

# ***RelExt: A Tool for Relation Extraction from Text in Ontology Extension***

Alexander Schutz and Paul Buitelaar

German Research Center for Artificial Intelligence (DFKI GmbH),  
Language Technology Lab,  
Stuhlsatzenhausweg 3,  
Saarbrücken, Germany  
{aschutz, paulb}@dfki.de

**Abstract.** Domain ontologies very rarely model verbs as relations holding between concepts. However, the role of the verb as a central connecting element between concepts is undeniable. Verbs specify the interaction between the participants of some action or event by expressing relations between them. In parallel, it can be argued from an ontology engineering point of view that verbs express a relation between two classes that specify domain and range. The work described here is concerned with relation extraction for ontology extension along these lines. We describe a system (RelExt) that is capable of automatically identifying highly relevant triples (pairs of concepts connected by a relation) over concepts from an existing ontology. RelExt works by extracting relevant verbs and their grammatical arguments (i.e. terms) from a domain-specific text collection and computing corresponding relations through a combination of linguistic and statistical processing. The paper includes a detailed description of the system architecture and evaluation results on a constructed benchmark. RelExt has been developed in the context of the SmartWeb project, which aims at providing intelligent information services via mobile broadband devices on the FIFA World Cup that will be hosted in Germany in 2006. Such services include location based navigational information as well as question answering in the football domain.

## **1 Introduction**

An investigation of the structure of existing ontologies via the Swoogle ontology search engine [1]<sup>1</sup> has shown that domain ontologies very occasionally model verbs as relations holding between their concepts. However, the role of the verb as a central connecting element between concepts is undeniable. Verbs specify the interaction between the participants of some action or event by expressing relations between them.

In parallel, it can be argued from an ontology engineering point of view, that verbs express a relation between two classes that specify the domain and range of some action or event. For instance, consider the following German sentence from the football domain:

*Ballack schiesst das Leder ins Netz.*

---

<sup>1</sup> <http://swoogle.umbc.edu/>

*(Ballack shoots the ball into the net.)*

A valuable contribution for an ontology in the football domain would be that the verb “*schiessen*” (to shoot) as a relation holds between the concept FOOTBALLPLAYER, instantiated as “Ballack”, as domain and the concept BALLOBJECT, instantiated as “Leder” (“leather, ball”) as range, that is:

Rel:SHOOT (Dom:FOOTBALLPLAYER, Range:BALLOBJECT)

The work described here is concerned with the extension of a football ontology along these lines, in the context of the SmartWeb <sup>2</sup> project. SmartWeb aims at providing services accessible via mobile broadband devices in the context of the FIFA World Cup, which is hosted in Germany in 2006. Such services include location based infrastructural information (i.e. “Show me the fastest route to the stadium.”) as well as question answering in the football domain (i.e. “Who caused the penalty that Ballack converted?”).

The ontology that was constructed for this purpose consists of the following components: the upper model DOLCE [2] as foundational ontology, SUMO [3] for describing cross-domain concepts, the domain-specific SportEventOntology, which was modelled by domain experts and is focused mainly on football, and other components such as the navigation and discourse ontology. As of now <sup>3</sup>, the ontology contains 1570 direct classes (concepts) and 487 direct relations. Relations relevant for the football domain are mostly properties of some class such as “hasAge”, “hasName”, “atMinute”. SUMO does model verbal relations between classes, for instance “causes”, “connects”, “knows” or “shows”, but these are still rather abstract. However, in domain ontologies relations need to be specified more precisely.

Therefore, in the approach we present here we implemented a system (RelExt) that is capable of automatically identifying highly relevant *triples* (pairs of *concepts* connected by a *relation*) that can be integrated in an (already existing) ontology. RelExt works by extracting relevant terms and verbs from a given text collection and computing relations between them through a combination of linguistic and statistical processing.

The remainder of this document is structured as follows. Section 2 compares our approach to other ongoing relevant research in relation extraction and ontology learning in general. In Sect. 3, we give a detailed overview of the components used in the system, as well as the processing steps undertaken. Section 4 describes the evaluation strategy and the methods we used for evaluation, and goes on with interpreting our results. Finally, Sect. 5 points out ideas for further work to be carried out in this direction and closes with concluding remarks.

## 2 Related Work

A large collection of methods for ontology learning from text have developed over recent years as witnessed by the proceedings of various workshops in this area, e.g.

<sup>2</sup> <http://www.smartweb-projekt.de/>

<sup>3</sup> July 2005.

at ECAI 2002<sup>4</sup>, ECAI 2004<sup>5</sup>. Unfortunately, there is not much consensus within the ontology learning community on the exact task they are concerned with, which makes a comparison of approaches difficult<sup>6</sup>.

