

# Towards Data-driven Ontologies: a Filtering Approach using Keywords and Natural Language Constructs

Maaïke H. T. de Boer and Jack P. C. Verhoosel

TNO

Anna van Buerenplein 1, 2595 DA, The Hague, The Netherlands

maaike.deboer@tno.nl, jack.verhoosel@tno.nl

## Abstract

Creating ontologies is an expensive task. Our vision is that we can automatically generate ontologies based on a set of relevant documents to create a kick-start in ontology creating sessions. In this paper, we focus on enhancing two often used methods, OpenIE and co-occurrences. We evaluate the methods on two document sets, one about pizza and one about the agriculture domain. The methods are evaluated using two types of F1-score (objective, quantitative) and through a human assessment (subjective, qualitative). The results show that 1) Cooc performs both objectively and subjectively better than OpenIE; 2) the filtering methods based on keywords and on Word2vec perform similarly; 3) the filtering methods both perform better compared to OpenIE and similar to Cooc; 4) Cooc-NVP performs best, especially considering the subjective evaluation. Although, the investigated methods provide a good start for extracting an ontology out of a set of domain documents, various improvements are still possible, especially in the natural language based methods.

**Keywords:** knowledge representation, ontologies, text mining

## 1. Introduction

In communities it is important to have a common language. Often information standards are created to capture the common language. The creation of such standards can be supported by an ontology that defines the common concepts in the domain. In order to create such an ontology, a common approach is for experts to come together in working sessions and discuss about the main domain concepts and their relations. These sessions need the dedicated effort from people to construct the model which takes a lot of time and effort. In this paper, we aim to generate a quick start with an ontology as a starting point for further improvement by experts in the domain. The field in which ontologies are learned from available knowledge using data is named *ontology learning*. This paper is an extension of our previous papers on this topic (de Boer and Verhoosel, 2019a) (de Boer and Verhoosel, 2019b). In the first paper (de Boer and Verhoosel, 2019b), eight different methods were evaluated and in the second paper (de Boer and Verhoosel, 2019a) this evaluation was enhanced with two additional document sets and new evaluation metrics. In this paper, we focus on two of the methods and further enhance them. We also added a qualitative evaluation by human expert assessors.

In the next section, the related work is described focusing on ontology learning, including open information extraction, and the evaluation of ontologies. In section 3, the experimental set-up with the document sets, characteristics of the resulting ontologies and our evaluation methodology is explained. Section 4 contains the results of the evaluation and some discussion and section 5 concludes the paper with the conclusions and future work.

## 2. Related Work

Ontology learning is focused on learning ontologies based on data (Cimiano et al., 2009)(Brewster, 2008). One of the most known concepts in ontology learning is the ontology learning layer cake (Buitelaar et al., 2005), as displayed in Figure 1.

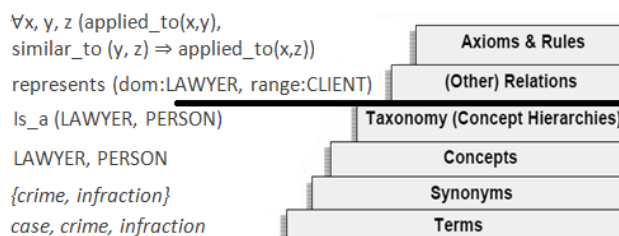


Figure 1: Layer cake taken from Buitelaar et al. (Buitelaar et al., 2005)

Starting from the bottom of the cake, the order from bottom to the top of the layer is: terms, synonyms, concept formation, concept hierarchy, relations, relation hierarchy, axiom schemata and finally general axioms. In this paper, the focus is on the lower levels of the layered cake, up to the relation hierarchy (as indicated by the line in Figure 1), i.e. the creation of triples with a subject, verb or relation, and object. The field that focuses only on the creation of these triples is named Open Information Extraction (OpenIE). The next subsection describes some tools from the OpenIE field.

### 2.1. OpenIE

The OpenIE field is quite advanced with many tools and techniques. According to a recent systematic mapping study by Glauber and Claro (Glauber and Claro, 2018), the two main steps in OpenIE methods are: 1) shallow analysis or dependency analysis for sentence annotation, such as Part of Speech (PoS) tagging or using the Stanford Dependency Parser; 2) machine learning or handcrafted rules for the extraction of relationship triples. Niklaus et al. (Niklaus et al., 2018) make the division between learning based systems, rule based systems, clause based systems and system capturing inter-propositional relationships.

One of the first OpenIE tools is TextRunner (Yates and et al., 2007). TextRunner tags sentences with PoS tags and noun phrase chunks, in a fast manner with one loop over all documents. TextRunner was followed by WOE (pos

and parse), ReVerb, KrakeN, EXEMPLAR, OLLIE, PredPatt, ClausIE, OpenIE4, CSD-IE, NESTIE, MinIE, and Graphene among others (Niklaus et al., 2018; Glauber and Claro, 2018).

Recently, deep learning methods, such as the encoder-decoder framework from Cui et al. (Cui et al., 2018), and the relation extraction method from Lin et al (Lin et al., 2016) have been proposed.

