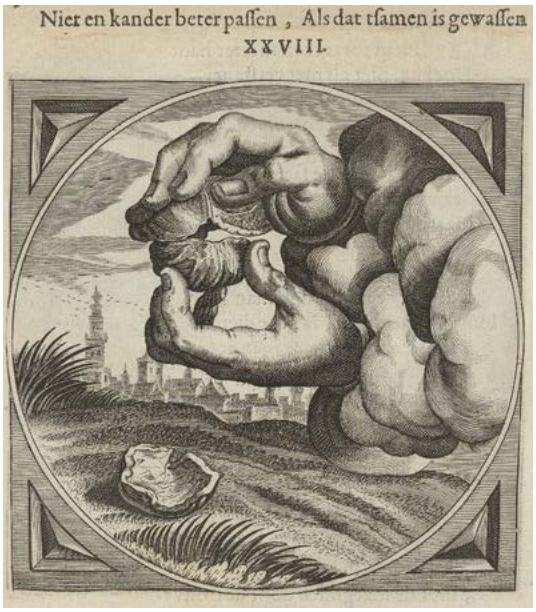
33A1 saluting – to put it somewhat differently - encompasses five ready-made labels for forms of non-aggressive social interaction between humans, including handshakes and embraces.
Applying these concepts as tags serves a modest purpose, comparable to the opening moves of a chess game. Tags are not the *end result* of iconographic research. They are *its beginning*. We assign them to bring our historical opponent into the light. When a sufficiently large number of images have been tagged, a wide field of analogies, similarities, parallels and differences opens up and motifs like handshakes and embraces can be studied in a variety of contexts. From an information system that uses Iconclass, then, these random examples - thanks to the hierarchical structure of the classification -will

automatically be retrieved together with the group of images of '*people baring the head*', because all tags wil start with **33A1**.
This grouping of analogous themes and motifs is a powerful feature of the use of a systematic classification. Still, that is merely a logical consequence of the general principle. It is simply what classifying is supposed to do.
Other features are more specific to Iconclass. The concept '*embracing each other, kissing*' is defined in Iconclass as an instance of '*saluting*', as we have just seen. However, Iconclass also contains quite a few other concepts of which '*embracing*' is an aspect. Here you can see embraces that are an aspect of

themes that are included in Iconclass in their own right: the meeting of Mary and Elisabeth (**73A622**), of David and Jonathan (**71H1733**), of the prodigal son and his father (**73C86461**), and of two lovers (**33C233**) (in this case: Pygmalion and the statue he made).
Even though these images were not tagged with the code **33A14 embracing each other, kissing**, they will still be retrieved with the word '*embracing*', because Iconclass also has an additional layer of keywords, pointing the user from the word *embracing* to the Visitation of Mary and Elisabeth, the departure of David, the return of the prodigal son and the embrace of lovers.



This screenshot from the Iconclass Browser illustrates this principle The word '*embracing*' – here shown with its variants – leads the user to a variety of concepts. These concepts are found in the general parts of the system, where we also found 'saluting'. They are also found among the narrative concepts, taken from e.g. the Bible and Classical Mythology.



Finally, in an actual application, in a rich web catalogue that uses Iconclass for the description of hundreds of thousands of images, many of those images will be tagged with multiple Iconclass concepts. And as one thing leads to another[20] when you query such a database for the word '*embracing*', the algorithm will smoothly direct you - with two clicks – to an emblem that does *not* show an *embrace* but a pair of hands holding two halves of a walnut. The walnut, however, is here used as a symbol of *Reciprocal Love* (Iconclass concept **56F22**) and *Fidelity in Marriage* (**42D30**). The accompanying text explains that the two halves of a walnut fit together perfectly because they belong together and have grown old together. Therefore it is a strong image ('*zinnebeeld*') conveying ideas about marriage and relations relevant for the Dutch society of the 17th century.
There is no embrace in sight, but it is hard to deny that this kind of *organized serendipity* – analogous to Cloud Vision's Entities list - can be a very helpful instrument of knowledge discovery.

3.c. Let's get practical: Iconclass is available, it works and it is free

It took many years and a small army of collaborators to build Iconclass. Its creator Henri van de Waal, who started to work on it in the late 1940's, tragically did not live to see more than the first volume published in 1972. After its transformation from a series of books to a computer system on CD-ROM in the 1990's, it now exists as an open access webservice, which anyone can use for free. Copying concepts from the system to use them as tags in an information system has been made very easy. So, anyone can produce information by applying Iconclass to a collection of images. A substantial group of institutions and individuals actually do. But, needless to say, many more do not. Otherwise there would be no point in explanations like this.

[20] As also illustrated by the earlier example of Samson's lion being interpreted as the devil

The next sections will therefore deal with the online system's basics, so you will understand how you can use it in your own information system or research project.