In order to estimate the state-of-the-art in ontology learning, we first need to establish the subtasks that together constitute the complex task of ontology development (either manual or with any level of automatic support). Ontology development is primarily concerned with the definition of concepts and relations between them. In our case this implies the acquisition of linguistic knowledge about the terms that are used to refer to a specific concept in text and possible synonyms of these terms. An ontology further consists of a taxonomy backbone (is-a relation) and other, non-hierarchical relations.

Recent work on relation extraction from text, other than the is-a relation, has been addressed primarily within the biomedical field as there are very large text collections readily available (e.g. PubMed<sup>7</sup>) for this area of research. The goal of this work is to discover new relationships between known concepts (i.e. symptoms, drugs, diseases) by analyzing large quantities of biomedical scientific articles (see e.g. [5] [6] [7]).

Most of the work on text mining combines statistical analysis with more or less complex levels of linguistic analysis, e.g. by exploiting syntactic structure and dependencies for relation extraction as reported for instance by [8], [9] and [10]. Relation extraction is therefore also very much related to the problem of acquiring selection restrictions for verb arguments in NLP (compare [11]), as witnessed for instance by the ASIUM system that enables an integrated acquisition of relations between concepts identified in text and so-called sub-categorization frames for the verbs that underlie these relations [12].

Relation extraction through text mining for ontology development was introduced in work on association rules in [13]. Of specific interest to the work described here is also recent research by Reinberger and Spyns [14], and by Sabou [15], both of which employ dependency structure for ontology learning.

While Reinberger and Spyns employ mainly statistical methods based on frequency information over linguistic dependencies (predicate-object, preposition-headnoun) in order to establish relations between entities from a corpus of the biomedical domain, they are not concerned with labelling the discovered relations, which moves their research more towards the work proposed by Maedche and Staab [13].

Sabou conducts her research on a corpus of controlled language from Web Service descriptions, that consists of simple sentence constructions from which ontology fragments can be extracted easily. Unfortunately, the proposed evaluation of Sabou's system cannot be performed automatically and needs a lot of manual interference.

### 3 Approach

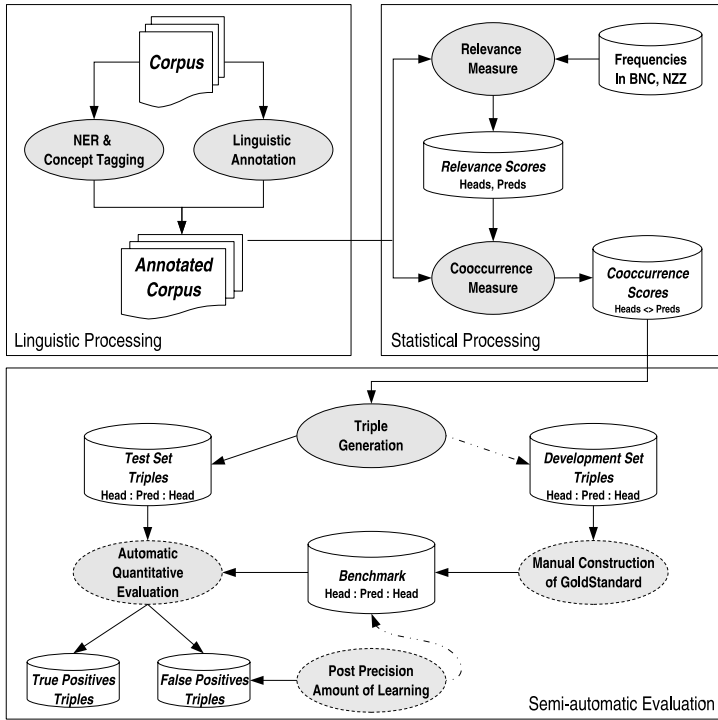
Here we describe an approach to relation extraction for ontology extension based on linguistic analysis and a predefined ontology that we intend to extend with relations derived from predicate argument structure.

<sup>4</sup> <http://www-sop.inria.fr/acacia/WORKSHOPS/ECAI2002-OLT/>

<sup>5</sup> <http://olp.dfki.de/ecai04/cfp.htm>

<sup>6</sup> A start towards surveying research in this area has been made by OntoWeb deliverable 1.5 [4].

<sup>7</sup> <http://www.pubmedcentral.nih.gov/>



**Fig. 1.** System processing architecture

What follows is the description of the corpus we used (3.1) and the system architecture, including linguistic annotation (3.2), the various stages of statistical processing and filtering (3.3), and finally, after the identification of relevant terms and relations, the construction of triples. The processing pipeline is sketched in Fig. 1.