Related to the OpenIE field, query expansion can also be used to find more concepts and relations (Alfred and et al., 2014). The most common method is to use WordNet (Song et al., 2007), but other knowledge sources such as ConceptNet can also be used (de Boer et al., 2016)(de Boer and et al., 2015). Recently, Word2vec is used in information retrieval (De Boer et al., 2017) and ontology enrichment (Pembeci, 2016)(Wohlgenannt and Minic, 2016).

In this paper, the deep learning methods are not used, but a combination of the state of the art tools with the Stanford OpenIE tool, POS tagging, noun and verb phrasing, a more rule based method using co-occurrences and word2vec are used.

## 2.2. Evaluating ontologies

Brank et al. (Brank et al., 2005) state that most approaches to evaluate ontologies can be placed in one of the following categories:

- Golden Standard: compare to "golden standard"
- Application based: use in application and evaluate results
- Data-driven: involve comparisons with a data source
- Assessment by humans: human evaluation based on a set of predefined criteria, standards, and / or requirements

In this paper, we focus on the data-driven evaluation as well as an assessment by humans. In the data-driven approach the ontology is often compared against existing data about the domain. Tiddi et al (Tiddi et al., 2012) use the F-measure and precision and recall to evaluate ontology correctness by checking 1) whether attribute values are correctly extracted and 2) how much of the existing knowledge is extracted (based on DBpedia). Rospocher et al (Rospocher et al., 2012) use the same performance metrics to compare an ontology with a list of automatically extracted keywords.

McDaniel et al. (McDaniel et al., 2018) introduced the DOORS framework in which ontologies can be ranked by using syntactic, semantic, pragmatic and social quality metrics. Other papers divide the evaluation of ontologies on different levels. Brank et al. (Brank et al., 2005) distinguishes lexical, hierarchical, other semantic relations, context, syntactic, and structure levels. Burton et al. (Burton-Jones et al., 2005) uses syntactic, semantic, pragmatic and social. Gangemi et al. (Gangemi and Presutti, 2009) use the distinction between structural, functional and usability-profiling. Burton et al. (Burton-Jones et al., 2005) use lawfulness, richness, interpretability, consistency, clarity, comprehensiveness, accuracy, relevance, authority, and history.

Lozano et al. (Lozano-Tello and Gómez-Pérez, 2004) even use a three-level framework of 117 criteria. Hlomani and Stacey (Hlomani and Stacey, 2014) make the distinction between ontology quality and ontology correctness views on ontology evaluation. For ontology quality, they focus on computational efficiency, adaptability and clarity. Ontology correctness uses accuracy, completeness, conciseness and consistency.

In this paper, we evaluate the generated ontologies in a quantitative as well as a qualitative manner. For the quantitative evaluation, the keyword based objective evaluation metric of Rospocher et al (Rospocher et al., 2012) and Tiddi et al (Tiddi et al., 2012) is used. Using keywords from the subject domain to evaluate an ontology enables a check of whether the nodes in the ontology are close to domain-specific concepts. For the qualitative evaluation, a subset of the subjective quality metrics of McDaniel et al. is used (McDaniel et al., 2018) in combination with an additional quality metric that defines the clarity of the ontology. This new quality metric is based on the work of Alexopoulos and Mylonas (Alexopoulos and Mylonas, 2014) who define vagueness-oriented quality measures for ontologies.

## 3. Experimental Setup

In our experiment, we compare five different methods to create ontologies containing subject-predicate-object relations. This is done using two different document sets. We evaluate our ontologies based on a quantitative subjective metric and a qualitative objective metric. The document sets, methods and evaluation metrics are explained in the following subsections.

### 3.1. Document Sets

In our experiment, we use two different document sets, of which one is dedicated to pizzas and one is focused on our application domain of agriculture. The document sets are described below. Each of the document sets is preprocessed in the same way. From each article we first extract the plain text from the PDF. On these plain texts we use sentence splitting, tokenizing, removing non-ascii and non-textual items and non-English sentences.

**Pizza** The pizza document set is based on the information on Wikipedia. The original description of pizza is used, as well as all descriptions of pizza varieties and cooking varieties that were present as a box in the pizza description (as of date July 4th, 2019). This resulted in a set of 45 documents about pizza. This dataset can easily be reproduced, but is also available upon request.

**Agri** Our experts collected 135 articles on the agriculture domain, including agrifood, agro-ecology, crop production and the food supply chain. This dataset is currently not publicly available, but can be requested by the authors.

### 3.2. Methods

**OpenIE** We use the OpenIE (Open Information Extraction) tool created by the CoreNLP group of Stanford (Angeli et al., 2015) to extract the relations. The extracted relations are often the verbs in the sentence, and this results in triples, multiple word concepts, and many different relations.

In previous research (de Boer and Verhoosel, 2019b), it is shown that OpenIE produces many relations, and these relations decrease the precision. In our research, we want to filter the result of OpenIE in order to keep a similar recall, but upgrade the precision. This means that we want to keep the good results, but delete the bad relations. To do so, we propose two novel methods: `Filtered_openie_word2vec` and `Filtered_openie_keywords`, described below.

