Evaluating Open Information Extraction for Ontologization of a Thematic Domain

Gopala Krishna Koduri[1], Siva Reddy[2], Bharat Ram Ambati[2], and Xavier Serra[1]

[1] Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain
[2] Institute for Language, Cognition and Computation, University of Edinburgh, UK
`gopala.koduri@upf.edu`

Abstract. In the past decade, domain-independent approaches to information extraction have paved way for its web-scale applications. Adapting them further to acquire knowledge from thematic domains can greatly reduce the need for manual knowledge engineering. This requires understanding how amenable the assertions extracted by such approaches are to ontologization. To this extent, we propose a framework for a comparative extrinsic evaluation of the open information extraction systems. The first part of the framework compares the volume of assertions along different dimensions with an aim to understand their coverage of the domain quantitatively. In the second part, the assertions are evaluated qualitatively by employing them in three of the fundamental tasks of ontologization: entity identification, concept identification and semantic relation extraction. The framework is demonstrated using the Indian art music domain. The combined observations are shown to be useful in gathering insights about not only the quality of the assertions, but also the nature of the approaches.

1 Introduction

The advent of the semantic web and the linked open data movements have not only resulted in a growing number of community-built structured data sources, but also catalyzed the development of domain-independent approaches for extracting information from unstructured text, further enriching them. Open information extraction (OIE) is one such paradigm that has emerged in the past decade, and has been used to extract assertions from unstructured data at web-scale with a considerable success [1]. Until recently, domain-specific approaches to information extraction from text required manual knowledge engineering as a prerequisite [2]. The OIE approaches, however, do not require a prespecified vocabulary and/or relation-specific input. Therefore, adapting them to information extraction from thematic domains would alleviate the need or manual knowledge engineering.

The process of channeling the assertions extracted from these approaches into a coherent knowledge-base itself poses certain challenges. There has been little work so far to identify and address such issues. The advances in OIE, including the recent systems such as NELL[3], are largely directed towards taking advantage of the volume of web data. In doing so, the recall of the systems is often traded off for a good precision. The

---

[3] http://rtw.ml.cmu.edu/rtw/

adaptation of such systems to acquire knowledge from a given domain is an exciting prospective direction. Soderland *et.al* [3] have first attempted to adapt TextRunner system [1] to populate the knowledge-base concerning facts in the football game domain. One of their findings is that the limited recall of the OIE systems is the major bottleneck in acquiring a good coverage of the relation types.

In this paper, we propose a comprehensive framework for a comparative evaluation of the open information extraction approaches for ontologization of a thematic domain. The first part of the framework compares the volume of extracted assertions along different aspects of a given domain with an aim to understand the coverage of the domain quantitatively. In the second part of the framework, the assertions are used in three fundamental tasks of ontologization: entity identification, concept identification, and semantic relation extraction. The framework is demonstrated using Indian art music domain, and the results from each task are validated against structured content in Wikipedia and/or are manually checked as necessary. The results from the two parts of the framework, when juxtaposed against each other, give us concrete insights into the differences between the performances and the nature of the approaches.

The remainder of the paper is organized as follows. In sec. 2, the OIE approaches that we chose to compare are discussed and in sec. 3, an overview of the data we work with is presented. In sec. 4, we present the framework with various quantitative and qualitative measures for analyzing the quality of assertions extracted, and in sec. 5, we demonstrate it on the music domain. Sec. 6 concludes the paper with our remarks and future direction of this work.

2 Open Information Extraction

Information extraction is the task of obtaining a set of assertions from the natural language text, featuring the entities and the relations of the corresponding domain. The approaches are diverse ranging from those which learn from the labeled training samples for the desired set of target relations, to those which operate in an unsupervised manner. An easy access to large volume of unstructured text on the web has necessitated approaches that scale appropriately to take advantage of this data. Open information extraction aims to extract the assertions from voluminous data without requiring a pre-specified vocabulary or labeled data for relations [1].

ReVerb & OpenIE 4.0 For demonstrating our evaluation framework, we choose two state-of-the-art OIE systems: ReVerb [4] and OpenIE 4.0 [5], which are shown to have outperformed the earlier systems such as TextRunner [1], woe^{pos} and woe^{parse} [6]. ReVerb addresses the issue of incoherent and uninformative extractions[4] found with the former systems, by using few syntactic and lexical constraints. OLLIE [5] is a successor of ReVerb, and includes the noun-mediated relations which are not handled by the latter. It also incorporates the context of the assertions in the form n-ary relations. OpenIE 4.0 employs a similar methodology to that of OLLIE, to retrieve assertions using semantic role labeling, also known as shallow semantic parsing. The implementations for both ReVerb and OpenIE 4.0 are available online[5].

---

[4] We have used the terms *assertions* and *extractions* analogously.
[5] Available at https://github.com/knowitall/

$$
\frac{
\begin{array}{ccc}
\text{John} & \text{plays} & \textit{guitar} \\
\hline
NP & (S\backslash NP)/NP & NP \\
\textit{john} & \lambda x\lambda y.\ \text{plays}(subj, y) & \textit{guitar} \\
& \wedge\ \text{plays}(obj, x) &
\end{array}
}{}
$$

$$
\frac{\begin{array}{c} S\backslash NP \\ \lambda y.\ \text{plays}(subj, y) \wedge \text{plays}(obj, guitar) \end{array}}{}>
$$

$$
\frac{S}{\text{plays}(subj, john) \wedge \text{plays}(obj, guitar)}<
$$

Fig. 1. An example showing the CCG syntactic and semantic derivation of `John plays guitar'.