4. Iconclass computing in the production and retrieval of information

4.a. Producing information with the Iconclass webservice

"Although the repertoire of human concepts is in a sense hierarchical ... it is nonetheless extremely different in nature from the precise and rigid way that concepts are built up systematically and strictly hierarchically in mathematics or computer science"[21]

This quote from Douglas Hofstadter is probably the best summary of the paradox at the core of most humanities computing. It also neatly describes the gap that had to be bridged when the Iconclass browser software was developed. This can be illustrated with the help of an everyday situation of verbal human communication, i.e. that of one person speaking to another person. To depict this situation with the 'silent means' of the visual arts, artists across the world developed a wide variety of pictorial means from the realm of 'visual rhetoric'. Here are a few random examples of one of those variants: one persons putting a hand forward in a gesture of speech.




Below is the Iconclass concept that can be used as a label for this gesture, with its hierachical chain of concepts starting at '*Human Being, Man in General*'.

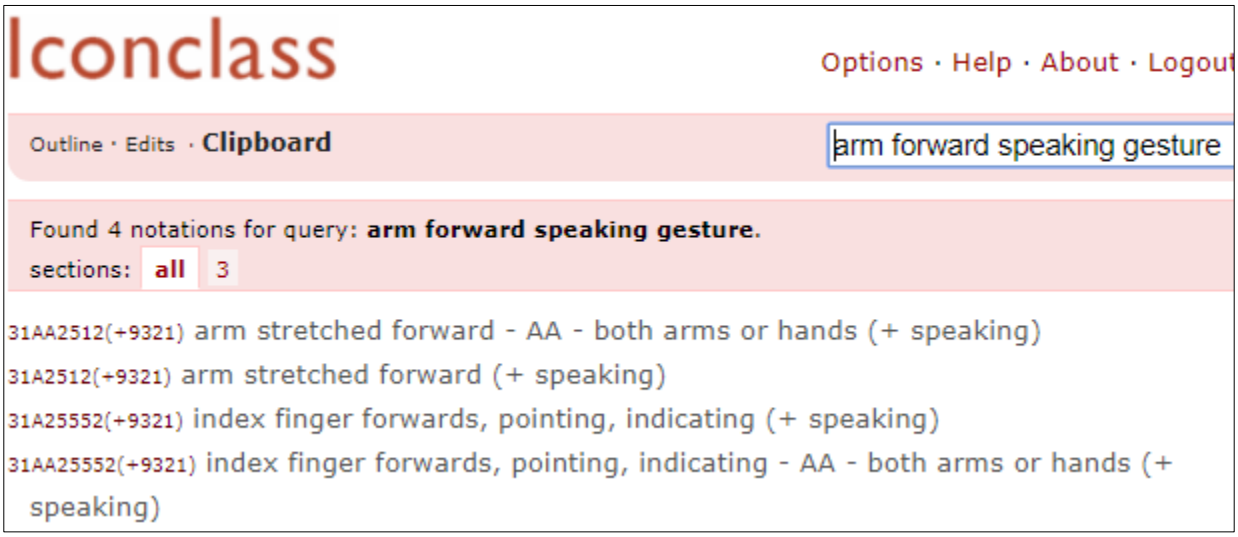
3 Human Being, Man in General
31 man in a general biological sense
31A the (nude) human figure; 'Corpo humano' (Ripa)
31A2 anatomy (non-medical)
31A25 postures and gestures of arms and hands
31A251 postures and gestures of arms and hands in general
31A2512 arm stretched forward
31A2512(+9) arm stretched forward (+ expressive connotations)
31A2512(+93) arm stretched forward (+ relations with neutral character (expressive connotations))
31A2512(+932) arm stretched forward (+ addressing)

31A2512(+9321) arm stretched forward (+ speaking)
Corpo humano · Ripa · addressing · anatomy · arm · arm stretched · biology · body · expression · forwards · gesture · hand · human being · human figure · man · nude · posture · speaking · woman

[21] D. Hofstadter & E. Sandler, 'Surfaces and Essences. Analogy as the fuel and fire of thinking' (Basic Books, New York, 2013), p.54.