**Filtered\_openie\_word2vec** Word2vec is a group of models, which produce semantic embeddings. These models create neural word embeddings using a shallow neural network that is trained on a large document set, such as Wikipedia, Google News or Twitter. Each word vector is trained to maximize the log probability of neighboring words, resulting in a good performance in associations, such as *king - man + woman = queen*. In this research, we use the skip-gram model with negative sampling (SGNS) (Mikolov et al., 2013) to create a semantic embedding of each of the document sets. With the set of extracted keywords, described in the following subsections, the top ten most similar words to these keywords in the word embedding are used filter the OpenIE relations on. If a keyword or one of the top ten words is present somewhere in the relation, the triple is kept, otherwise the triple is deleted for this ontology.

**Filtered\_openie\_keywords** Keywords can be extracted using several methods. Instead of the keyword extraction method by Rospocher et al (Rospocher et al., 2012), which uses KX (Pianta and Tonelli, 2010) to get an ordered list of keywords, we combine the Term Frequency (TF) and the term extraction method from Verberne et al. (Verberne et al., 2016). The standard Wikipedia corpus from the paper is used as background set. The extracted keywords are manually inspected and all subjectively determined non-relevant terms are deleted, resulting in the following set of 12 keywords for Agri: *Data, Food, Information, Drones, Agriculture, Crop, Technology, Agricultural, Production, Development, Farmers, Supply Chain*. And for the pizza case the following 13 keywords are selected: *cheese, pizza, sauce, peppers, chicken, mozzarella, onion, tomato, pepperoni, mushroom, bacon, olive, italian*. All OpenIE relations are filtered on the keywords. If a keyword is present somewhere in the relation, the triple is kept, otherwise the triple is deleted for this ontology.

**Cooc** Besides OpenIE, it is shown that co-occurrences can create objectively good ontologies (de Boer and Verhoosel, 2019a). In co-occurrences all type of relations are extracted, in which the number of times words co-occur with each other, for example in the same sentence, are counted (Matsuo and Ishizuka, 2004). The distance between the set of pairs of different words that co-occur in the sentences of the document set is set to a maximum distance of 4 words. In the implementation the N-gram generator of the CountVectorizer module of the Scikit-learn package (Pedregosa et al., 2011) is used and cleaned with the built-in English stopword list. Because this set of co-occurring pairs of words will be very large, the set is further pruned using a threshold on the minimum number of times a pair of words co-occurs. This threshold is defined as a percent-

age of the maximum number of times a co-occurring pair of words is found. In the experiments, this number is set to 10. This number is based on experimentation with several values (ranging from 1 to 50) and overall performance seems best with 10 in our case. The ontology based on these co-occurring pairs of words will have only one vague ‘co-occurrence’ relation, indicating that the words that co-occur with each other in the document set. The specific type of relation is not determined, whereas this is specified in the OpenIE relations.

**Cooc-NVP** Although the Cooc algorithm seems to perform fairly well with respect to the concepts that are extracted from a document set (de Boer and Verhoosel, 2019a), a further improvement is necessary on the extracted relations between these concepts. Therefore, a noun phrasing and verb phrasing technique is introduced in a new algorithm, called Cooc-NVP. The algorithm starts using KL-div and NLTK POS-tagging to extract a set of noun keywords  $N$  from the document set. Then, the Cooc algorithm and NLTK POS-tagging is used to extend this set with other noun concepts that have a maximum distance in the document set of 4 words to one of the keywords. As a next step, the algorithm uses the NLTK RegexParser to parse every sentence in the document set and extracts noun phrases and verb phrases that adhere to the following simple grammar:

$NP = \langle DT \rangle ? \langle JJ. * \rangle ? \langle NN. * \rangle *$

$VP = \langle TO \rangle ? \langle VB. * \rangle +$

Thus, a noun phrase is defined as an optional Determiner optionally followed by one or more adjectives followed by one or more nouns. Additionally, the last word in a noun phrase must be one of the nouns in the constructed set of nouns  $N$ . Finally, for every combination in a sentence of two such noun phrases with last words *noun1* and *noun2* and each verb phrase  $VP$  that is positioned between these noun phrases, a triple  $\langle lemma(noun1), VP, lemma(noun2) \rangle$  is added to the ontology that is generated, where  $lemma(n)$  is the WordNetLemmatizer. The idea behind this filtering approach is that a verb phrase between two nouns is most probably the verb that expresses the natural language relation between these nouns. In addition, the last word in a noun phrase is usually the core noun of the phrase and thus forms the subject or object of the sentence. Future work needs to be done in order to further improve this approach according to natural language grammar rules.

### 3.3. Evaluation

To evaluate the created ontologies, we use a quantitative objective metric and a qualitative subjective metric. For the objective metric, we calculate the F1-score, which is based on a precision and a recall score, in two different ways. The first way is based on the formulas proposed by Rospocher et al. (Rospocher et al., 2012) (equation 1; Node-based F1), while the second way additionally takes the relations between concepts into account (equation 2; Relation-based F1). The weighted node-based metric proposed by Rospocher et al. (Rospocher et al., 2012) is not reported here, because of our evaluation showed only very small differences and similar trends compared to the node-

based metric of equation 1.

$$\begin{aligned}
 Prec_{node} &= \frac{\#k \in K_{correct}}{\#k \in Onto} \\
 Rec_{node} &= \frac{\#k \in K_{correct}}{\#k \in K} \\
 F1_{node} &= 2 * \frac{(Rec * Prec)}{Rec + Prec}
 \end{aligned}
 \tag{1}$$

where  $k$  is a keyword, which can be found in the set of correct keywords ( $K_{correct}$ ), the total set of extracted keywords ( $K$ ) and in the ontology ( $Onto$ ) to be evaluated.