Semantic parsing On the other hand, deep semantic parsing is an active research topic in the natural language processing (NLP) community, which aims to obtain a complete logical form of a given sentence. It is used in applications such as question-answering systems, robotic navigation and further has several direct implications for OIE as it is domain-independent and is shown to be web-scalable [7]. To our knowledge, there is no existing literature that compares semantic parsing with the likes of ReVerb and OpenIE 4.0. We therefore built an information extraction wrapper around a state-of-the-art semantic parser and compare with the selected OIE systems. What follows is a brief description of this system.

We use Combinatory Categorial Grammar (CCG) [8] as our grammatical framework to parse natural language sentences to logical representation. CCG is known for its transparency between syntax and semantics, i.e. given the syntactic structure (CCG derivation) of a sentence, a semantic representation can be built deterministically from its derivation. Each word in a sentence is first assigned a CCG category based on its context. Each category represents the syntactic constraints that the word has to satisfy. For example, in Fig. 1, the word *plays* is assigned a syntactic category $(S\backslash NP)/NP$ implying that *plays* take a noun (NP) argument on its right, and a noun argument (NP) on its left to form a sentence (S). An equivalent semantic category in terms of a lambda function is constructed from the syntactic category, here $\lambda x.\lambda y.\text{plays}(subj,\ y) \wedge \text{plays}(obj,\ x)$ with *plays* representing the predicate, x and y representing the object (guitar) and subject (John) arguments. CCG defines a set of combinators using which the adjacent categories combine to form syntactic categories of larger text units like phrases (*e.g:* plays guitar), from there on leading to parsing a whole sentence. Correspondingly, the lambda functions of the categories compose, eventually leading to the semantic representation of the sentence. The advantage with CCG is that the complexity of obtaining a logical representation of a sentence is simplified into the task of assigning categories to words. We use a modified version of Boxer [9] to further convert our sentences of interest to the triple form (subject, relation phrase, object).

3 Data

A major challenge in developing technologies for the exploration and the navigation of music repertoires from around the world lies in obtaining and using their cultural context. The vocabulary used for describing and relating the entities (musical concepts,

roles of people involved etc) differs to a great extent from music to music. Most commercial platforms have a limited view of such context, resulting in poor navigation and exploration systems that fail to address the cultural diversity of the world. Within the music information research community, there is a growing interest for developing culture-aware approaches to address this problem [10]. Such approaches are diverse in terms of the data they work with (audio, metadata and contextual-data) and methodologies they employ [11].

However, to our knowledge, there are no major attempts that use web text, arguably the largest openly available data source. As a first step in this direction, we choose to demonstrate our framework in the music domain. Indian art music traditions: Carnatic and Hindustani, have a very distinct character, especially when compared to the popular music styles that drive the music market worldwide. The terminology and the structuring of the knowledge in these music traditions differs substantially from what people are accustomed to, on most commercial platforms [12]. Therefore, they make a suitable yet challenging thematic domain to analyze the quality of the assertions for ontologization.

Text corpus Our data consists of the plain text obtained from the Wikipedia pages corresponding to the Carnatic and Hindustani music traditions, after removing the tables, listings, figures, infoboxes and other structured content. Pages from the parent categories corresponding to both music traditions are obtained recursively. Text from each page is tokenized to sentences, which are further filtered using the following constraints: a minimum number of 3 words and a maximum of 21 words per sentence, with each word not exceeding 30 characters in length. These constraints are empirically found to reduce the number of malformed and highly complex sentences. In the semantic parsing based system, to address the multiword named entities, we identify consecutive NNPs and merge them into one in a pre-processing step. In ReVerb and OpenIE4.0, they are handled in their respective implementations.

We observed that a majority of the sentences featured pronouns. The resulting assertions only partially contribute to ontologization. For instance, consider the sentence `She is a composer'. The resulting assertion would be (She, is a, composer). A few such sentences might help us learn that there exists a concept called *composer*. However, such assertions are helpless in identifying entities of the corresponding concept. Therefore, the pronouns in the text from each page are resolved using the deterministic coreference resolution described in [13][6]. There were a few false assertions as a result. However, there is a substantial rise in the recall of the entities in the domain. Table. 1 lists the total number of sentences, and the number of assertions extracted using the OIE systems. ReVerb and Open IE 4.0 associate a confidence score with the extracted assertions. We did not however choose to filter them based on this score, as [3] advocates that a system with a better recall at the cost of lower precision is actually preferred for knowledge-base population using open information extraction. All the assertions are converted to the triple form.

Gold standard We use the ontologies manually engineered with the help of domain experts (one musicologist and three musicians) as groundtruth for the evaluation in

---

[6] Available online at http://nlp.stanford.edu/software/dcoref.shtml

Music	#Sentences	#ReVerb	#OpenIE 4.0	#Sem. Parsing
Carnatic	10284	9844	15013	19241
Hindustani	10724	9944	15777	18496

Table 1. The number of sentences for each music, and the number of extractions obtained from the OIE systems.

the tasks of concept identification and semantic relation extraction[7]. The concepts and relation-types used in our evaluation framework belong to the top most level in the ontologies and are agreed upon by all the experts. The groundtruth for entities for each concept correspond to the page-titles in the respective subcategory in the Wikipedia (eg: Carnatic_musicians).

