Extracting concept descriptions from the Web: the importance of attributes and values

Massimo POESIO a and Abdulrahman ALMUHAREB b

a Center for Mind / Brain Sciences, Università di Trento, Italy and Language and Computation Group, University of Essex, UK
b King Abdulaziz City for Science and Technology (KACST), Saudi Arabia

Abstract. When extracting information about concepts from the Web, the problem is not recall, but precision: trying to identify which properties of a concept are genuinely distinctive. We discuss a series of experiments in empirical ontology using both unsupervised and supervised methods, showing that not all semantic relations we can extract from text are equally useful, and suggesting that attempting to identify concept attributes (parts, qualities, and the like) and their values results in better concept descriptions than those obtained by being less selective.

Keywords. Ontology learning from text, relation extraction, attributes and values

1. Introduction

The availability of huge amounts of textual data about concepts on the Web and in other web-size corpora has kickstarted a new area of research usually called ontology learning from text, which has the ambitious goal of developing methods for extracting from text full ontologies—or at least, taxonomies of conceptual knowledge. As this volume testifies, ontology learning from text is a very active area of research, and very many systems performing this task have been developed, including OntoLearn [1], Text2Onto [2] and other systems discussed in the volumes in this collection.

Our research is at the boundaries between ontology learning proper and work on (nominal) lexical acquisition in Computational Linguistics CL [3,4,5,6,7,8], in that the approach we are pursuing is perhaps best described as ‘lexical acquisition meets relation extraction’. In some sense, our goals are more modest than those of systems like OntoLearn or Text2Onto: we are only interested in clustering what in psychology are called (nominal) basic categories such as dog, cat, car, and truck into what psychologists call their superordinate categories, such as animal or vehicle. From another point of view, however, our goal is quite ambitious: to carry out a form of what we call empirical ontology—namely, to use computational methods to gain insights into concepts that can supplement the use of evidence from psychology and the neural sciences in providing a test for theories of ontology.
Perhaps the most distinctive feature of the methods discussed in this paper is that we defined the relations to be extracted on the basis of views developed in Artificial Intelligence (AI), linguistics, and philosophy about the 'intrinsic' properties of concepts— that we will call attributes. In other words, we believe that not all information is equally important to build good concept descriptions: from the clustering perspective at least—the perspective of identifying which concepts are similar, and which ones have distinct superordinate categories—'less is more': not all information that can be gathered from a corpus is equally important, not even all semantic information.

The structure of this paper is as follows. We begin with a quick overview of work on concepts, on lexical acquisition, and on relation extraction. We then discuss first our unsupervised, then our supervised methods for building concept descriptions. A discussion follows.

2. Background

Concepts are viewed as complex mental objects characterized by a number of attributes or features in most theories of concepts developed in philosophy, psychology, linguistics and Artificial Intelligence (AI), even in those theories derived from the work of Wittgenstein and Rosch that do not subscribe to the view that concepts can be 'defined' in the Aristotelian sense.

The notion of 'attribute' assumed in philosophical work is typically semantic. For instance, according to Aristotle (in *Metaphysics*) the nature of an object can be described by four 'causes': the *material cause* (the material of which the object is composed), the *agentive cause* (what causes the object’s movement or creation), the *formal cause* (what a thing is planned and intended to be—its essence and form), and the *final cause* (“that for the sake of which a thing exists, or is done”). This view of the nature of objects was adopted in linguistics by Pustejovsky [9], who developed Generative Lexicon theory according to which an integral part of a lexical entry is its *qualia structure* in the sense of Aristotle. Pustejovsky’s qualia structure consists of four (types of) roles corresponding to Aristotle’s four causes. The *formal role* is a complex of attributes specifying what type of object the concept denotes—its ‘intrinsic qualities’. These include both its supertypes (its ⟨isa⟩ relations and attributes specifying its form. For instance, in the case of the concept book, the formal roles include the fact that a book is a physical object with certain qualities such as ⟨shape⟩ and ⟨color⟩. The second role is the *constitutive role*, specifying the stuff and parts that an object consists of. Again, in the case of book, the constitutive roles include the fact that a book is made of paper, that it has chapters and an index, etc.. The *telic role* specifies the purpose of the object—e.g., in the case of a book, reading. Finally, the *agentive role* specifies how the object was created: e.g., in the case of a book, by writing.