$$\begin{aligned}
 Prec_{rel} &= \frac{\#r \in R \text{ with } k \in K}{\#r \in R} \\
 Rec_{rel} &= \frac{\#k \in K \text{ found in } R}{\#k \in K} \\
 F1_{rel} &= 2 * \frac{(Rec * Prec)}{Rec + Prec}
 \end{aligned}
 \tag{2}$$

where  $k$  is keyword in set of Keywords ( $K$ ),  $r$  is relation in set of Relations ( $R$ ). The set of selected items is the set of relations  $R$  (precision), and the set of relevant items is the set of keywords  $K$  (recall).

Because the ground truth is not available for both document sets, we use the KLdiv method to generate a set of evaluation-keywords as ground truth. The performance is then calculated with a varying number of generated evaluation-keywords. These evaluation-keywords will be different from the keywords used in the keyword based methods Filtered\_openie\_keywords and Cooc-NVP, because these methods only use 12 to 13 keywords generated using a different technique combined with human selection. Therefore, the results are not influenced due to the use of the same keywords in the evaluation as well as the extraction methods.

For the qualitative subjective evaluation metric, the 5 quality criteria of the DOORS framework by McDaniel et al. (McDaniel et al., 2018) and 1 new quality criterion are used. The quality criteria from McDaniel et al. used are:

- syntactic: structure and richness
- semantic: precision
- pragmatic: accuracy and adaptability.

The definitions of these criteria can be found in McDaniel et al. (McDaniel et al., 2018). In addition, we define another pragmatic criterion based on the idea of vagueness of an ontology as described by Alexopoulos et al. (Alexopoulos and Mylonas, 2014). The vagueness of an ontology is defined as the percentage of vague concepts, according to a human expert, in relation to the total number of concepts plus relations. However, we define the opposite, more positive, quality criterion, that we call "clarity", as the percentage of non-vague concepts in relation to the total number of concepts plus relations.

Each of the extracted ontologies were scored on these 6 quality criteria by two external assessors. They used the viewer of the WebVOWL tool (<http://www.visualdataweb.de/webvowl/>) to generate a graphical view of an ontology. Using zooming and searching

functionality of this tool, an impression of the ontology was created in order to score it. Scoring was done on the 5-point Likert scale, where 1 is low and 5 is high. The scoring results were then normalized to get an equally weighted score between 0 and 1, in order to be able to compare it with the quantitative evaluation results.

## 4. Results and Discussion

In this section, the results for the Pizza and Agri document sets on the different metrics are shown and discussed.

### 4.1. Pizza document set

Table 1 shows the number of classes and number of relations for the five different ontologies created for the pizza document set. The table shows that the filtering methods on OpenIE reduce the number of classes and relations considerably. The Cooc-NVP adds some classes and relations compared to Cooc. This can be explained by the fact that the number of keyword-nouns is increased by Cooc-NVP and the number of relations is determined not only by co-occurring of nouns, but also by the number of verb phrase in between these nouns.

Table 1: INSIGHTS IN THE PIZZA ONTOLOGIES

OntologyName	#Classes	#Relations
OpenIE	5,690	8,160
Filtered_openie_word2vec	539	692
Filtered_openie_keywords	492	629
Cooc	113	164
Cooc-NVP	324	1,917

Figure 2 shows the node-based F1-score for 15, 30, 50, 100, 150 and 200 automatically created keywords. Figure 3 shows the results for the relation-based F1-score. The figures show that Cooc and Cooc-NVP outperform the OpenIE methods, in which Cooc has a higher F1-score compared to Cooc-NVP in the node-based metric, whereas the difference is not that clear in the relation-based metric. The filtered methods have a higher performance in both node-based and relation-based metric compared to the OpenIE method. Interestingly, the trend of OpenIE in the relation-based metric is upwards, whereas the trend of the other methods is downwards. Looking deeper into the results, this can be clarified because the precision in the Cooc methods does not go up much, whereas the recall decreases. In the OpenIE methods, the precision increases, whereas the recall does not decrease as much.

Table 2 shows the precision and recall for the node-based and relation-based metrics for 50 keywords. The table shows that there is a slight increase in precision for the filtered methods, but a major decrease in recall for both the node and relation-based metrics. Cooc-NVP has a lower precision and recall compared to Cooc in the node-based metric and a higher precision and lower recall in the relation-based metric. This can be explained because the Cooc-NVP algorithm is targeted to only add relations for which at least one concept is a keyword.