### 3.1 Corpus Description

We worked on a document collection compiled from the football domain, consisting of 0.5 mio tokens in 1219 documents.<sup>8</sup> The documents comprise minute-by-minute (live-ticker) reports on football matches from the first and second German division. Figure 2 shows an example document of the corpus.

The benefits of this kind of text compared to the much more detailed match reports are twofold: Firstly, the sentences are rather concise, which significantly reduces the error rate of grammatical function assignment of our parser. Secondly, the language used in the minute-by-minute texts is not as prosaic as the language used in the detailed match reports, which reduces the amount of (sometimes newly invented) synonyms for domain specific terms. The average sentence length of the corpus is approximately 13 words.

<sup>8</sup> The corpus was compiled from <http://www.kicker.de>

---

Anpfiff

16: Überraschende Führung für Energie Cottbus: Miriuta zirkelt einen Freistoß über die Bremer Mauer ins rechte obere Toreck.

34: Nach einem öffnenden Zuspiel von Skripnik kommt auf der rechten Seite Stalteri an den Ball, dringt in den Strafraum ein und überwindet mit einem Rechtsschuss den herauseilenden Gäste-Keeper Piplica.

...

Schlusspfiff

---

Kickoff

16: Energie Cottbus surprisingly take the lead: Miriuta curls a freekick over the Bremen wall into the top corner.

34: From a penetrating pass by Skripnik, Stalteri takes possession on the right wing, moves into the penalty area and beats the on rushing visitor's keeper Piplica with a right footer.

...

Final whistle

---

**Fig. 2.** Example document from the corpus

### 3.2 Linguistic Analysis

For the linguistic annotation, we used the SCHUG-system [16] [17], which provides a multi-layered XML-format for a given text, specifying *dependency structure* along with *grammatical function assignment*, *phrase structure*, *part-of-speech* and *lemmatization* (including *decomposition*, which is useful in particular for German where compound nouns are often used). Figure 3 provides an example.

**Dependency Structure.** As mentioned before in Sect. 1, verbs specify an action or event, whereas the semantic classes of their syntactic arguments account for the class of participants in that event. Exploiting this information could be very useful when it comes to restricting a relation to hold only between a small set of semantic classes.

On the phrase level, SCHUG is able to provide a detailed analysis of syntactic arguments, which involves decomposition of complex NPs into nominal head, pre- and post modifier. Considering the whole NP as a candidate term for relation extraction would introduce data sparseness, and therefore, it is important to normalise a complex NP to its headnoun. Using only headnouns can be seen as a step towards normalisation, which eases concept tagging and therefore, broadening coverage.

**Named Entity Recognition / Concept Tagging.** In order to map instances of football players in the corpus to existing ontology class labels, we performed Named Entity Recognition (NER), based on gazetteer lists. The gazetteers were automatically generated from semi-structured documents about football matches in the first and second German division, containing formal data, such as team lineup, referee names and further information about a given match. For instance, if we encountered the string “Oliver Kahn”<sup>9</sup> (or “Kahn”, or “O. Kahn”), we tagged the tokens with the correct named entity type, in this case GOALKEEPER. For NER, we distinguished only between 4 different ontology classes: GOALKEEPER, FOOTBALLPLAYER, TEAM and COACH.

---

<sup>9</sup> A German goalkeeper.

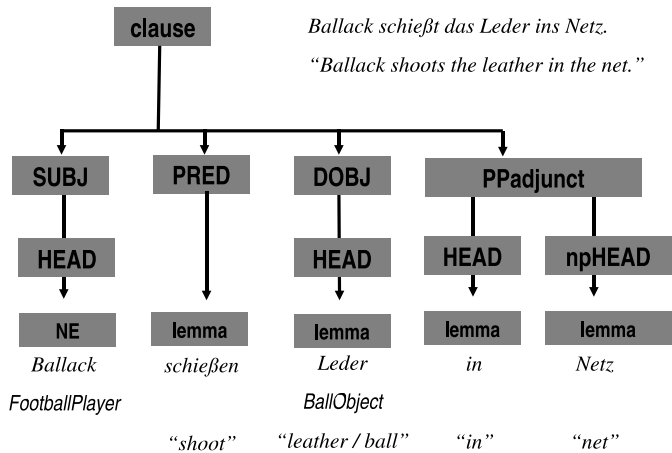


Fig. 3. Analyzed dependency structure for “Ballack schießt das Leder ins Netz.”

