

Open Relation Extraction for Support Passage Retrieval: Merit and Open Issues

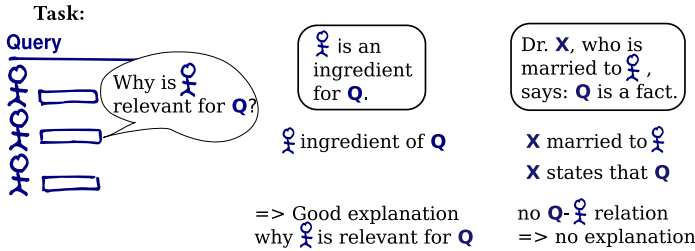
Amina Kadry a.kadry@dymatrix.de and Laura Dietz dietz@cs.unh.edu



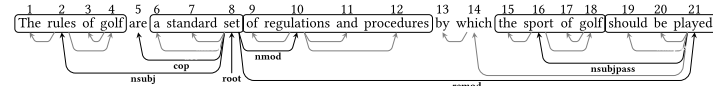
ABSTRACT

Our goal is to complement an entity ranking with human-readable explanations of how those retrieved entities are connected to the information need. Relation extraction technology should aid in finding such support passages, especially in combination with entities and query terms. This work explores how, the ClausIE system, a current state of the art method for unsupervised relation extraction (OpenIE) contributes to a solution for the task, assessing potential, limitations, and avenues for further investigation.

1 TASK & IDEA



2 FOUNDATION: CLAUSIE



Phase 1. Clause types.

Complementary clause SVC(*C:sets₈*, *V:are₅*, *S:rules₂*, *A?:of₉*)
 Adverbial clause SVA(*V:played₂₁*, *S:sport₁₆*, *S:by₁₃*)

Phase 2. Propositions of relation tuples.

The rules₂ of golf are₅ a standard set₈ of regulations
 The rules₂ of golf are₅ a standard set₈ of procedures
 The rules₂ of golf are₅ a standard set₈
 the sport₁₆ of golf should be played₂₁ by₁₃ a standard set₈ of regulations
 the sport₁₆ of golf should be played₂₁ by₁₃ a standard set₈ of procedures

Assumption: Whenever entity and query terms are contained in the same proposition, this sentence is likely to explain the connection between query and entity.

3 FEATURES FOR RELEVANCE EXPLANATIONS

Feat.	Description
Text	
1	sentence length measured in number of words
2	sentence position measured as a fraction of the document
3	fraction words that are stop words
4	fraction of query terms covered by sentence
5	sum of ISF of query terms (ISF is inverse sentence frequency)
6	average of ISF of query terms
7	sum of TF-ISF of query terms
8	number of entities mentioned
NLP	
9-12	for nouns/verbs/adjectives/adverbs: fraction of words with POS tag
13	whether sentence contains a named entity
14-16	for NER types PER/LOC/ORG: whether NER of type is contained
DP	
17	number of edges on the path between two entities in dependency tree
18	indicator whether path goes through root node
19	indicator whether path goes through query term
ClausIE	
20	whether ClausIE generated an extraction from this sentence
21-27	for all seven clause types: whether clause of this type is extracted
28	proposition length measured in tokens
29	maximum constituent length (size of dependency tree) in proposition
30-32	for subject/object/both: if another entity is in subject and/or object position of the proposition
33-34	for subject/object position: if given entity is in position of proposition
35-36	for subject/object position: if an entity link is in position of proposition
37-38	for subject/object position: if an entity link is in position of prop.
39-41	for subject/verb/object position: if a query term (ignoring stopwords) is in position of proposition
42-43	for subject/object position: if a named entity (NER) is in position of proposition

4 EXPERIMENTAL EVALUATION

Test collection, available: www.cs.unh.edu/~dietz/appendix/openie4ir

- ten 2013/2014 TREC Web track queries
- and (up to) 10 relevant entities per query from REWQ gold standard.¹
- resulting in 75 support passage rankings (one per query and entity).
- 22,731 annotations for candidate sentences from entity's Wiki page
- 2,906 are marked as relevant for AQ1

Assessors imagine writing knowledge article on the topic Q . Which information to include about the given entity e_i ?

AQ1) Explanation: Does the sentence explain the relevance of entity e_i ?

AQ2) Relation: Does the sentence mention any relationship involving e_i ?

AQ3) Rel rel: Is this relationship relevant for the explanation?

AQ4) ClausIE: Does ClausIE extract a valid relationship from sentence?

AQ5) ClausIE rel: Is ClausIE's extraction relevant for the explanation?

We study these annotations in combination with two heuristics:

Qterm: Does the sentence include query terms (stopwords ignored)?

Name: Does the sentence include the entity's name?

4.1 Experiment 1: Any hope?

Table 1: Performance of AQ1–5 as predictors for explanations.

	Relation	Rel rel	ClausIE	ClausIE rel	Qterm (*)	Name
Prec(*)	0.46 ±0.05	0.52 ±0.05 *	0.45 ±0.05	0.49 ±0.05 *	0.38 ±0.04	0.33 ±0.05
Recall	0.28 ±0.03	0.21 ±0.03	0.20 ±0.02	0.14 ±0.02	0.49 ±0.04	0.43 ±0.04
ρ	0.27	0.52	0.33	0.49	0.47	0.35
Count	1767 (8%)	935 (4%)	1172 (5%)	636 (3%)	4476 (20%)	6173 (27%)

4.2 Experiment 2: Automatic Method with LTR

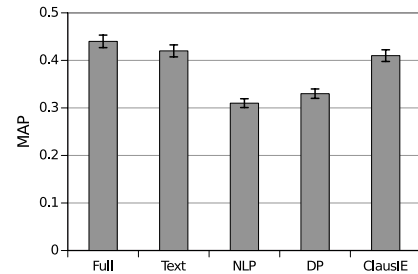


Figure 1: Ranking of sentences explaining entity relevance.

4.3 Experiment 3: Open Issues

- Among all sentences that express a relation, ClausIE is missing this relation in 32% of the cases.
- Only half of the sentences with relation expressions actually actually contain a relation that is relevant for the query-entity pair
- Only 636 sentences with relevant ClausIE extractions (3%) of all 22731 annotated sentences.
- In contrast, our data set contains 2906 sentences (13%) with explanations of relevance.
- An ideal ranking would obtain a MAP value of $\frac{1172}{2906} = 0.41$ (theoretical upper bound). Equals ClausIE feature set alone, MAP 0.41, cf. Figure 1.
- We conclude that our approach obtains an optimal ranking under limitations imposed by the off-the-shelf OpenIE system.
- Improving coverage of OpenIE systems is likely to translate to immediate quality improvements for text-ranking tasks.

¹<http://mschuhma.github.io/rewq/>