Table 3 shows the result from the human assessment. The results show that the ontology based on OpenIE is on all aspects the worst ontology. From the filtering methods,

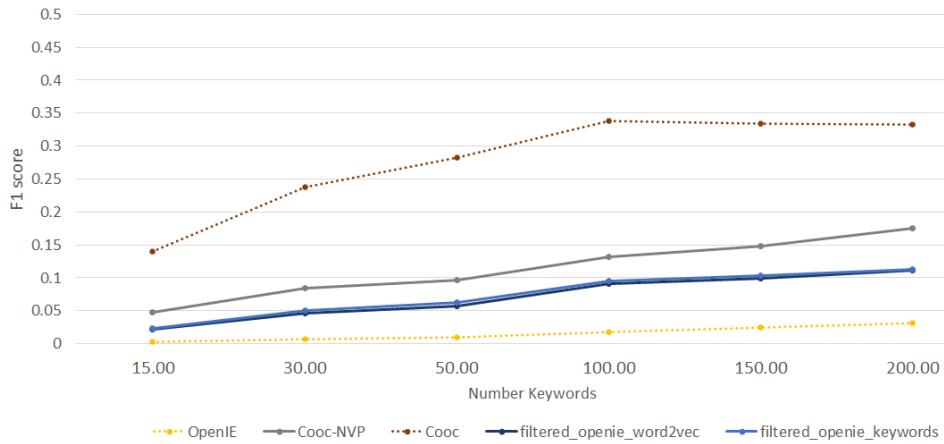


Figure 2: Node based F1 score for Pizza

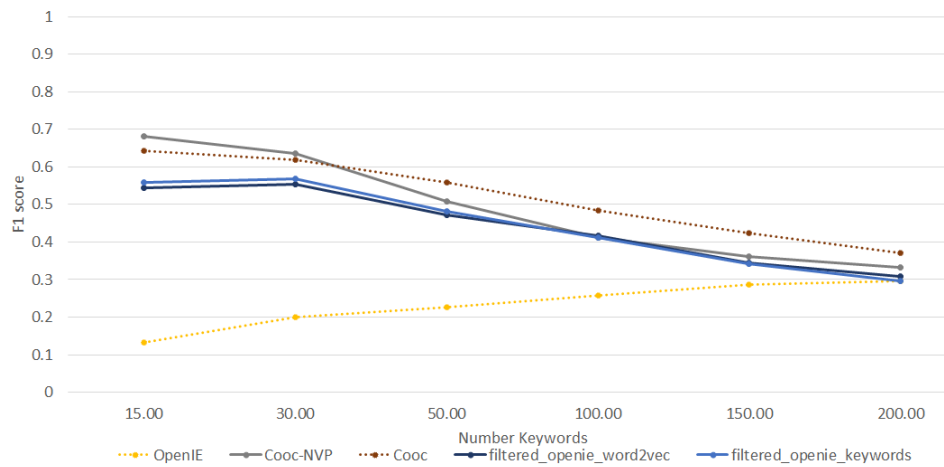


Figure 3: Relation based F1 score for Pizza

Table 2: PIZZA PRECISION AND RECALL AT K = 50

OntologyName	$Prec_{node}$	$Rec_{node}$	$Prec_{rel}$	$Rec_{rel}$
OpenIE	0.005	0.560	0.144	0.540
Filtered_openie_word2vec	0.0316	0.340	0.903	0.320
Filtered_openie_keywords	0.346	0.340	0.981	0.320
Cooc	0.203	0.460	0.768	0.440
Cooc-NVP	0.056	0.360	1.0	0.340

Table 3: SUBJECTIVE ASSESSMENT OF THE PIZZA ONTOLOGIES

OntologyName	Structure	Richness	Precision	Accuracy	Clarity	Adapt.	Normalized
OpenIE	1.0 ( $\pm 0.0$ )	1.0 ( $\pm 0.0$ )	2.0 ( $\pm 0.0$ )	1.0	2.0 ( $\pm 0.0$ )	2.0	0.3
Filtered_openie_word2vec	3.5 ( $\pm 0.5$ )	3.5 ( $\pm 0.5$ )	2.0 ( $\pm 0.0$ )	2.5 ( $\pm 0.5$ )	2.0 ( $\pm 0.0$ )	4.0	0.583
Filtered_openie_keywords	2.5 ( $\pm 0.5$ )	3.0 ( $\pm 1.0$ )	2.5 ( $\pm 0.5$ )	2.0 ( $\pm 0.0$ )	2.0 ( $\pm 0.0$ )	4.0	0.533
Cooc	2.5 ( $\pm 0.5$ )	2.5 ( $\pm 0.5$ )	2.5 ( $\pm 1.5$ )	4.0 ( $\pm 1.0$ )	4.0 ( $\pm 0.0$ )	2.0	0.583
Cooc-NVP	4.5 ( $\pm 0.5$ )	5.0 ( $\pm 0.0$ )	4.0 ( $\pm 0.0$ )	3.5 ( $\pm 0.5$ )	4.5 ( $\pm 0.5$ )	4.0	0.85

Word2vec seems slightly better compared to keywords. Cooc-NVP is overall the best method, although Cooc is the best on accuracy. The results seem not very different from the relation based results. In the comments, the assessors indicated that the OpenIE ontology has no good structure and contains many false statements. In the keyword filtering method most relations are not relevant and the concepts

are only related to pizza. In the Word2vec filtering method ‘it’ was a strong node, there was a lot of noise and only a few good concepts and relations are present. Cooc only has co-occurrence as relation type and the ontology contains many abstract nouns. In the Cooc-NVP some concepts are not related to the domain (such as ‘web’), the relations between the concepts are not all true (many are) and the leaves

of the ontology tree do not have many ancestors.