Furthermore, a concept tagging step was undertaken, in order to map synonyms for given terms to the corresponding ontology concepts. For this purpose we used synonyms that the SportEventOntology specifies for a given *concept label* in German and English. For instance, the concept DEFENDER has a synonym list with the following elements: [DE: *Abwehrspieler, Abwehr, Verteidiger*] [EN: *Defender, Defense, Back, Fullback, Defenceman*]. We exploited this information also for mapping terms from more specific subclasses to more general superclasses, i.e. if we encountered the token “Manndecker” (stopper / DEFENDER), we tagged the token with the more general term/label FOOTBALLPLAYER instead, in order to reduce the sparse data problem. Ambiguity in concept tagging is not really an issue here as we are working with a domain specific corpus and ontology in which there is mostly a one-to-one mapping between terms and concepts.

3.3 Statistical Processing

In order to identify the most relevant terms and relations for the football domain, it is necessary to filter out more general terms. As our goal was not only to find single relevant terms, but highly relevant triples, a single statistical ranking step was not sufficient in order to produce satisfactory results. In fact, we had to perform several computations on the extracted data, starting from *relevance ranking*, and *cross-referencing* relevant nouns and verbs with the predicate-argument-pairs, to computing *co-occurrence-scores* in order to construct triples that are specifically used in the football domain.

**Relevance Measure.** In the context of *ontology learning*, a promising approach is to select the *nominal heads* of noun phrases as candidates to be modelled as classes in an ontology, while verbs (or rather the *predicates* they express) bear information about the relationship between two classes. We therefore exploit the rich linguistic information provided by SCHUG and extracted two lists from the processed corpus. The first list contains lemmatized headnouns, while the second list consists of lemmatized predicates.

Adopting the methods of [18](chapter 5.3), a  $\chi^2$  test was used to compute a relevance ranking, comparing the observed frequencies of headnouns in the domain specific corpus with the frequencies of the same headnouns in a larger and more general corpus. As a general corpus, we relied on the British National Corpus (90 mio tokens) for English texts, and a corpus compiled from Swiss newspapers (9 mio tokens) for German texts. The same procedure was used to rank the predicates. The formula for  $\chi^2$  is given below.

$$\chi^2 = \sum_{i,j} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}} . \quad (1)$$

However, since we are dealing with 2x2 contingency tables only, it simplifies to

$$\chi^2 = \frac{N(O_{11}O_{22} - O_{12}O_{21})^2}{(O_{11} + O_{12})(O_{11} + O_{21})(O_{12} + O_{22})(O_{21} + O_{22})} . \quad (2)$$

where the indices refer to the column and row of the table,  $O$  is the observed frequency and  $N$  the sample size.

For instance, the noun “Ball” (*ball*) occurred 6849 times in our corpus of approx 0.5 mio tokens and only 511 times in the approx 9 mio tokens general corpus, obtaining a higher  $\chi^2$ -score than “Tor” (*goal*), which occurred more frequently than “Ball” in our corpus, due to the squared sums of the mean error.

According to this ranking, we obtained three lists ordered by relevance, one for headnouns, a second list for headnouns mapped to ontology class labels, and a third list for predicates. To illustrate highly relevant terms for the football-domain, Table 1 lists the top 10 for headnouns and Table 2 gives the top 10 class labels after mapping headnouns to the ontology. Table 3 displays highly relevant lemmatized verbs used in the football-domain.

**Table 1.**  $\chi^2$ -top lemmatized headnouns

Rank	$\chi^2$ -score	Headnoun	Frequency
1	125245.24	Ball ( <i>ball</i> )	6849
2	121888.52	Tor ( <i>goal</i> )	7767
3	95003.21	Meter ( <i>meters</i> )	5967
4	64157.18	Schuss ( <i>shot / drive</i> )	3575
5	57185.76	Eck ( <i>corner</i> )	3132
6	45474.96	Strafraum ( <i>penalty area</i> )	2298
7	34668.11	Freistoss ( <i>freekick</i> )	1752
8	30017.75	Leder ( <i>leather / ball</i> )	1561
9	27989.09	Flanke ( <i>cross</i> )	1479
10	27414.66	Pfosten ( <i>post</i> )	1457

**Co-occurrence Measure.** After filtering out those elements from the  $\chi^2$ -sorted lists where the score did not indicate strong relevance for the football-domain, we examined the dependency structure of the remaining predicates. We considered only those predicate-argument-pairs for further investigation, where a highly ranked predicate co-occurred with a highly ranked headnoun. We then ranked the resulting list of predicate-argument-pairs again by further statistical processing. Assuming a headnoun together with its grammatical function as one unit, *co-occurrence-scores* were computed again with  $\chi^2$ , as described below.