In Artificial Intelligence, theories of concepts based on a semantic notion of attribute have been developed in the area of formal ontology. For instance, Guarino [10] developed a theory of attributes according to which there are two types of attributes: *relational* and *non-relational*. Relational attributes include *qualities* such as ⟨color⟩ and ⟨position⟩ and *relational roles* such as ⟨son⟩ and ⟨spouse⟩. Non-relational attributes include *parts* such as ⟨wheel⟩ and ⟨engine⟩. Activities such as reading and writing for a book are not viewed as attributes in Guarino’s classification. *Description logics* [11], the latest form
of the approach to semantic networks developed by Brachman, Levesque, and colleagues in papers such as [12], are logics developed to define concepts in terms of subsumption relations (isa links) and attributes.

Work on (nominal) concept acquisition from corpora in computational linguistics, on the other hand, is usually based on the distributional view of meaning derived from Wittgenstein via Firth [13], according to which the meaning of a word is specified by the other words with which this word co-occurs, not necessarily through the mediation of deeper semantic notions. This view of meaning led to the vector space model of semantic representation, proposed by Salton et al. [14] to represent the meaning of documents in Information Retrieval, but then adopted as a view of lexical meaning in the pioneering work of Schuetze [6], and then in work such as [4,5]. These models have proven rather successful at modelling synonymy, as shown e.g., by the results obtained on the TOEFL test [15]. Starting at least with Grefenstette [3], however, a modified view of vectorial spaces have been developing, in which the syntactic relation between a word and its neighbors (in the sense of dependency grammar) is taken into account in defining proximity [7,8].

In parallel with these efforts, however, there has been in CL a line of work dedicated to work on (supervised and unsupervised) extraction of semantic relations, beginning with the seminal work of Hearst [16] who developed unsupervised methods for the extraction of hyponymy relations using patterns. Hearst’s proposals on hyponymy were followed up in [17,18], and additional work was carried out on part-of relations [19,20,21] and on other relations as part of the ACE program.2

More recently still, these methods for relation extraction have begun to be applied to ontology construction and population [22,23,24]; see also the papers in this volume, particularly the chapters by Aussenac et al., Pantel and Pennacchiotti, and Voelker et al. Our own approach [25,26,27,28,29] belongs to this line of work attempting to combine the two types of research: using relation extraction techniques to extract the ‘features’ to be used to describe vectors, but for ontology learning rather than ontology population. (We discuss related work in a later section.)

3. Attribute-based concept descriptions: an unsupervised approach

3.1. Attributes and values in concept descriptions

When less data were available—e.g., when extracting relational information from the Penn Treebank as done by Caraballo [17], the 64 million-word corpus used by Lin [7], or even from the 100 million-word British National Corpus used by [20]—one cannot afford to be choosy: every bit of information was necessary. But now, working with the Web or Web-sized corpora,3 too much information is the problem.

Our approach to concept extraction is therefore based on the principle that sometimes ‘less is more’: extract only ‘intrinsic’ properties of concepts, as opposed to all information about them that can be found in a corpus. (This, of course, is predicated on the

---

2http://www.nist.gov/speech/tests/ace/index.htm
3We are not aware of any study attempting to establish what exactly is the threshold at which the phenomenon we are pointing out—elimination or reduction of the data sparsity problem, excessive data richness—begins to be displayed; we would guess not before 10 giga words.
assumption that there are more ‘essential’ properties of concepts—a point to which we will return in the discussion.) Specifically, what distinguishes our own approach to using relation extraction to acquire concept descriptions is the attempt to extract from corpora ‘attributes’ of concepts and their ‘values’, instead of all relations, being guided in this by works such as [10] and [9].

The methods for extracting attributes proposed in [25, 26] are based on the observation that many such properties are dependent concepts realized using relational or functional nouns, and can therefore be used in constructions of the form “the X of the Y”. Parts such as ⟨wheels⟩, for instance, are generally dependent concepts, hence the fact that cars have ⟨wheels⟩ can be expressed using the construction “the WHEEL of the CAR”. This point can already be found in [30], where we can find the following test for attribute-hood. According to Woods, A is an attribute of concept C if we can say:

VALUE is a/the ATTRIBUTE of CONCEPT,

For example, the fact that we can say:

brown is a color of dogs

suggests that ⟨color⟩ may be an attribute of dog. Relational nouns potentially expressing attributes also occur in possessive constructions, as in “the CAR’s WHEEL”. On the basis of this observation, we extracted possible concept attributes searching in the Web for the two constructions above.