4.a.i. Computing and the Iconclass schedules

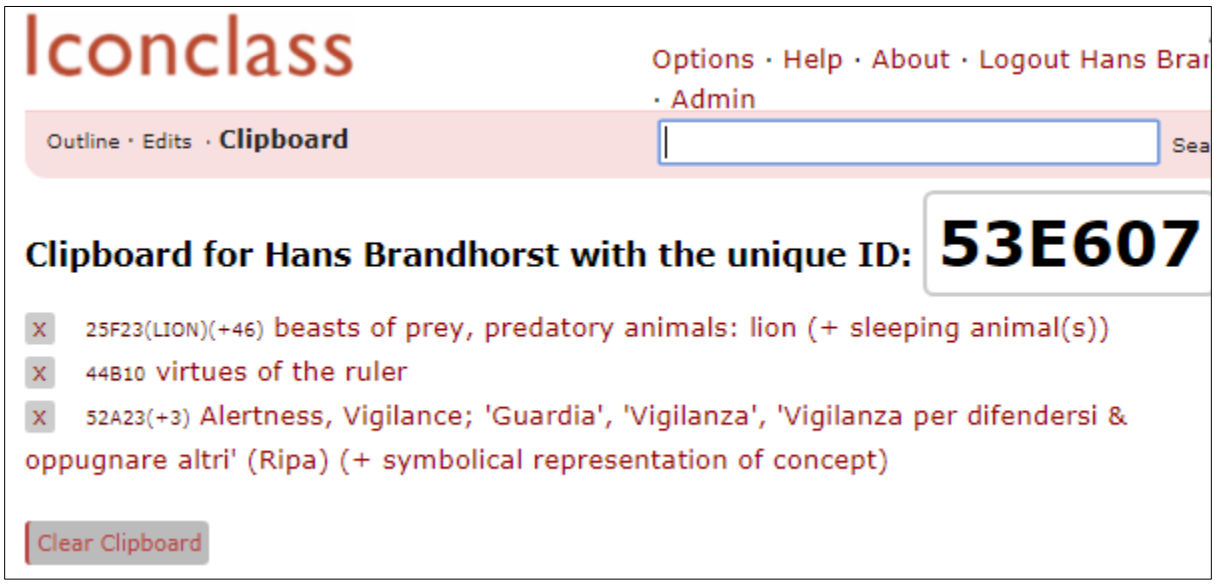
To follow this chain of concepts from '*Human Being, Man in General*' to '*arm stretched forward, speaking*' comes natural to a human observer. Just start reading at the top and with every step down the alphanumeric notation grows with a letter or digit and the concept definition increases in specificity. Also notice that halfway down the list the shape of the notation changes a little: a digit between brackets preceded by a + sign is added: **(+9)** , **(+93)**, etcetera.
To write a computer algorithm that mimicks this behaviour of the human eye (and brain) is a challenge, to put it mildly. Especially if you realize that '*arm stretched forward*' is one out of nine different pre-defined gestures and '*speaking*' is only one of well over a hundred possible connotations, all of which can be combined with those gestures. Exactly which connotation can validly be combined with each gesture is determined by an internal 'grammar' of Iconclass rules.
Similar secondary hierarchies ('**Key lists**') are made available in many different places in the Iconclass schedules. They are an important feature of the system, but they are not the only tool in the Iconclass toolbox. Other instruments, which it would take us too far to discuss here[22], ensure that the core set of circa 30,000 concepts can be combined and expanded virtually *ad infinitum*, but at least to 1.5 million concepts.
Although the purpose of these tools is easy to grasp and they are easy to work with, it is far from trivial to implement them correctly in Iconclass browser software. Van de Waal had high expectations of computers, already in the late 1940's, but the Iconclass toolbox and the Iconclass 'grammar' were created in an exclusively paper environment. The original editors of the system were happily unaware of the effort it would eventually cost to reconcile the Iconclass schedules and the additional features to the demands of binary logic. But precisely because they were unaware, the system could successfully accommodate the 'analogue' logic of the humanities.
Using Iconclass in an environment dominated by card files and print, users were expected to acquire some in-depth knowledge of the schedules before applying the system. In a digital world it has become very much easier to find the appropriate concept through keyword queries, as illustrated by this screenshot, so these expectations have been modified.



[22] Among these we mention 'Bracketed text', 'Structural digits', and 'Systematic references', all of which have been discussed in numerous publications.

4.a.ii. Computing and the online browser

Building the software that correctly interprets the Iconclass 'grammar' and facilitates easy searching and browsing is one thing. Implementing the software to create an online Iconclass system and getting this version accepted as the authoritative Iconclass system is something else. And then getting this online Browser system to be actually used for tagging images is yet another thing.
Over the past two decades the Iconclass Browser offered at the URI http://www.iconclass.org, has been accepted *de facto* as the standard Iconclass system by the international community of users. It is accepted as the digital equivalent of the original printed publication. It is used across the world for cataloguing on a daily basis by circa 100 institutions, whose suggestions and comments have led to an expansion of the schedules with more than 1,500 new concepts, keywords and cross references.
The typical usage of the system consists of looking up the appropriate concepts to tag an image and then copy those concepts to a local file or database system. In addition to the straightforward copy-and-paste method, a dedicated Clipboard function was written which allows users to gather multiple concepts in a list for collective export.