4   Evaluation framework

In the information extraction literature, different approaches are evaluated by measuring their performances on a set of labeled sentences or by employing human judges [4, 5]. Our goal, however, is to evaluate them by the usefulness of the assertions extracted. We quantify this using a series of tasks that help in understanding the coverage of the entities and the relation types of a given domain in the extracted assertions, quantitatively and qualitatively. The tasks discussed in the first part of the evaluation compare the volume of the assertions. While in the second part, we validate to what extent the assertions yield to be structured. We then juxtapose and compare the results from both parts of the evaluation.

4.1   Quantitative assessment

We study the distribution of the extracted assertions along four different aspects to gain an insight into their coverage of the domain with respect to each of them: **sentences**, **entities**, **relation types** and **concepts**. For the purpose of analyses discussed in this section, the subject of a triple is taken for an entity, and the object of a triple featuring a subsumption relation phrase[8] (e.g: is a, be etc..) is taken for a concept. For instance, in the triplet (*Tyagaraja*, *is a*, *composer*), *Tyagaraja* is taken for an entity, and as the triple has a subsumption relation type *is a*, *composer* is taken for a concept.

Observations from the distribution of the number of extractions for sentences give a crude perspective of the modularity of the information extraction approach, which is its ability to identify multiple, distinct relations from a given sentence. The distribution of extractions for the entities allows us to gain an overview of the scope of the extracted relations in identifying the entities in the given domain as well as in describing a given entity. The distribution of extractions for relation types allows us to understand the relevance and coverage of an identified relation-type in the domain.

As we will see, a large majority of the assertions from all the OIE systems correspond to the subsumption relation type, often outnumbering the other relation types

---

[7] Available at https://github.com/gopalkoduri/ontologies
[8] A subsumption relation phrase is that which indicate a hierarchy in the taxonomy of the domain.

by orders of magnitude. Therefore, it is important to further analyze this relation type. These relations mainly inform us about the concept membership of the entities. Hence, they assume importance for ontologization as they are resourceful in defining the taxonomy of the given domain. The distribution of extractions for concepts would reveal to what extent the assertions actually carry the required information.

4.2 Qualitative assessment

Though the number of relation-types, concepts and entities are representative of the size of the ontology learned and the volume by which it is populated, the numbers alone might be misleading as not all the extractions from a given system are unique and meaningful. The differences between relative performances in quantitative and qualitative analysis expose those systems which overgenerate wrong/redundant relation-types.

Consequently, the tasks discussed in this section are complementary to those presented in the former, validating whether the quantitative observations correlate with the performances of OIE systems on various tasks in ontologization. For this, we consider the three fundamental tasks of ontologization: entity identification, concept identification and semantic relation extraction [14].

Concept identification. It is the task of identifying the concepts in the domain. Objects from all the triples featuring a subsumption relation phrase are collected. They are disambiguated based on their spellings, mostly automatically using string matching and edit distance measures[9] with minimal manual intervention where necessary. The resulting objects are taken to be the candidate concepts of the given domain. We compare the coverage of these against the concepts[10] in the manually built ontologies.

Entity identification. It concerns with finding the entities of a given domain and assigning a concept to them. The set of subjects from all the triples are considered as candidates entities in the domain. A list of titles of the Wikpedia pages in the domain along with the categories each page belong to, is acquired. The page titles correspond to entities, and the categories are manually mapped to concepts in our ontology. This constitutes the reference with which we compare the results from the two subtasks of entity identification.

For evaluating the first subtask of entity identification, i.e., finding the entities of the given domain, we measure the overlapping (O) and the residual (R) portions of the candidate entities from each system with respect to the reference set. If X is the set of candidate entities and Y is the reference set, O and R are defined as:

$$O(X,Y) = \frac{|X \cap Y|}{|Y|} \tag{1}$$

$$R(X,Y) = \frac{|X - Y|}{|X|}$$

The overlap and residual measures are preferred to other standard measures such as precision and recall as there are legitimate entities recognized in the assertions that are

---

[9] Available at https://github.com/gopalkoduri/string-matching

[10] The term concept is used analogous to the term class in ontologies.

not part of the groundtruth. For technical correctness, we chose to evaluate using custom measures that convey the same information.

The second subtask, concept assignment, is evaluated using two methods. In the first method, we manually build a set of rules over subsumption relation type for each concept. For instance, for an entity to belong to the concept *singers*, it must have either of the words *vocalist* or *singer* in the object of the corresponding triples with subsumption relation type. All the entities satisfying the rules for a given concept are assigned to it.

In the second method, a given entity is reduced to be represented by a term vector corresponding to the words from objects of the assertions it is part of. Following this, each concept is initiated with a seedset of entities belonging to it. A given concept is taken to be an abstract super entity, represented by the union of the term vectors of the constituting entities. A bootstrapping mechanism is started, which in a given iteration, finds the closest entity to the given concept and adds it to the seedset, and recomputes its representation. The distance between given two entities corresponds to the cosine similarity between the term vectors transformed using TF-IDF, followed by Latent Semantic Indexing [15]. Unlike the first method, which is constrained to assertions with subsumption relation type, this method takes advantage of the full spectrum of relation types. Results from the both methods are evaluated using O and R measures from eq. 1, where X and Y correspond to the candidate set of entities obtained using one of the methods for a given concept, and the reference set of entities respectively.