Many concepts are characterized by special values of certain attributes: e.g., while all fruit have a ⟨color⟩ attribute, bananas are typically yellow, whereas strawberries are typically red. Thus we aimed to include in our concept descriptions, in addition to information about possible attributes of concepts, information about values of such attributes which are particularly distinctive of that concept.

It turns out that the construction proposed by Woods above is not very common, even when the Web is used as a corpus. We found less than 500 instances of the constructions “VAL is the ATTR of CONCEPT” or “the ATTR of the CONCEPT is VAL” [29]. The construction suggested in [32],

“the VAL1 or VAL2 ATTR”,

(as in “the RED or WHITE COLOR”) has very high precision (almost 80%) but does not have very high recall either: we could not find in the Web any values for 3 out of 10 very common attributes [29]. So in the end we settled for considering as potential ‘values’ all prenominal modifiers occurring in constructions of the form:

“the VALUE CONCEPT is”

as in “the RED CAR is”. This of course meant that we included among the ‘values’ a number of modifiers that could not really be considered values of any particular attribute (e.g., “trained” as a modifier of “horse”) as well as a lot of information that is best described as collocational (e.g., “Trojan, Arabian, hobby” for “horse” again); we return to this point in the discussion.

4 We looked for other constructions using the methods proposed by Hearst [31], but such constructions resulted in much lower precision for little recall gain.
3.2. Experimental results: text patterns

We tested the hypothesis that we would obtain better concept descriptions by concentrating on the constructions expressing information about attributes and values in a series of experiments discussed in [25,26,29,33]. We tested two ways of identifying these constructions: using simple textual patterns, as done in [31,19], and using a (dependency) parser—specifically the RASP parser developed by Briscoe and Carroll [34].

In all of the experiments described in this section we used the t-test as defined by [8] as a ranking function, defined as follows, where $t_{i,j}$ is the output weighted frequency for the $concept_i$ and the $feature_j$, $N$ is the overall frequency, and $C$ is a count function:

$$t_{i,j} \approx \frac{C(concept_i,feature_j)}{N} - \frac{C(concept_i) \times C(feature_j)}{N^2} \sqrt{\frac{C(concept_i,feature_j)}{N^2}}$$

Our own tests confirmed Curran and Moens’ finding that this measure outperformed other measures including simple frequency, mutual information, log likelihood ratio, and $\chi^2$. As a similarity function we used the version of the Jaccard coefficient defined by [8] and shown below, where $t_{m,i}$ and $t_{n,i}$ are the weighted co-occurrence frequencies between $concept_m$ and $concept_n$ with $feature_i$, respectively.

$$Jaccard(concept_m, concept_n) = \frac{\sum_i(t_{m,i} \times t_{n,i})}{\sum_i(t_{m,i} + t_{n,i})}$$

Again, our results confirmed those by Grefenstette and by Curran and Moens that extended Jaccard outperforms other similarity measures including the cosine and Lin’s similarity function [7]. We also always used as our clustering algorithm Repeated Bisections [35], a variant of K-means clustering, as implemented in the CLUTO clustering tool [36], as again we found it outperformed other clustering algorithms that we tested including EM, agglomerative clustering, and COBWEB.

In a preliminary experiment [25], we only used text patterns to extract information about the attributes and values of 214 relatively common nouns associated with synset belonging to 13 different WordNet classes. For finding attributes we used the following two Google patterns:

"the * of the C R"

"the C’s * R"

where $C$ is a concept, $R$ is a restrictor such as is and was, and the wildcard denotes an unspecified attribute. For values we used the following Google pattern:

"[a|an|the] * C R"

In this experiment we found that attributes were as informative as values but that to achieve the best performance it was necessary to include in the concept descriptions a combination of the highest ranked attributes and values.