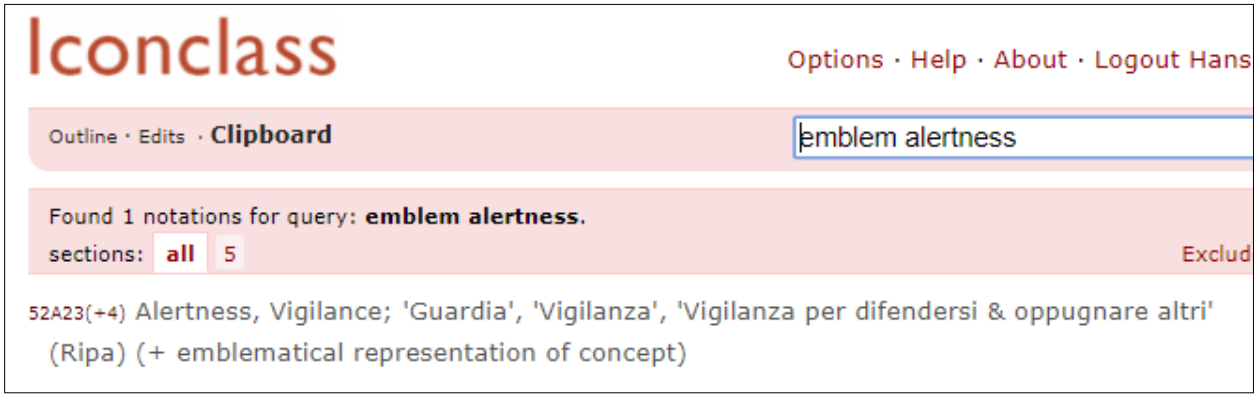


Iconclass is also made available as Linked Open Data, in RDF and JSON formats, so storing the alphanumeric notations suffices – in addition to some basic scripting - to offer end users the textual explanations of the codes in one or all of the Iconclass languages.
Users can register with the system for free, and create their own personal Clipboard with its own, unique ID and URI.

4.b. Solving the authority paradox: retrieving information with the Iconclass HIM service

Anyone using a classification, a thesaurus or some other form of controlled vocabulary for cataloguing, will be confronted sooner or later by what could be called the *authority paradox.* The core of this paradox is that using a terminology authority in essence means selecting specific terms from the authority system and copying them to records in an external database. In this process they are not only lifted from their context and separated from the keywords they are linked to, but – more generally - the instruments that are available to the cataloguer to select the most appropriate descriptors from the

authority system when creating the catalogue entries, are not automatically available to the end user of the catalogue.
Let us look at a specific example to illustrate this. It might be a little technical, but it should not be too difficult to see that the Iconclass concept in the screenshot is a composite concept built from two elements: the concept '*Alertnes, Vigilance...* (**52A23**)' and the concept '*emblematical representation of...* (**+4**)'. It should also be fairly easy to grasp that the second element can be used as an extension of thousands of abstract concepts in the section of Iconclass listing Abstract Ideas and Concepts (**Iconclass section 5**). These concept vary widely: from Vigilance to Eloquence, from Jealousy to Honesty, and from Justice to Stupidity. As you can see in the screenshot, to find such a concept the user of the Iconclass browser can do a simple keyword query, in this case for the words *emblem* and *alertness.*



The algorithm necessary to retrieve this concept, though not the most sophisticated implementation of the Iconclass rule base, is still more complex than you might expect. It would take us too far to explain this in detail. Suffice it to say that the two words in the query '*emblem*' and '*alertness*' are found in two separate records in the database and that these records are considered related only because the algorithm understands the Iconclass 'grammar' defining relationships in Iconclass.
The exact details of the algorithm, however, are not the point here. The point is that the end user – e.g. a researcher searching a webcatalogue for representations of animals symbolizing Vigilance other than the lion with one eye open – would want to query the image database with the exact same words '*emblem*'and '*alertness*' which the original cataloguer used to find the concept in Iconclass in the first place. Not really interested in the technicalities, the researcher wants to retrieve images of cranes, cocks or dogs that are as watchful as the lions we saw earlier on.

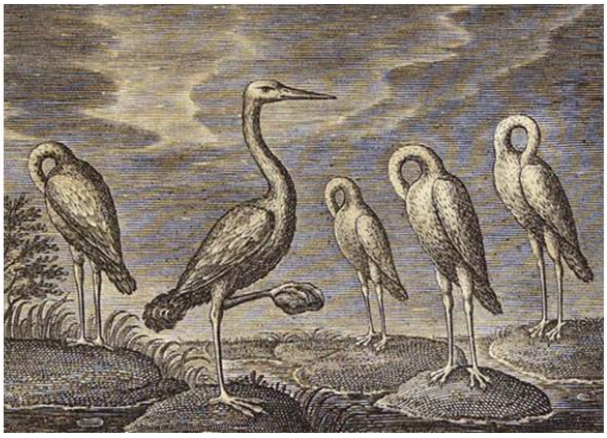
To facilitate this, a special variant of the Iconclass software was written, baptized the **Harvester of Iconclass Metadata** (or **HIM)** service. The HIM software in essence turns a datafile that contains two

core metadata elements – an object identifier and the Iconclass codes that were used as tags for the object – into a standalone iconographic retrieval application that can also be added as a plugin to any webcatalogue. This dedicated retrieval application ensures that the end user can use the same combination of words to retrieve the Iconclass concepts with which the powerful searches are actually done. The HIM service thus provides a generic solution to a crucial problem related to the use of authority files like controlled vocabularies and thesauri, especially when they rely on more than simple hierarchical subordination.
In the next section we shall discuss a few aspects of an image databank that is built around this HIM principle, albeit that Arkyves is somewhat of a special case, because it aggregates several web catalogues and acts as a single access point to a series of databases using the HIM service.