Semantic relation extraction. It refers to the relation types other than those which convey concept hierarchies. The assertion shown in Fig. 1 is one such example, where *plays* is a relation that connects *person* and *musical instrument* concepts. We compare the OIE systems in this task by two measures: breadth and depth of the identified relation types. Breadth corresponds to the absolute number of valid relation types identified for each concept, and the depth corresponds to the number of assertions for a given relation type that consist of the identified entities. The valid relation types are manually marked from among the relation phrases in the assertions.

5 Results and discussion

5.1 Quantitative assessment

The results of quantitative assessment for Carnatic and Hindustani music correlate with each other, which shows that results are consistent[11]. Fig. 2(a) shows the distribution of the number of extracted assertions using each of the OIE systems. Notice that the y-axis is a log scale. Between ReVerb and OpenIE 4.0, the latter seem to perform better, which can be attributed to the noun mediated relations. The semantic parsing based system, however, retrieves substantially more relations per sentence than these two. A tight coupling between syntax and semantics proves to be advantageous in chunking different types of assertions, as well as relating entities far off each other in a given sentence. For

---

[11] Due to space constraints, we present results only for Carnatic music. Results for Hindustani music can be found here: https://github.com/gopalkoduri/openie_eval
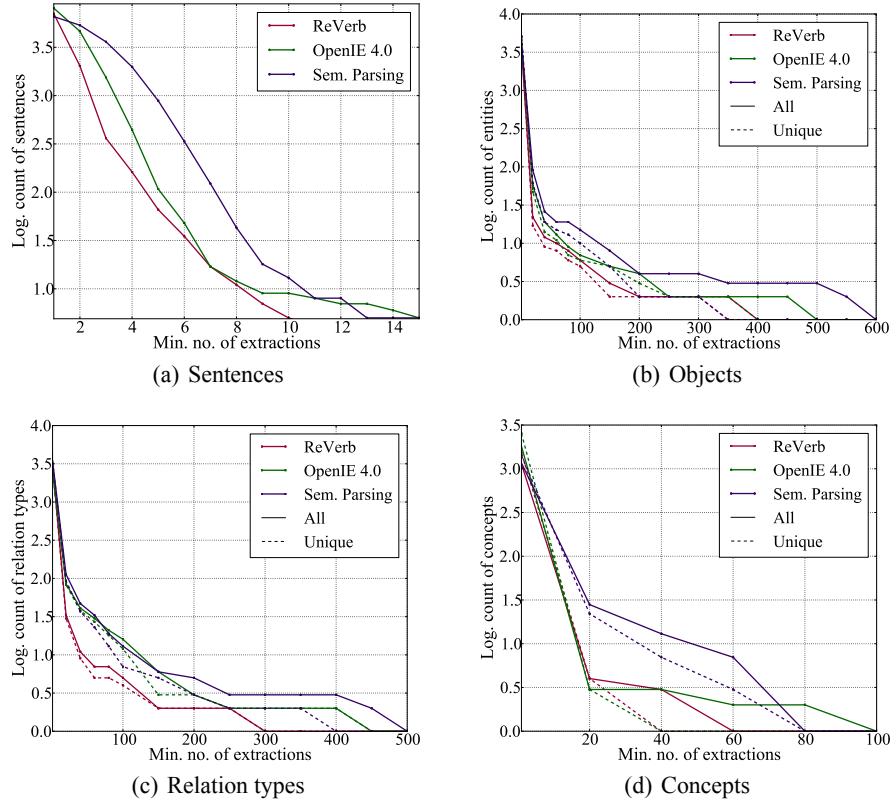
Fig. 2. Distribution of no. of extractions from OIE systems for Carnatic music shown along different *aspects*. For a given number of extractions on x-axis, the y-axis shows the logarithmic count of the instances in the *aspect*, which have at least those many extractions.

instance, in sentences which feature a single subject, but multiple relations (e.g: Chittibabu is a renowned Veena player, born in Kakinada to Ranga Rao and Sundaramma.), it performed thoroughly well compared to others.

Fig. 2(b) shows the corresponding distribution for entities. Recall that we defined an entity to be the subject of a triple. The few entities with a disproportionately high number of extractions are usually the pronouns (despite resolving most of them), followed by musical terms. The semantic parsing based system retrieves slightly more number of assertions per entity compared to OpenIE 4.0, which in turn performs better than Re-Verb. We observed some redundancy in assertions for a given entity, which is beneficial as this can be used as a measure of confidence in asserting the corresponding relation. In order to analyze this, we have also plotted the distributions of number of unique extractions for entities (shown in dashed lines in the figures).We can observe that the semantic parsing based system retrieves substantially more number of redundant assertions per entity compared to the other two.

Fig. 2(c) shows the distribution of the number of extracted assertions for the relation types. The results for semantic parsing based system and OpenIE 4.0 are more or less the same, both of which are substantially better than ReVerb. The redundancy in assertions, seen as the difference between the distributions shown by solid and dashed lines, is not as pronounced as it is for entities. The decline in the total number of relation-types as the number of extractions go higher, is less steep than in the case of entities. Unless the vocabulary in the domain is itself limited, this may indicate a slightly better coverage of relation types compared to that of the entities in the domain.