In a second experiment [26] we compared attribute / value extraction using text patterns and using a parser. For this experiment we used a dataset of 402 concepts covering
all 21 WordNet unique beginners, and balanced for frequency (1/3 of the concepts in the set are high frequency as measured from the BNC, 1/3 are medium-frequency, 1/3 are low frequency) and ambiguity (1/3 of the concepts are highly ambiguous in the sense of having more than 4 senses in WordNet, 1/3 have between 2 and 3 senses, 1/3 have only 1 sense). (The dataset is shown in Appendix A.) The results were measured using purity and entropy, both of which measure the extent to which clusters are ‘uniform’, and are defined as follows.

Let $S_r$ be a cluster, $n_r$ the size of cluster $S_r$, $q$ the number of classes in the dataset, $n_i^r$ the number of concepts from the $i$th class that were assigned to the $r$th cluster, $n$ the total number of concepts, and $k$ is the number of clusters. Then the purity of cluster $S_r$, $P(S_r)$, is the ratio of the number $\max_i(n_i^r)$ of elements in the ‘dominant’ category for $S_r$—the category with the greatest number of elements in that cluster—to the number $n_r$ of elements in that cluster. A cluster containing only elements from one class will have purity 1. The entropy of cluster $S_r$, $E(S_r)$, is the standard entropy—a more comprehensive measure, that takes into account the entire distribution of categories in the cluster. Overall entropy and purity are the weighted sum of individual cluster entropies and purities respectively.

$$\text{Entropy} = \sum_{r=1}^{k} \frac{n_r}{n} E(S_r), \quad \text{where} \quad E(S_r) = -\frac{1}{\log q} \sum_{i=1}^{q} \frac{n_i^r}{n_r} \log \frac{n_i^r}{n_r}$$

$$\text{Purity} = \sum_{r=1}^{k} \frac{n_r}{n} P(S_r), \quad \text{where} \quad P(S_r) = \frac{1}{n_r} \max_i(n_i^r)$$

The results using textual patterns (the same as in the first experiment) are shown in Table 1.

<table>
<thead>
<tr>
<th>Measures</th>
<th>Attributes and Values</th>
<th>Attributes and Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>21 Classes (402 Concepts)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purity</td>
<td>0.657</td>
<td>0.567</td>
</tr>
<tr>
<td>Entropy</td>
<td>0.335</td>
<td>0.384</td>
</tr>
<tr>
<td>Vector Size</td>
<td>24,178</td>
<td>94,989</td>
</tr>
<tr>
<td></td>
<td></td>
<td>119,167</td>
</tr>
</tbody>
</table>

Table 1. Clustering results using textual patterns

Here, we find that using vectors of attributes we can get better results than using vectors of ‘values’, even with vectors of a quarter of the size. And again, the best results are obtained by combining attributes and ‘values’.

Looking at per-category purity, we find that we obtain perfect purity for edible fruit, vehicle, illness. Purity is above .8 for animal, chemical element, creator, feeling and monetary unit—almost all concrete categories. The worse purity is for abstract categories; among these, the worse two are motivation and time, with purity less than .4.

3.3. Experimental results: using a dependency parser

Text patterns are rigid: a pattern like “the * of the C is” cannot match, for instance, the construction “the SPEED of John’s CAR is”. Using a parser and finding instances of the
construction by searching in its output would allow us to find more constructions. On the other hand, this might introduce more errors, as well as making the method more language-dependent as we don’t have dependency parsers for all languages.

For the second part of the experiment, we collected the documents found using the text patterns in the first part of the experiment, extracted the sentences that contained instances of the desired concept, and parsed them with RASP, obtaining results such as those shown in Figure 1.

| (detmod, strawberry, the)          | (dobj, like, strawberry)          |
| (detmod, strawberry, a)           | (ncmod, fruit, strawberry)        |
| (ncsubj, be, strawberry)          | (ncmod, strawberry, wood)         |
| (ncmod, strawberry, fresh)        | (ncsubj, grow, strawberry)        |
| (ncmod, strawberry, wild)         | (dobj, eat, strawberry)           |
| (ncmod, plant, strawberry)        | (ncsubj, have, strawberry)        |
| (xcomp, be, strawberry)           | (ncmod, strawberry, frozen)       |
| (ncmod, strawberry, whole)        | (ncmod, strawberry, cup)          |
| (ncmod, strawberry, ripe)         | (ncmod, strawberry, red)          |
| (ncmod, strawberry, cultivated)   | (dobj, have, strawberry)          |
| (iobj, with, garnish, strawberry) | (ncmod, of, variety, strawberry)  |
| (detmod, strawberry, an)          | (ncmod, strawberry, modern)       |
| (ncmod, variety, strawberry)      | (ncmod, cultivar, strawberry)     |
| (dobj, slice, strawberry)         | (ncsubj, grow, strawberry, obj)   |
| (ncmod, strawberry, large)        | (ncmod, strawberry, big)          |