4.c. An aggregated corpus of Iconclass-tagged objects: Mnemosyne and Arkyves

4.c.i. Practical aspects: the basics of the HIM implementation

The basic principle of the HIM service is simple but it is radically different from what software developers in a library and museum context would normally expect. The difference is in the direction the data will flow. It is possible to enrich local catalogues by absorbing major parts - or even the whole - of the Iconclass data. However, merely adding the Iconclass data without also implementing the rule base, the 'grammar' of the system, will not do the job. By far the easier strategy is to reverse the procedure and enrich an Iconclass system with data exported from the local catalogues. What makes it easy to export information from a catalogue that uses Iconclass for its subject access, is the simple fact



that Iconclass is a classification system. Therefore, even the most complex concept in the system is expressed in a compact code, or '*notation*'. Like barcodes, QR-codes, or ISBN- numbers these notations are very concise containers of information.

The principle of the HIM service can be caught in a simple mantra: "*Don't move the system to your notations, but your notations to the system*".
What does that mean in actual practice? To explain this we shall use a hypothetical example. These pictures of cranes are in fact catalogued. The first is from a 13th century Bestiary in the British Library. The second and third one are from 17th century emblem books in the university libraries of Glasgow and Illinois. In all cases we see a crane holding a stone, a well known and ancient representation of vigilance and responsible leadership. The bird is holding a stone in one claw. Should he fall asleep, the stone will drop and the noise will wake him up.

The meaning can be summarized with the help of Iconclass concepts; to keep things simple we'll use only two:

x 25F37(CRANE)(+5245) shore-birds and wading-birds: crane (+ animal(s) holding something)
x 52A23(+4) Alertness, Vigilance; 'Guardia', 'Vigilanza', 'Vigilanza per difendersi & oppugnare altri' (Ripa) (+ emblematical representation of concept)

Now, imagine these three libraries cataloguing their collection with Iconclass. Their databases will then contain records with an identifier, a local recordnumber, unique to their image. All records will also contain the Iconclass notations expressing this concept of the crane as an emblem of Vigilance. This bare minimum of metadata would look something like this:

IDENTIFIER localrecordnumber_0001
ICONCLASS 25F37(CRANE)(+5245)
ICONCLASS 52A23(+4)

In this form the information is useless to any end user who is not an Iconclass expert – and even then... What the end user wants is to be able to initiate a query with words like '*crane'* and '*vigilance*', at least to get things going. The user may also want to cast the net wider and search for '*birds*' and '*vigilance*' or even '*animal*' and '*vigilance*'.
Moreover, French, Italian and German (and Chinese...) end users would appreciate it if they could do this query in their mother tongue. To offer this option, as we said above, these libraries would need to implement the Iconclass 'grammar' in combination with their copy of the raw Iconclass data. To grasp the implications – and again apologizing if this gets a little technical - have a look at this hierarchical chain of concepts, starting with **25F animals** and ending with **25F37(CRANE)(+5245) shore-birds and wading-birds: crane (+animal(s) holding something)**. Notice that at the bottom of this screenshot there is a set of words in italics, among which are ***animal*** and ***holding***. This is the cumulative set of hand-picked keywords with the help of which concepts can be retrieved from the Iconclass system, in addition to the words of the concept definitions.

25F animals
25F3 birds
25F37 shore-birds and wading-birds
25F37(...) shore-birds and wading-birds (with NAME)
25F37(CRANE) shore-birds and wading-birds: crane
25F37(CRANE)(+5) shore-birds and wading-birds: crane (+ animal(s) in motion; positions, expressions of animals)
25F37(CRANE)(+52) shore-birds and wading-birds: crane (+ movements of animal(s))
25F37(CRANE)(+524) shore-birds and wading-birds: crane (+ movement of animal(s) in relation to an other animal, human figure, or object)
25F37(CRANE)(+5245) shore-birds and wading-birds: crane (+ animal(s) holding something)
animal · bird · crane · earth · holding · motion · nature · posture · shore-bird wading-bird · world

Now scan the list of concepts from top to bottom, and you will understand that the keyword ***animal*** refers to the concept at the top of this branch of the hierarchy and therefore – by implication – to the complete chain: a crane holding something still is an animal... On the other hand, ***holding*** only refers to the final and most specific concept in the chain.
You do not have to be a programmer to understand that an algorithm has to be written to ensure that a word search using '***animal***' and '***holding***' will retrieve records that have been tagged with **25F37(CRANE)(+5245)**. Even though this is accomplished by one of the more trivial algorithms implementing the Iconclass rule base, each of the libraries cataloguing their pictures of vigilant cranes will have to create its own local version, using its own local copy of the Iconclass data. And what is more: they will also have to do this for all the more complex transactions, and if they decide to offer their

2 Nature
25 la terre, le monde en tant que corps céleste
25F animaux
25F3 oiseaux
25F37 échassiers et oiseaux riverains
25F37(...) échassiers et oiseaux riverains (avec NOM)
25F37(CRANE) échassiers et oiseaux riverains : grue
25F37(CRANE)(+5) échassiers et oiseaux riverains : grue (+ animal(animaux) en mouvement; positions, expressions de l'animal)
25F37(CRANE)(+52) échassiers et oiseaux riverains : grue (+ mouvements de l'animal)
25F37(CRANE)(+524) échassiers et oiseaux riverains : grue (+ mouvement de l'animal par rapport à un autre animal, une personne ou un objet)
25F37(CRANE)(+5245) échassiers et oiseaux riverains : grue (+ animal(animaux) tenant quelque chose)
animal · grue · monde · mouvement · nature · oiseau · oiseau riverain · position · tenir · terre · échassier

information in multiple languages, they will need to import the translated versions of Iconclass as well – such as the French translation shown here.