## 4.2. Agri document set

Table 4 shows the number of classes and number of relations for the different ontologies created with the agricultural document set. The table shows that the filtering methods on OpenIE significantly reduce the number of classes and relations. The Cooc-NVP increases the number of classes and relations compared to Cooc. As for the Pizza document set, this can be explained by the fact that the number of keyword-nouns is increased by Cooc-NVP and the number of relations is determined not only by co-occurring of nouns, but also by the number of verb phrase in between these nouns.

Table 4: INSIGHTS IN THE AGRI TAXONOMIES

OntologyName	#Classes	#Relations
OpenIE	280,063	535,380
Filtered_openie_word2vec	36,623	62,479
Filtered_openie_keywords	11,910	17,591
Cooc	506	1,360
Cooc-NVP	2,897	105,943

Figure 4 shows the node based F1-score for 15, 30, 50, 100, 150 and 200 automatically created keywords. Figure 5 shows a similar result for the relation based F1-score. The figures show different trends. Cooc is very prominently better compared to the other methods, whereas in the relation based F1 score the Filtered-openie-keywords, Cooc-NVP and Cooc are best performing.

Table 5 shows the precision and recall for the node based and relation based metrics for 50 keywords. In the node based metric, the filtered methods slightly increase in precision, but decrease more in recall, creating a worse F1-score. Cooc-NVP has a slight decrease in precision, but a major decrease in precision of the nodes. Reviewing the relation based metric, the filtered methods also slightly decrease in recall, but have a major increase in precision. Cooc-NVP has a precision of 1.0, but has a slight decrease in recall compared to Cooc. In all cases, Cooc has a lower recall and higher precision compared to the OpenIE based methods. This can be explained by the lower number of classes and relations.

Table 6 shows the results from the human assessment. Some values are missing compared to the pizza assessment. The assessors indicated that WebVOWL could be used for the Pizza ontologies, but that the Agri ontologies were often too big. Therefore, a normal text editor was used to evaluate the ontology. A drawback of this approach is that it was much more difficult to evaluate the ontology. Especially the adaptability was hard to judge in that way, so only one value was filled in. The results show that OpenIE is considered the least good ontology, whereas the filtered methods are slightly better, especially in richness. Cooc and Cooc-NVP seem to be the best of our five methods, especially on richness and clarity.

## 5. Conclusion and Future Work

The aim of this paper is to present and compare methods to automatically extract ontologies from a set of relevant

documents. We compare five methods: two state of the art methods, the Stanford OpenIE parser and a Co-occurrence algorithm, and three enhancement methods based on them. Two quantitative, objective metrics for automatic evaluation and one qualitative, subjective metric for human assessment are used to evaluate the created ontologies. The methods have been applied to two document sets, one for the Pizza domain and one for the Agri(cultural) domain.

The results show that the well-known information extraction method of Stanford creates a lot of concepts in an ontology that partly overlap and have, therefore, a lot of redundancy. This brings down its precision and F1-score drastically. The human evaluation supports this conclusion. The Filtered-OpenIE methods show similar performance, which is both objectively and subjectively almost comparable to the Cooc based methods. The state of the art Cooc method that extracts the most occurring words that are at a distance of maximum 4 words, surprisingly behaves fairly well according to the quantitative keyword based evaluation. The human evaluation shows that the Cooc method has a better performance compared to OpenIE, but an average performance with respect to the other methods. The newly introduced Cooc-NVP method performs on average when looking at the objective evaluation metrics, but scores highest in the human evaluation. This can most probably be explained by the fact that the extracted ontologies contain correct noun concepts of the domain and have verb phrase relations, which represent the way of thinking when designing ontologies.

When comparing the results for the two document sets, we can see that the node based F1-scores have similar increasing curves. The main difference is that the Cooc curve increases much faster for the Pizza docset than for the Agri docset. This might be because the Pizza docset is smaller than the Agri docset and thus will the curve reach its peak much earlier for increasing evaluation-keywords. In addition, the relation based F1-score decreases for the Pizza docset, while it maintains its level for the Agri docset. This can also be explained by the size of the docset. When we further increase the number of evaluation-keywords also the relation based F1-score for the Agri docset decreases.

Although the Cooc-NVP method performs fairly well, especially in the subjective evaluation, various improvements on the NLP parts of the algorithm can be added. For instance, the grammar used to define noun phrases and verb phrases can be extended considerably with other natural language grammar patterns. In addition, the combining of nouns and verb phrases to generate triples can be made more intelligent based on natural language patterns. The amount of relations between two concepts (nouns) is very large and needs to be decreased using natural language patterns. Also, the filtering approaches could be used to extend Cooc or other methods, or filter them. Summarizing, the investigated methods provide a good start for extracting an ontology out of a set of domain documents. Nevertheless, various improvements are still possible, especially in the natural language based methods.