Figure 1. The most frequent grammatical relations for strawberry

RASP gives us the opportunity to study the usefulness for concept description of many other types of constructions in addition to those tested in the experiments with text patterns. The text patterns essentially extracted from text two (approximated) instances of what in RASP is called the ncmod grammatical relations: attribute patterns extracted cases which RASP would parse as

(ncmod, of, ATTR, CONCEPT),

(as in (ncmod, of, color, strawberry)), whereas value patterns extract cases that RASP would parse as

(ncmod, CONCEPT, VAL)

as in (ncmod, strawberry, red). However, many other types of grammatical relations have been included in concept descriptions in the literature on using syntax-based relations for concept descriptions [3,7,8], including in particular direct objects and subjects. We could now compare the effect of using these additional grammatical relations. The results are shown in Table 2.

As the table shows, using all grammatical relations as part of the concept descriptions results in less purity than using only attributes and values—around a third of the features collected using RASP— or indeed just using attributes—less than a tenth of all features. In other words, these results indicate that “less is more”: adding more information not necessarily results in better concept descriptions. The table also shows that using a parser results in slightly better performance than using text patterns, at the expense of less language-independence.
Table 2. Clustering results using different subsets of grammatical relations

<table>
<thead>
<tr>
<th>Measure</th>
<th>Grammatical relations subset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attribute</td>
<td>Value</td>
</tr>
<tr>
<td>Purity</td>
<td>0.656</td>
</tr>
<tr>
<td>Entropy</td>
<td>0.320</td>
</tr>
<tr>
<td>Vector Size</td>
<td>20,285</td>
</tr>
</tbody>
</table>

4. Supervised attribute extraction

4.1. A first classification scheme for attributes

The patterns used for attributes in the experiments above match constructions expressing all sorts of semantic relations other than attribute-concept. The range of semantic relations expressed by the “the X of the Y” construction is illustrated in Figure 2 with examples of ‘attributes’ of deer collected by our unsupervised system discussed in Section 3. As the figure shows, this construction is used to express partitions of sets (hence the name partitive construction), ordinals, ‘picture of’ relations, naming, and other more complex relations. While some of this information may be considered highly distinctive of a concept like deer (e.g., it is hard to imagine that one would find discussions of the meaning of bismuth), none of it can be considered as expressing an ‘attribute’ of deer, in the intuitive sense of a ‘defining property’ of the concept.

the rest / majority of the deer
the first / last of the deer
the picture / image / photos of the deer
the cave / mountain / lake of the deer
the meaning of the deer [in Western philosophy / ... ]

Figure 2. Semantic relations expressed using the “the X of the Y” construction.

In order to make further progress towards the goal of extracting concept descriptions containing only ‘proper’ attributes, and excluding information such as those in Figure 2, it is necessary to be more clear about what we mean with ‘attributes’—i.e., to have a theory of attributes, or at least a classification scheme specifying which relations count as proper attributes of concepts and which ones instead do not. Having this theory, or the classification, would allow us to develop supervised or unsupervised theories of concept extraction.

Unfortunately although the notion of ‘attribute’ has been part of philosophical theories of concepts and knowledge in philosophy since at least Aristotle, no fully worked out theory of attributes exists, nor a classification of attributes covering all concepts, in part also because philosophy and psychology tend to concentrate on a few categories of concepts. Partial theories can however be found in works such as [10] and [9] discussed earlier; these works can serve as a starting point for our study. (Inversely, one would hope of course that work on ontology extraction from text might contribute to formal work on ontology.)
Out of the two proposals of Guarino and Pustejovsky we developed a classification scheme for attributes that considers a relation in which a concept participates as an attribute if it belongs to one of the following types:

- **qualities.** These relations include Guarino’s qualities and (some of) Pustejovsky’s formal roles: e.g., the ⟨weight⟩/⟨fusibility⟩/⟨solubility in Aqua Regia⟩ of gold.
- **parts.** These relations include Guarino’s non-relational attributes, and Pustejovsky’s constitutive roles, including that is parts such as the ⟨wheel⟩ of the car or the ⟨leg⟩ of the animal, and ‘stuff’ such as the ⟨gold⟩ of the ring.
- **related objects.** These relations relate objects to other independent objects with which they are in strong association, such as the ⟨nest⟩ of the bird, and include Guarino’s non-relational attributes other than parts and relational roles.
- **activities.** These relations include Pustejovsky’s telic and agentive roles such as the ⟨reading⟩ and ⟨writing⟩ of the book, but also other important activities such as the ⟨publication⟩ of the book.
- **related agents.** Finally, these relations relate objects to agents that perform the activities above, such as the ⟨writer⟩, ⟨reader⟩ or ⟨publisher⟩ of a book.

Notice that ⟨isa⟩ relations were not included among attributes, although most work on ontology learning concentrates on this type of information [31,17]. This is in part precisely because there is no lack of insightful work on this topic, in part because this information plays a different epistemological function in the definition of concepts, in the sense of [37].

### 4.2. New experiments

In [28] we discussed the results obtained using a supervised classifier to classify potential attributes extracted from the Web in the five classes above, as well as the class ‘not-an-attribute’.

We collected from the Web 20,000 candidate attributes for the 402 concepts in the dataset, kept the 4,728 that occurred more than 20 times, and collected for all of them four types of information:

- **morphological features,** extracted through heuristics, such as the information that a particular noun might be derived from an adjective or a verb, which is useful to identify qualities and activities respectively [38];
- **question patterns,** that is, the frequencies obtained by querying the Web with questions of the form “What is the ATTR of CONCEPT” or “When is the ATTR of CONCEPT”;
- **features of features,** i.e., the top attributes of these potential attributes, extracted using the same patterns that we used to extract attributes; and finally
- **feature use,** that is, information about the respective frequency of the use of these nouns as attributes (“the ATTR of the *”) or concepts (“the * of the CONCEPT”).

We then hand-classified 1,155 of these feature vectors into six classes (the five classes above, and ‘not an attribute’), and we trained two classifiers: a binary one just making the decision attribute / not attribute, and one classifying attributes into one of five classes. We evaluated these classifiers in three ways: (i) through cross-validation over the 1,155
hand-annotated features, (ii) by running them over around 400 additional feature vectors and hand-evaluating the results, and (iii) by using them to filter the potential attributes and clustering the original concepts using only the remaining attributes.

The binary classifier achieved an accuracy of 81.82% as evaluated through cross-validation, which corresponds to an F value of .892 at recognizing attributes and .417 at recognizing non-attributes. About the same results were obtained over the additional 400 attributes.

The 5-way classifier achieved an accuracy of around 80% at cross-validation, corresponding to an F value over .8 for quality, activity, and part / related object, of .95 for related agent, and .538 for not-an-attribute. Accuracy over the additional 400 attributes however was significantly lower at around 70%.

Table 3 compares the results obtained when clustering concepts using all attributes—i.e., the results with all attributes shown in Table 1—(second column) with the results obtained by filtering attributes using simple heuristics (third column) and, finally, the results obtained when clustering after having removed the attributes classified as ‘not-attributes’ by the binary classifier.

<table>
<thead>
<tr>
<th></th>
<th>All Candidate Attributes</th>
<th>Heuristic filtering</th>
<th>Filtering by classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purity</td>
<td>0.657</td>
<td>0.672</td>
<td>0.693</td>
</tr>
<tr>
<td>Entropy</td>
<td>0.335</td>
<td>0.319</td>
<td>0.302</td>
</tr>
<tr>
<td>Vector size</td>
<td>24,178</td>
<td>4,296</td>
<td>3,824</td>
</tr>
</tbody>
</table>

Table 3. Clustering with a supervised classifier

As the table shows, removing ‘non-attributes’ resulted in a significant improvement in the purity of the clustering. (The difference between the value of purity using all candidate attributes and using filtering by classification is significant.)
5. Related work

As mentioned above, relation extraction techniques have been used in work on ontology population such as [22,23,39]. The techniques used in these systems are primarily unsupervised, and the focus is on developing methods for acquiring new patterns for extracting the relevant information, as in KNOWITALL [23]. [39] propose an unsupervised method, but introduce a measure of the reliability of pattern used to identify which of the patterns found by the system are reliable.