Needless to say that if it can be done once, it can be done again. Everyone is entitled to have a go at the invention of a better version of the wheel. However, Iconclass's natural habitat is the field of the Arts and Humanities, where funding is always problematic, even for larger museums and libraries. Preventing the waste of resources therefore was an important reason the HIM service was created. The essential idea was that institutions using Iconclass deposit a core set of metadata, similar to the set shown above, to a central database that also contains a complete Iconclass system. For each deposited collection an instance of the Iconclass HIM browser would be created, that automatically takes care of all Iconclass retrieval functions. Such a browser could then be included as a plugin to any web catalogue without additional programming, i.e. without extra costs. By simply adding one more piece of information – the name of the collection - this HIM browser would always be tailored to a specific collection and a specific institution. So, the core metadata set for our vigilant crane in the Glasgow University Library (**GUL**) collection would look like this:

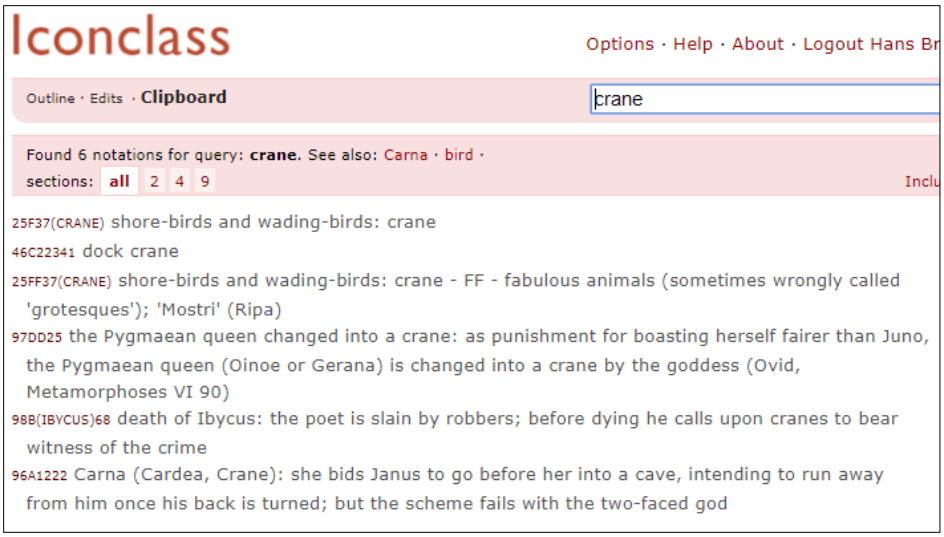
IDENTIFIER localrecordnumber_0001
HIM GUL
ICONCLASS 25F37(CRANE)(+5245)
ICONCLASS 52A23(+4)

Obviously, the HIM data element would change with each deposited collection, and for every HIM value a separate instance of the Iconclass browser would come into existence in an automated process. Simultaneously gathering all of these datasets in one common data pool would set up a HIM browser that acts as a single access point to the iconography of all contributing collections. The aggregator site that does this job is called Arkyves.

4.c.ii. Starting the serendipity engine: making the most of co-occurring concepts