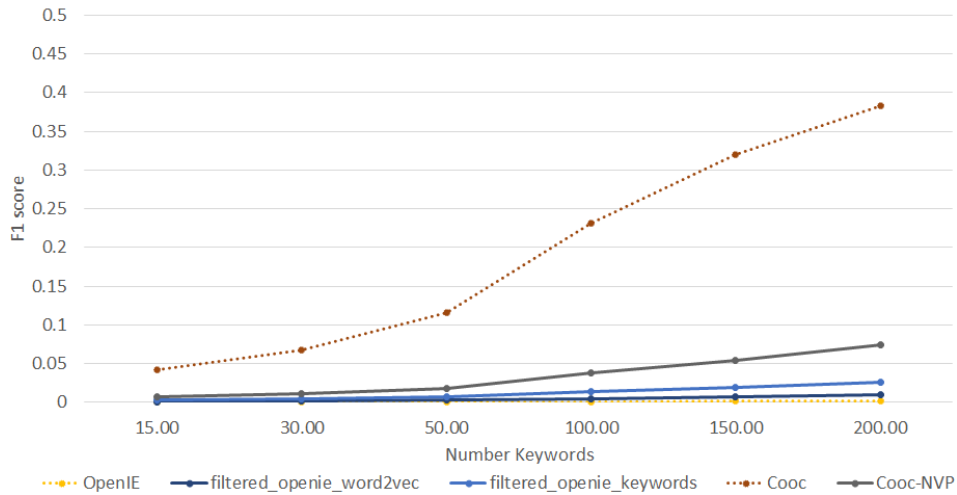


Figure 4: Node based F1 score for Agri

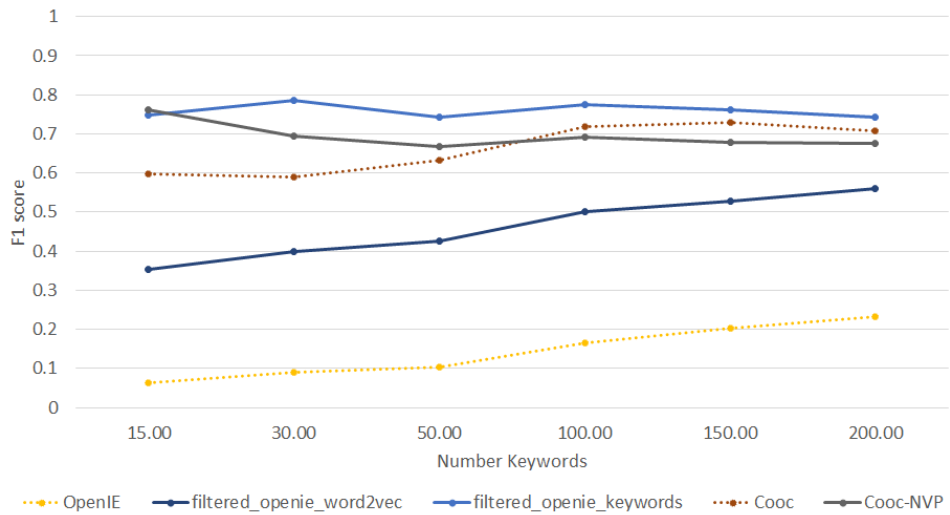


Figure 5: Relation based F1 score for Agri

Table 5: AGRI PRECISION AND RECALL AT K = 50

OntologyName	$Prec_{node}$	$Rec_{node}$	$Prec_{rel}$	$Rec_{rel}$
OpenIE	0.000	0.920	0.055	0.880
Filtered_openie_word2vec	0.001	0.860	0.286	0.820
Filtered_openie_keywords	0.003	0.780	0.746	0.740
Cooc	0.063	0.640	0.644	0.620
Cooc-NVP	0.009	0.540	1.0	0.5

Table 6: SUBJECTIVE ASSESSMENT OF THE AGRI ONTOLOGIES

OntologyName	Structure	Richness	Precision	Accuracy	Clarity	Adapt.	Normalized
OpenIE	2.0	1.0	2.0 ( $\pm 0.0$ )	2.0	2.5 ( $\pm 0.0$ )	-	0.38
Filtered_openie_word2vec	2.0 ( $\pm 1.0$ )	4.0 ( $\pm 0.0$ )	2.0 ( $\pm 0.5$ )	2.0 ( $\pm 0.0$ )	3.5 ( $\pm 1.5$ )	3.0	0.55
Filtered_openie_keywords	3.0	4.0	3.0 ( $\pm 0.0$ )	3 ( $\pm 0.0$ )	2.5 ( $\pm 0.5$ )	-	0.62
Cooc	3.0	2.0	4.0 ( $\pm 0.0$ )	3.5 ( $\pm 1.5$ )	4.5 ( $\pm 0.5$ )	-	0.68
Cooc-NVP	3.0	4.0	3.0 ( $\pm 0.0$ )	3.5 ( $\pm 0.5$ )	4.5 ( $\pm 0.5$ )	-	0.72

## 6. Acknowledgements

This work has been executed as part of the Interreg Smart-Green project (<https://northsearegion.eu/smartgreen/>). The

authors would like to thank Christopher Brewster providing a representative agriculture-related document set and Roos Bakker and Romy van Drie for the human assessment.