The work most closely related to ours is probably Cimiano and Wenderoth’s work on extracting qualia structures [40]. Cimiano and Wenderoth are also concerned with ontology learning rather than ontology population, and developed a completely unsupervised approach to extracting qualia structures by developing specific patterns for each of Pustejovsky’s qualia. Apart from their approach being unsupervised, the main difference from the present work is that Cimiano and Wenderoth do not use the extracted information to build vectors for clustering purposes, hence the main evaluation is not in terms of clustering performance, but by comparing the information thus extracted with what is found in the literature, and by human evaluation.

6. Discussion

In a sense, the most exciting aspect of this type of research is that we can revisit all sorts of old philosophical chestnuts, but with the guidance of empirical evidence.

One example of ‘philosophical’ question is the question of what is an ‘attribute’. We most certainly do not think that we found the definitive characterization of the notion of attribute. What we do think we have is evidence that attempting to identify the types of semantic relations that go under the name of ‘attributes’ or ‘qualia’ does seem useful from an ontology learning perspective (i.e., to draw a distinction between ‘basic categories’). One might even wonder whether this work provides empirical evidence for Aristotle’s notion of ‘essence’–in the sense that concepts may have ‘essential properties’ that are best in distinguishing them from other concepts. (Of course, for some applications one may want to collect all available data about concepts.)

Another important question is how far to go in the direction of obtaining purely ‘semantic’ concept descriptions. The concept descriptions obtained with the methods discussed in this paper mix semantic and distributional information; would it make sense to try to ‘clean up’ these descriptions further? Again, perhaps a distinction needs to be drawn between information to be used for clustering concepts (which may be in part distributional) and information about the concepts to be used for question-answering purposes.

Up until now the terms ‘concept’ and ‘noun’ have been treated as synonymous, but of course this is not the case. A noun like palm will be associated with many concepts (e.g., 4 synsets in WordNet). This is to say that the concept descriptions obtained with the methods discussed here need to be discriminated in order to obtain actual concepts; we propose methods for doing this in [41,29].

Future work will include improving upon our theory of attributes, extracting additional types of information, trying out recent developments in semi-supervised relation extraction, and developing better evaluation methods.
References


A. The 402 concepts dataset

<table>
<thead>
<tr>
<th>WordNet Unique Beginner</th>
<th>Category</th>
<th>Concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>animal</td>
<td>animal</td>
<td>bear, bull, camel, cat, cow, deer, dog, elephant, horse, kitten, lion, monkey, mouse, oyster, puppy, rat, sheep, tiger, turtle, zebra</td>
</tr>
<tr>
<td>possession</td>
<td>assets</td>
<td>allocation, allotment, capital, credit, dispensation, fund, gain, gold, hoard, income, interest, investment, margin, mortgage, payoff, profit, quota, taxation, trove, venture, wager</td>
</tr>
<tr>
<td>natural phenomenon</td>
<td>atmospheric phenomenon</td>
<td>airstream, aurora, blast, clemency, cloud, cloudburst, crosswind, cyclone, drizzle, fog, hurricane, lightning, rainstorm, sandstorm, shower, snowfall, thunderstorm, tornado, twister, typhoon, wind</td>
</tr>
<tr>
<td>substance</td>
<td>chemical element</td>
<td>aluminium, bismuth, cadmium, calcium, carbon, charcoal, copper, germanium, helium, hydrogen, iron, lithium, magnesium, neon, nitrogen, oxygen, platinum, potassium, silver, titanium, zinc</td>
</tr>
<tr>
<td>person</td>
<td>creator</td>
<td>architect, artist, builder, constructor, craftsman, designer, developer, farmer, inventor, maker, manufacturer, musician, originator, painter, photographer, producer, tailor</td>
</tr>
<tr>
<td>location</td>
<td>district</td>
<td>anchorage, borderland, borough, caliphate, canton, city, country, county, kingdom, land, metropolis, parish, prefecture, riverside, seafront, shire, state, suburb, sultanate, town, village</td>
</tr>
<tr>
<td>natural object</td>
<td>edible fruit</td>
<td>apple, banana, berry, cherry, grape, kiwi, lemon, mango, melon, olive, orange, peach, pear, pineapple, strawberry, watermelon</td>
</tr>
</tbody>
</table>