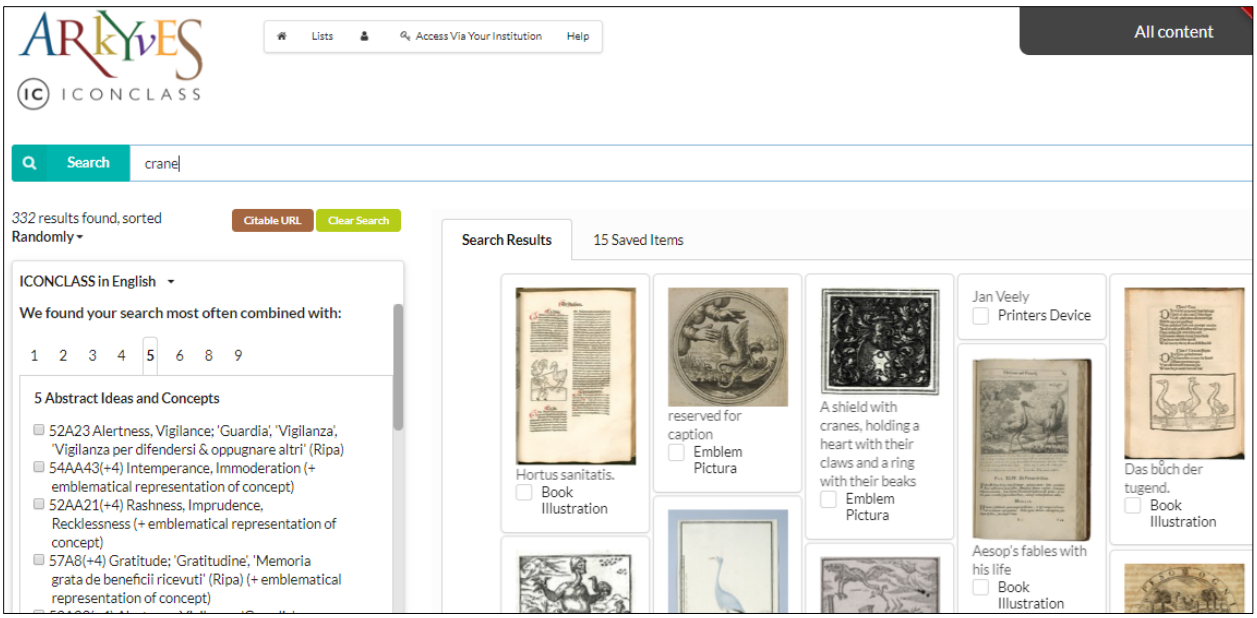
But there is more to Arkyves. Although Arkyves does bring collections together, it is not merely an aggregator site. Precisely because it is a *collection of collections* that use Iconclass for subject retrieval, its sum is greater than its parts. The accumulation of Iconclass concepts facilitates an additional level of knowledge discovery. To illustrate this we shall return once more to our vigilant crane. It speaks for itself that *vigilant crane* is a composite concept. It combines the representation of a bird with an abstract idea. Obviously, the crane cannot be equated with vigilance: not every representation of a crane expresses vigilance; nor is vigilance always expressed by a crane. However, it is easy to imagine that a database contains multiple instances of this theme and thus of records that combine these Iconclass

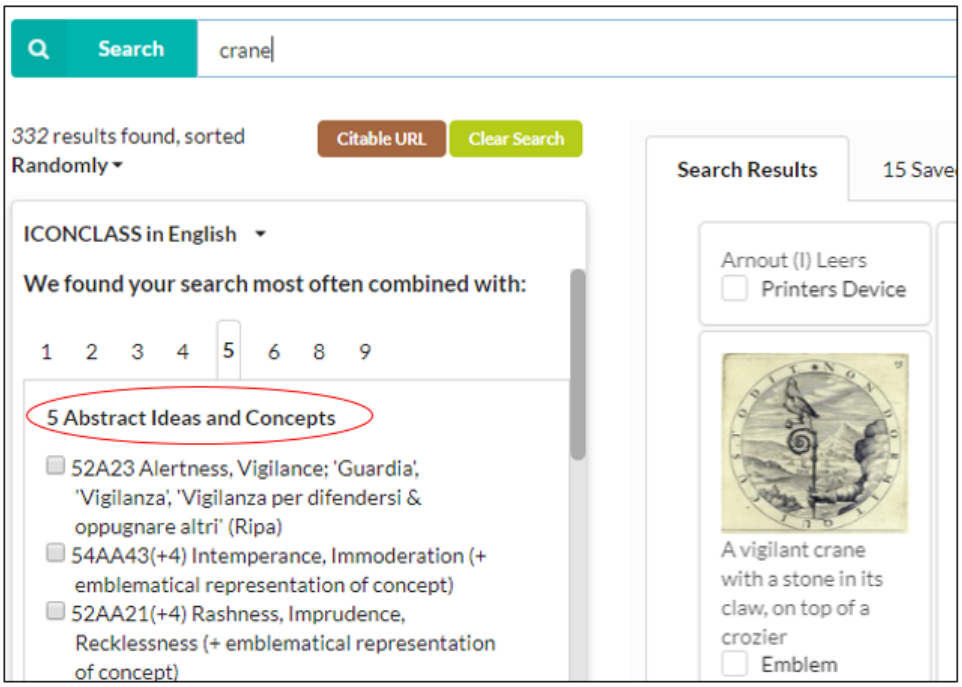
concepts. In fact that is exactly the case for Arkyves: it contains a few dozen instances of these two concepts – **crane** and **vigilance –** *co-occurring*.



Now, by exploiting this *co-occurrence* of concepts - i.e. of different iconographic concepts tagging the same image - a search for one of those concepts can be used to point us in other directions as well. In other words: the information retrieved through a search for 'crane' alerts us to the fact that this bird can connotate 'vigilance'.

The obvious question: can we find out if it was used in other contexts as well?