Fig. 2(d) shows the distribution of the number of extracted assertions for the concepts. Recall that a concept is defined to be the object of a triple with subsumption relation phrase. The difference between the semantic parsing based system and the other two is quite marked, with the former retrieving more assertions per concept. For Hindustani music, the coverage of concepts in the assertions of ReVerb and OpenIE 4.0 is very low, with no concept having more than 20 assertions.

To summarize, the results indicate that the semantic parsing based system has a better coverage of entities, concepts and relation types of the domain than OpenIE 4.0, which is followed by ReVerb. It also retrieves more assertions per sentence compared to the other two. Note that this section has only provided the quantitative information, which by no means is complete in itself. The results we discuss in the following section complement these providing qualitative observations along the same dimensions (i.e., entities, concepts and relation types).
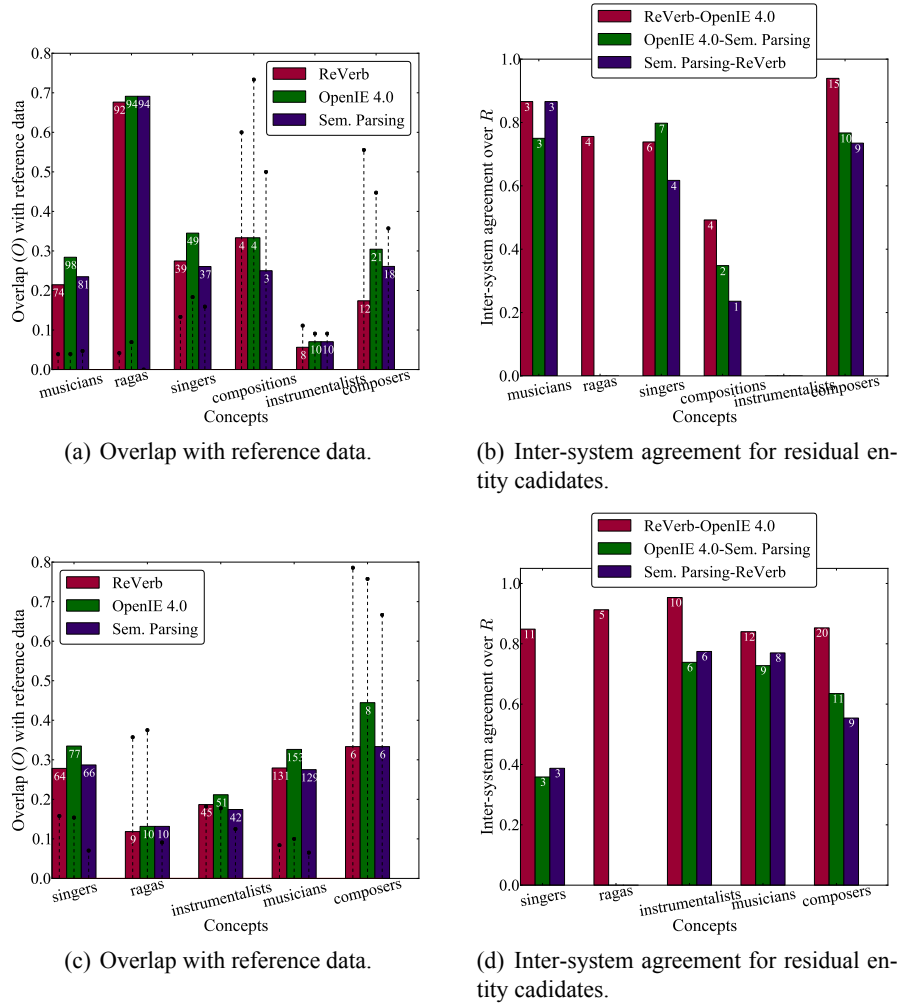
5.2 Qualitative assessment

In this section, we present the results for various tasks in the ontologization of Indian art music domain: concept identification, entity identification and semantic relation extraction.

Concept identification. Recall that we define candidate concepts to be the collection of objects from the triples with subsumption relation type. We map these to the concepts in the ontologies as described in sec. 4. Table. 2 shows the number of concepts in the ontolgies for each music, and the number of concepts mapped from the assertions of the OIE systems. These results are consistent with our earlier observations of the results shown in fig. 2(d).

Music	#Ontology	#ReVerb	#OpenIE 4.0	#Sem. Parsing
Carnatic	53	4	4	22
Hindustani	55	1	2	9

Table 2. The number of concepts in the ontologies for each music, and those mapped from the assertions of the OIE systems.

Entity identification. The first subtask in this part is to find the entities in the domain. The candidate entities from each OIE system are defined to be the collection of subjects from all its triples. Table. 3 shows the total number of entities in the reference data taken from Wikipedia for each music, and the number of entities in the intersection of these with the candidate entities of each OIE system. There is no marked difference between

(a) Overlap with reference data.

(b) Inter-system agreement for residual entity cadidates.

(c) Overlap with reference data.

(d) Inter-system agreement for residual entity cadidates.

Fig. 3. Results for rule-based concept assignment of entities identified in Carnatic (top) and Hindustani (bottom) music.

the results, with nearly all the systems having about 60% of the entities from reference data in their assertions. However, it is observed that these correspond to only about 7% of all the possible candidate entities.