Table 4. The balanced dataset of 402 concepts used in the experiments (1)
<table>
<thead>
<tr>
<th>WordNet Unique Beginner</th>
<th>Category</th>
<th>Concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>plant</td>
<td>tree</td>
<td>acacia, casuarina, chestnut, cinchona, coco, conifer, fig, hornbeam, jacaranda, lime, mandarin, mangrove, oak, palm, pine, pistachio, rowan, samba, sapling, sycamore, walnut</td>
</tr>
<tr>
<td>artifact</td>
<td>vehicle</td>
<td>aircraft, airplane, automobile, bicycle, boat, car, cruiser, helicopter, motorcycle, pickup, rocket, slap, truck, van</td>
</tr>
<tr>
<td>feeling</td>
<td>feeling</td>
<td>anger, desire, fear, happiness, joy, love, pain, passion, pleasure, sadness, sensitivity, shame, wonder</td>
</tr>
<tr>
<td>act</td>
<td>game</td>
<td>baccarat, basketball, beano, bowling, chess, curling, faro, football, golf, handball, keno, lotto, nap, raffle, rugby, soccer, softball, tennis, volleyball, whist</td>
</tr>
<tr>
<td>state</td>
<td>illness</td>
<td>acne, anthrax, arthritis, asthma, cancer, cholera, cirrhosis, diabetes, eczema, flu, glaucoma, hepatitis, leukemia, malnutrition, meningitis, plague, rheumatism, smallpox</td>
</tr>
<tr>
<td>relation</td>
<td>legal document</td>
<td>acceptance, assignment, bill, bond, check, cheque, constitution, convention, decree, draft, floater, law, licence, obligation, opinion, rescript, sequestration, share, statute, straddle, treaty</td>
</tr>
<tr>
<td>quantity</td>
<td>monetary unit</td>
<td>cent, cordoba, dinar, dirham, dollar, drachma, escudo, fen, franc, guilder, lira, mark, penny, peso, pound, riel, rouble, rupee, shilling, yuan, zloty</td>
</tr>
<tr>
<td>motivation</td>
<td>motivation</td>
<td>compulsion, conscience, deterrence, disincentive, dynamic, ethics, impulse, incentive, incitement, inducement, life, mania, morality, motivator, obsession, occasion, possession, superego, urge, wanderlust</td>
</tr>
</tbody>
</table>

Table 5. The balanced dataset of 402 concepts used in the experiments (II)
<table>
<thead>
<tr>
<th>WordNet Unique Beginner</th>
<th>Category</th>
<th>Concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>cognition</td>
<td>pain</td>
<td>ache, backache, bellyache, burn, earache, headache, lumbago, migraine, neuralgia, sciatica, soreness, sting, stinging, stitch, suffering, tenderness, throb, toothache, torment</td>
</tr>
<tr>
<td>attribute</td>
<td>physical property</td>
<td>chill, coolness, deflection, diameter, extension, glow, heaviness, length, mass, momentum, plasticity, poundage, radius, reflexion, shortness, snap, stretch, temperature, visibility, weight</td>
</tr>
<tr>
<td>event</td>
<td>social occasion</td>
<td>ball, celebration, ceremony, commemoration, commencement, coronation, dance, enthronement, feast, fete, fiesta, fundraiser, funeral, graduation, inaugural, pageantry, party, prom, rededication, wedding</td>
</tr>
<tr>
<td>group</td>
<td>social unit</td>
<td>agency, branch, brigade, bureau, club, committee, company, confederacy, department, divan, family, house, household, league, legion, nation, office, platoon, team, tribe, troop</td>
</tr>
<tr>
<td>shape</td>
<td>solid</td>
<td>concavity, corner, crinkle, cube, cuboid, cylinder, dodecahedron, dome, droop, fluting, icosahedron, indentation, jag, knob, octahedron, ovoid, ring, salient, taper, tetrahedron</td>
</tr>
<tr>
<td>time</td>
<td>time</td>
<td>aeon, date, day, epoch, future, gestation, hereafter, menopause, moment, nonce, period, quaternary, today, tomorrow, tonight, yesterday, yesteryear</td>
</tr>
</tbody>
</table>

Table 6. The balanced dataset of 402 concepts used in the experiments (III)