## 7. Bibliographical References

- Alexopoulos, P. and Mylonas, P. (2014). Towards vagueness-oriented quality assessment of ontologies. In *Hellenic Conference on Artificial Intelligence*, pages 448–453. Springer.
- Alfred, R. and et al. (2014). Ontology-based query expansion for supporting information retrieval in agriculture. In *The 8th International Conference on Knowledge Management in Organizations*, pages 299–311. Springer.
- Angeli, G., Premkumar, M. J. J., and Manning, C. D. (2015). Leveraging linguistic structure for open domain information extraction. In *Proc. of 53 ACL and 7th Int. Joint Conf. on NLP (Vol 1: Long Papers)*, volume 1, pages 344–354.
- Brank, J., Grobelnik, M., and Mladenić, D. (2005). A survey of ontology evaluation techniques.
- Brewster, C. A. (2008). *Mind the gap: Bridging from text to ontological knowledge*. Ph.D. thesis, University of Sheffield.
- Buitelaar, P., Cimiano, P., and Magnini, B. (2005). *Ontology learning from text: methods, evaluation and applications*, volume 123. IOS press.
- Burton-Jones, A., Storey, V. C., Sugumaran, V., and Ahluwalia, P. (2005). A semiotic metrics suite for assessing the quality of ontologies. *Data & Knowledge Engineering*, 55(1):84–102.
- Cimiano, P., Mädche, A., Staab, S., and Völker, J. (2009). Ontology learning. In *Handbook on ontologies*, pages 245–267. Springer.
- Cui, L., Wei, F., and Zhou, M. (2018). Neural open information extraction. *arXiv preprint arXiv:1805.04270*.
- de Boer, M. H. and et al. (2015). Query interpretation—an application of semiotics in image retrieval. *International Journal on Advances in Software*, 3 4:435–449.
- de Boer, M. and Verhoosel, J. P. (2019a). Creating and evaluating data-driven ontologies. *to appear*.
- de Boer, M. and Verhoosel, J. (2019b). Creating data-driven ontologies: An agriculture use case. In *ALL-DATA 2019: the Fifth International Conference on Big Data, Small Data, Linked Data and Open Data, Valencia, Spain 24-28 march 2019*, 52-57.
- de Boer, M., Schutte, K., and Kraaij, W. (2016). Knowledge based query expansion in complex multimedia event detection. *Multimedia Tools and Applications*, 75(15):9025–9043.
- De Boer, M. H., Lu, Y.-J., Zhang, H., Schutte, K., Ngo, C.-W., and Kraaij, W. (2017). Semantic reasoning in zero example video event retrieval. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 13(4):60.
- Gangemi, A. and Presutti, V. (2009). Ontology design patterns. In *Handbook on ontologies*, pages 221–243. Springer.
- Glauber, R. and Claro, D. B. (2018). A systematic mapping study on open information extraction. *Expert Systems with Applications*, 112:372–387.
- Hlomani, H. and Stacey, D. (2014). Approaches, methods, metrics, measures, and subjectivity in ontology evaluation: A survey. *Semantic Web Journal*, 1(5):1–11.
- Lin, Y., Shen, S., Liu, Z., Luan, H., and Sun, M. (2016). Neural relation extraction with selective attention over instances. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 2124–2133.
- Lozano-Tello, A. and Gómez-Pérez, A. (2004). Ontometric: A method to choose the appropriate ontology. *Journal of Database Management (JDM)*, 15(2):1–18.
- Matsuo, Y. and Ishizuka, M. (2004). Keyword extraction from a single document using word co-occurrence statistical information. *International Journal on Artificial Intelligence Tools*, 13(01):157–169.
- McDaniel, M., Storey, V. C., and Sugumaran, V. (2018). Assessing the quality of domain ontologies: Metrics and an automated ranking system. *Data & Knowledge Engineering*, 115:32–47.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Adv. in neural information processing systems*, pages 3111–3119.
- Niklaus, C., Cetto, M., Freitas, A., and Handschuh, S. (2018). A survey on open information extraction. *arXiv preprint arXiv:1806.05599*.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Pembeci, İ. (2016). Using word embeddings for ontology enrichment. *International Journal of Intelligent Systems and Applications in Engineering*, 4(3):49–56.
- Pianta, E. and Tonelli, S. (2010). Kx: A flexible system for keyphrase extraction. In *Proceedings of the 5th international workshop on semantic evaluation*, pages 170–173. Association for Computational Linguistics.
- Rospoche, M., Tonelli, S., Serafini, L., and Pianta, E. (2012). Corpus-based terminological evaluation of ontologies. *Applied Ontology*, 7(4):429–448.
- Song, M., Song, I.-Y., Hu, X., and Allen, R. B. (2007). Integration of association rules and ontologies for semantic query expansion. *Data & Knowledge Engineering*, 63(1):63–75.
- Tiddi, I., Mustapha, N. B., Vanrompay, Y., and Aufaure, M.-A. (2012). Ontology learning from open linked data and web snippets. In *OTM Confederated International Conferences "On the Move to Meaningful Internet Systems"*, pages 434–443. Springer.
- Verberne, S., Sappelli, M., Hiemstra, D., and Kraaij, W. (2016). Evaluation and analysis of term scoring methods for term extraction. *Information Retrieval Journal*, 19(5):510–545.
- Wohlgenannt, G. and Minic, F. (2016). Using word2vec to build a simple ontology learning system. In *ISWC*.
- Yates, A. and et al. (2007). Textrunner: open information extraction on the web. In *Proc. of Human Language Technologies: The Annual Conf. of the N.-A. Chapter of the Association for Comp. Ling.: Demonstrations*, pages 25–26. Association for Computational Linguistics.