For the second subtask of entity identification, i.e., assigning entities to concepts, we have considered those concepts from our ontology for which there is a corresponding category on Wikipedia, each having at least 20 pages. This was done to avoid manual labeling of entities. We found 5 such concepts for Hindustani music: musicians, singers, instrumentalists, composers and ragas[12]. For Carnatic music, in addition to these, we

---

[12] Raga is the melodic framework for both the Indian art music traditions.

Music	#Reference	#ReVerb	#OpenIE 4.0	#Sem. Parsing
Carnatic	618	349	364	364
Hindustani	697	396	410	399

**Table 3.** The number of entities in the reference data for each music, and those identified using the OIE systems.

have found another concept: compositions. As discussed, we evaluate this task using two methods: rule-based and bootstrapping-based method.

Figs. 3(a) and 3(c) show the overlap (O, see eq. 1) on rule-based concept assignment for entities found in Carnatic and Hindustani music using the OIE systems. The stem plots in the figures show residual portion of entities (R). The most notable performances are seen for the raga concept in Carnatic music. This can be attributed to two specific reasons: most Carnatic ragas on Wikipedia are described using a template, and the terminology consists mainly of Sanskrit terms which set them apart from the rest (mainly people). On the other hand, the description for Hindustani ragas varied a lot from raga to raga, and often the Sanskrit terms are inconsistently romanized making it hard for OIE systems to retrieve meaningful assertions. In theory, there is a template for almost every category, but there is a lot of variability, such as this, in describing the corresponding entities, except in the case of Carnatic ragas. OpenIE 4.0 seems to perform slightly better in terms of overlap, compared to the semantic parsing based system and ReVerb.

It is noteworthy to observe that residual entity candidates are consistently less in number for the semantic parsing based system. There are two possibilities with them: they can be either false positives, or true positives which are not found in the reference data. In most cases, they are observed to be false positives. However, there are also a few of the latter. In order to understand them further, we have plotted the inter-system agreement in figs. 3(b) and 3(d), which is given by the cosine similarity between R of different systems. ReVerb and OpenIE 4.0 agree with each other consistently higher over many of the concepts. We have observed that the cases where two or more systems agree on the candidature of a given entity, it is highly probable that the entity actually belongs to the concept. All the figures also show absolute numbers to put into perspective the proportion of residual entities where the systems agree with each other.