To this end an algorithm was developed that uses, among other things, the *frequency of co-occurrences* of Iconclass concepts to create an organized web of cross references. It does so in a two-step process. In the Iconclass system, as illustrated in the screenshot from the Browser, the keyword 'crane' points us to a handful of concepts. In a first step Arkyves is searched with the word 'crane', and records are retrieved if they contain any of these Iconclass concepts. In a second step *all of the other Iconclass concepts* that are found in these records are retrieved and then displayed in order of frequency. These concepts are offered as an organized list, organized that is according to the 9 broad Iconclass categories. For example: Iconclass category '**5**' contains **Abstract Ideas and Concepts.** The concepts in this category most frequently found in combination with (Iconclass concepts linked to the word) **crane** will be found under the little tab '**5**'- as shown in this screenshot. So for a simple word search for '**crane**', Arkyves shows - in two steps - that in the aggregated collections cranes are combined not just with the concept **52A23 Alertness and Vigilance**, but also with, for example, **54AA43 Intemperance, Inmoderation** and **52AA21 Rashness, Imprudence.** Obviously it is not just the Iconclass concepts as such that are offered as hints for further research. They are supported by the actual source documents – images and texts - where we encounter cranes in a variety of contexts, as can be glimpsed in this Arkyves screenshot.

The association of cranes and **Intemperance**, for example, is demonstrated by Alciatus's emblem '**Gula**', where the man symbolizing an insatiable appetite has the neck of a crane. The man aiming for a flying

crane, not noticing the snake at his feet, illustrates the **Imprudence** of those who are dangerously curious about 'higher things' (**Qui alta contemplantur, cadere**).

Now, using the term 'serendipity' suggests that these links between crane and vigilance, intemperance or imprudence are unintentional, 'lucky' discoveries. As our somewhat technical explanation has hopefully demonstrated, that is not the case. These references may be unexpected, they may point users in directions that they had not been looking, but they are not random as they depend on the latent features of a well-organized system.

5. When Henri met Aby: Case studies

Introduction

At the start of our discussion about the role of images as carriers of historical information, I stated that image analysis by default moves in two directions. On the one hand we zoom in on details, registering the properties of an individual image. On the other hand we zoom out, as we attempt to discover more general patterns shared by groups of images. We move smoothly across a continuum of abstraction levels as we recognize details, identify motifs and narratives and label themes and *topoi*. Inevitably our observations are guided by our hypotheses and hunches, by what we know and have seen before. When we describe what we see, we make choices. We have to. We choose to register certain details of the image in front of us; other details we do not register. We may not have the time to describe them, we may consider them insignificant, or we may simply overlook them. In addition we may misinterpret aspects of what we see until a new piece of visual of textual evidence puts us on the right track. Whatever we do, however much effort we invest, we know from the outset that our descriptions can never be "complete" and they will never be "definitive". It has always been that way. What is seen as relevant in an image always shifts with the researcher's perspective. Diverging interpretations of the sources are exactly what the scholarly debate in cultural history is about.

In our context, however, the past decades have seen changes that force us to reconsider the way we look at cataloguing, at iconographic description and at subject metadata. Massive digitization has put (digital representations of) historical sources at our disposal on a completely uprecedented scale. Whether we are studying medieval book illumination, early Netherlandish panel or baroque wall paintings, European, Chinese or Japanese prints, we are confronted by an explosive growth of the number of visual resources at our fingertips.

We cannot expect our libraries, museums and archives, overwhelmed and understaffed, to accompany this tidal wave of images with iconographic information at the level of detail we have been talking about here. While this may sound fatalistic, the contrary is true. Unlike the boy in this emblem, deliberately looking

through the wrong end of the telescope[23] at some ravens in a dead tree, because he "*does not want to know them*", we should not feel intimidated by what we see. Digitization did not increase the gap between the detailed information researchers of iconography need and the amount of raw data 'out there'. The sources we are talking about existed independent of digitization. They survived on the shelves of our libraries and in the halls of our museums; digitization simply is making them visible without us having to travel to see them[24]. This second image supports our point. It shows the personification of the human soul staring at the symbols of Death and the Last Judgement – again *through the wrong end of a telescope*. The message of both images is similar: we are fooling ourselves if we think that we can keep Death at a distance by turning around the telescope through which we are looking at it.

The point is that establishing a connection between these images depends on a human cataloguing effort. It does now, in this digital era, as it did in analogue times. A human observer has to notice the analogy and catalogue it.

The revolutionary difference – the paradigm shift, if you will - is that in a digital environment one researcher can pick up where another researcher left off and store new information on the same platform. Any researcher, in other words, can contribute new subject metadata to a common pool of iconographic information. Technology facilitates this new form of collaboration, and, if anything, this seems to me what 'digital humanities' is about. It will not come as a surprise that we also argue that the accumulation and exchange of iconographic information will benefit from the use of a common language, *in casu* that of the Iconclass system.

The best way to illustrate this is with the help of a few case studies. These studies usually started as conference papers and they combine the methodological points we have made above with observations about sources of – mainly – early modern iconography.

<to be continued ...>

Hans Brandhorst
Voorschoten, October 2019

[23] As the German explanation says "*Ein Genius siehet durch ein ausgezogenes aber umgekertes Perspectiv, nach einem mit Raben besetzten kahlen Baum*". A dead tree with some ravens -

[24] Obviously simplifying things here: autopsy of the physical object may still be required.