The second method for the evaluation of concept assignment employs bootstrapping as discussed in sec. 4. This process involves selection of a seedset and determining the number of bootstrapping iterations. For the sake of brevity, we have set the size of seedset to be the same for all the concepts, which is 3. The entities in the seedset are randomly chosen from the ones among the reference data taken from Wikipedia. However, as the bootstrapping process itself can be sensitive to the initial selection of the entities in the seedset, the whole process is repeated 5 times with randomly chosen seedsets. The bootstrapping method is terminated once the size of seedset reaches that of the corresponding concept in the reference data. After every 5 instances added during the process, we measure the overlap (O) and residual (R) portions of the seedset with respect to the reference data. Figs. 4 and 5 show their mean over 5 runs.
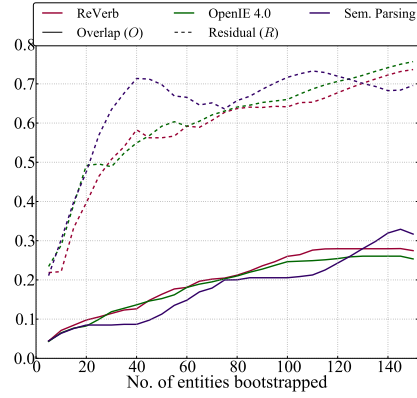
In most categories and for all the three systems, it can be seen that R grows quickly over iterations, making the residual portion the majority among the candidate entities, which brings the precision down. The semantic parsing based system consistently out-
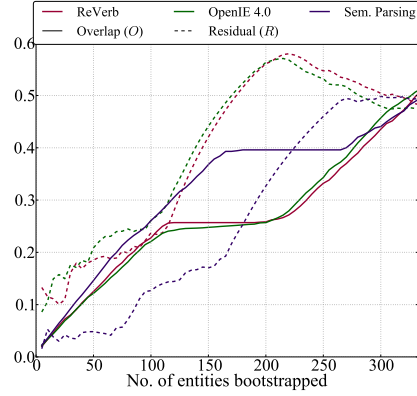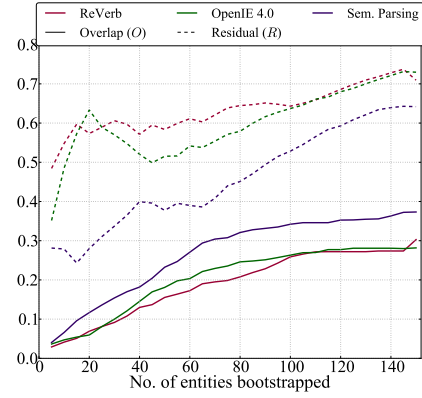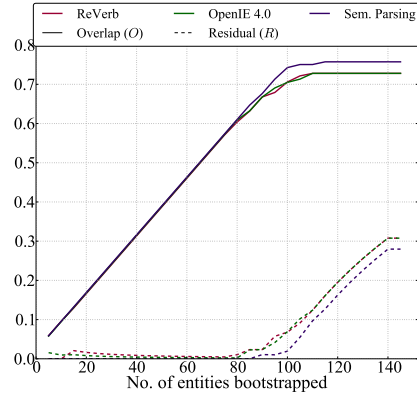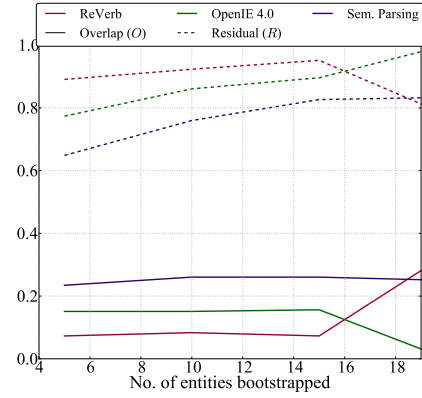
(a) Musicians
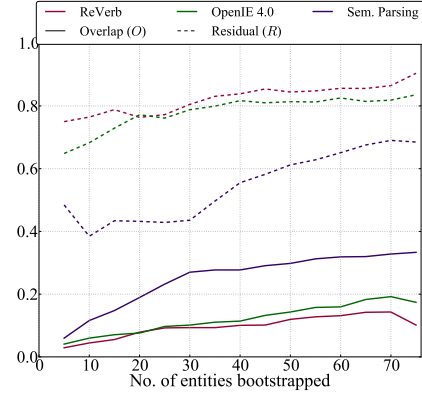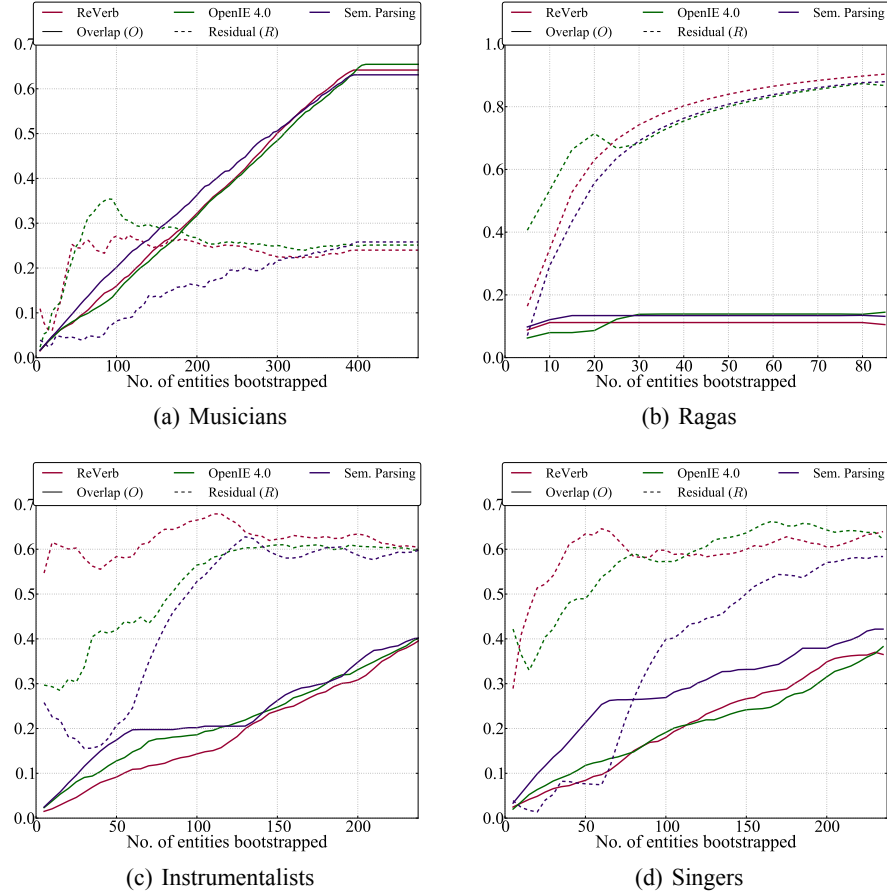
(b) Composers

(c) Instrumentalists

(d) Singers

(e) Ragas

(f) Compositions

Fig. 4. Results for bootstrapping-based concept assignment of entities identified in Carnatic music

Fig. 5. Results for bootstrapping-based concept assignment of entities identified in Hindustani music. The results for hindustani composers were not included due to space constraints.

performs the other two methods, both in terms of having higher O, and lower R. Between ReVerb and OpenIE 4.0, there is no substantial difference in terms of O. For Carnatic singer and instrumentalist categories, however, the latter results in a lower R, and a slightly higher O compared to the former. These results partly contrast with those obtained for the rule-based concept assignment (fig. 3). However, remember that the coverage of concepts in the domain is observed to be remarkably better in the case of the semantic parsing based system (fig. 2(d)). As the bootstrapping method uses the objects from the triples (which are the candidate concepts), it seems logical that the semantic parsing based system has performed substantially better than the other two.

Semantic relation extraction. For the purpose of this task, the subsumption relation types and also those relation phrases which do not have a consistent meaning across the assertions were discarded. Then, following the procedure discussed in sec.4, we have marked the valid relation types for each concept, and obtained the corresponding asser-

			Mus.	Ragas	Sing.	Compositions	Instru.	Composers	Overall
Carnatic	**#Rel.types**	R	10	5	3	0	5	3	17
		O	10	6	5	0	10	6	19
		S	12	7	10	0	9	7	27
	#Assertions	R	100	30	60	0	70	50	310
		O	240	120	160	0	230	160	910
		S	220	190	180	0	170	180	940
Hindustani	**#Rel.types**	R	8	0	7	n/a	12	1	18
		O	8	0	9	n/a	9	2	12
		S	12	0	12	n/a	12	2	22
	#Assertions	R	140	0	130	n/a	150	10	430
		O	310	0	310	n/a	320	150	1090
		S	290	0	260	n/a	280	150	880

Table 4. Semantic relation extraction task: The number of valid relation types marked for each concept, and the number of corresponding assertions that include the entities in the domain. R, O and S stand for ReVerb, OpenIE4.0 and Sem. Parsing systems respectively.

tions featuring the entities in the domain. Table. 4 shows the results, for both Carnatic and Hindustani music.

In terms of the breadth of the valid relation types (see sec. 4), the semantic parsing based system performs better than OpenIE 4.0, which in most cases fares better than ReVerb. However, in terms of the depth of relation types, both the semantic parsing based system and OpenIE 4.0 perform competitively, with the former scoring high in a few categories while the latter in few others. Compared to these two, ReVerb scores substantially less in this task. This can be attributed to two reasons: the former two handle noun-mediated relations, whereas ReVerb does not, the average number of assertions for a noun-mediated relation type is observed to be usually higher than the verb-mediated relations. The second reason can be specific to the domain, where the relations with musical concepts are mostly noun-mediated (e.g: Abhogi, is a raga in, Carnatic music). Notice that there is a strong concurrence between these results and those shown in fig. 2(c), with a noticeable correlation in the differences between the performance of the OIE systems.

6 Conclusions and Future work

In this paper, we have presented a framework for comparative evaluation of open information extraction systems for ontologization of thematic domains. We have demonstrated it using three OIE systems in ontologizing the Indian art music domain. The results lead us to better understand the behavior of the systems from different perspectives, which can be used to guide the future research in adapting OIE to thematic domains. The source code for the framework, links to various software components used in the demonstration, the ontologies and the data are made available online[13].

---

[13] https://github.com/gopalkoduri/openie_eval
Documentation and code streamlining are underway.

A particular limitation of the framework that is of concern is the availability of groudtruth for qualitative evaluation. The current framework hinges on to the structured content in Wikipedia/DBpedia to this extent. However, as our case study with Indian art music shows, they lack finer segregation of concepts (refer to the ontologies linked to in this paper). As we have seen, the results from the quantitative and the qualitative evaluation have a strong agreement with each other. We would like to further explore this correlation using other domains, in order to better understand it and adapt the framework to meet this challenge.

References

[1] Etzioni, O., Banko, M.: Open Information Extraction from the Web. Communications of the ACM **51**(12) (2008) 68--74

[2] Sarawagi, S.: Information Extraction. Foundations and Trends in Databases **1**(3) (2008) 261--377

[3] Soderland, S., Roof, B., Qin, B., Xu, S., Etzioni, O.: Adapting open information extraction to domain-specific relations. AI Magazine (2010) 93--102

[4] Fader, A., Soderland, S., Etzioni, O.: Identifying Relations for Open Information Extraction. In: Empirical Methods in Natural Language Processing. (2011)

[5] Mausam, Schmitz, M., Bart, R., Soderland, S., Etzioni, O.: Open Language Learning for Information Extraction. In: Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. (2012)

[6] Wu, F., Weld, D.: Open information extraction using Wikipedia. In: Association for Computational Linguistics. Number July (2010) 118--127

[7] Harrington, B., Clark, S.: Asknet: Automated semantic knowledge network. In: AAAI. (2007) 889--894

[8] Steedman, M.: The Syntactic Process. MIT Press, Cambridge, MA, USA (2000)

[9] Bos, J., Clark, S., Steedman, M., Curran, J.R., Hockenmaier, J.: Wide-coverage semantic representations from a CCG parser. In: COLING, Morristown, NJ, USA, Association for Computational Linguistics (2004) 1240--1246

[10] Serra, X.: A Multicultural Approach in Music Information Research. In: ISMIR. (2011) 151--156

[11] Serra, X., Magas, M., Benetos, E., Chudy, M., Dixon, S., Flexer, A., Gómez, E., Gouyon, F., Herrera, P., Jordà, S., Paytuvi, O., Peeters, G., Schlüter, J., Vinet, H., Widmer, G.: Roadmap for Music Information Research. (2013)

[12] Krishna, T.M., Ishwar, V.: Karṇāṭik Music : Svara, Gamaka, Phraseology And Rāga Identity. In: 2nd CompMusic Workshop. (2012) 12--18

[13] Lee, H., Chang, A., Peirsman, Y., Chambers, N., Surdeanu, M., Jurafsky, D.: Deterministic Coreference Resolution Based on Entity-Centric, Precision-Ranked Rules. Computational Linguistics **39**(4) (December 2013) 885--916

[14] Petasis, G., Karkaletsis, V.: Ontology population and enrichment: State of the art. In: Knowledge-driven multimedia information extraction and ontology evolution. Springer-Verlag (2011) 134--166

[15] Řehůřek, R., Sojka, P.: Software framework for topic modelling with large corpora. In: Workshop on New Challenges for NLP Frameworks, LREC, Valletta, Malta, ELRA (May 2010) 45--50