**Table 2.** Top ontology class labels after NER and concept tagging

Rank	Class Label	Frequency
1	FOOTBALLPLAYER	28494
2	GOALOBJECT	8188
3	BALLOBJECT	7249
4	GOALKEEPER	6887
5	SHOOT	3578
6	TEAM	2477
7	PENALTYAREA	2298
8	FREEKICK	1752
9	WING	1482
10	POST	1457

**Table 3.**  $\chi^2$ -top lemmatized predicates

Rank	$\chi^2$ -score	Predicate	Frequency
1	27167.41	flanken ( <i>to cross / to centre</i> )	1373
2	22045.39	klaeren ( <i>to clear</i> )	1435
3	21908.37	schiessen ( <i>to shot</i> )	1503
4	20439.09	koepfen ( <i>to head</i> )	1033
5	16342.99	lassen ( <i>to let / to leave</i> )	826
6	9563.41	ziehen ( <i>to pull / to drag</i> )	1548
7	9468.57	passen ( <i>to pass / to play</i> )	814
8	7752.84	spielen ( <i>to play / to pass</i> )	1559
9	7653.68	lenken ( <i>to divert</i> )	537
10	7637.45	parieren ( <i>to parry / to save</i> )	405

We obtained a ranked (by  $\chi^2$ -score) list, consisting of predicates paired up with one of their arguments in a specific grammatical function. A higher score for a predicate-headnoun-pair with a particular grammatical function means that this headnoun occurring with this particular grammatical function is statistically more likely to appear for that predicate than the same headnoun in any other grammatical function.

By this computation, we determined the *selectional preferences* of each predicate, which are semantic restrictions on syntactic arguments of the grammatical function for a given predicate and headnoun, and which in turn were used for the construction of triples. Selectional preferences have been used also in previous research on ontology learning [19].

Table 4 illustrates some of the selectional preferences for the verb “flanken” (*to cross*) and the verb “prüfen” (*to try / to test*), computed by the  $\chi^2$ -algorithm. We exploited the computed selectional preferences in order to find the most preferred *subjects* for a given verb as well as the most preferably selected *direct* and *indirect objects*.

**Relation Extraction.** The triples were constructed from the selected headnoun-predicate pairs, where the subject was chosen as the domain of the relation, while the (direct and indirect) objects as well as the adjuncts defined the range, as shown in Table 5. The steps undertaken for each verb (in order to combine it with appropriate terms for domain and range) were as follows:

1. compose a sub-unit consisting of a predicate and a highly ranked OBJ or NP-Head of PP\_ADJUNCT
2. glue a highly ranked SUBJ to the lefthand side of the sub-unit
3. SUBJ + sub-unit constitutes a triple



**Table 4.** Selectional preferences for “flanken” (*to cross*) vs. “pruefen” (*to try / to test*)

Predicate	ARG-CLASS	GF	$\chi^2$
flanken ( <i>to cross</i> )	FOOTBALLPLAYER	SUBJ	25.03
flanken	REFEREE	SUBJ	0.05
flanken	FOOTBALLPLAYER	DOBJ	34.77
flanken	REFEREE	DOBJ	0.01
flanken	FOOTBALLPLAYER	IOBJ	10.63
flanken	REFEREE	IOBJ	0.01
pruefen ( <i>to try / to test</i> )	FOOTBALLPLAYER	SUBJ	3.09
pruefen	GOALKEEPER	SUBJ	0.63
pruefen	FOOTBALLPLAYER	DOBJ	0.20
pruefen	GOALKEEPER	DOBJ	20.60
pruefen	FOOTBALLPLAYER	IOBJ	0.96
pruefen	GOALKEEPER	IOBJ	7.69

**Table 5.** Examples of constructed triples

Domain	Relation	Range
FOOTBALLPLAYER	flanken ( <i>to cross</i> )	FOOTBALLPLAYER
FOOTBALLPLAYER	flanken_auf ( <i>to cross to</i> )	FOOTBALLPLAYER
FOOTBALLPLAYER	flanken_zu ( <i>to cross to</i> )	FOOTBALLPLAYER
FOOTBALLPLAYER	pruefen ( <i>to test</i> )	GOALKEEPER

Our inspections however showed that the huge amount of single occurring predicate-headnoun-pairs (which unfortunately obtained a high  $\chi^2$  value) biased the construction of accurate, relevant triples to a large extent. We therefore introduced a threshold, considering only those elements for triple construction that co-occurred more than once.

## 4 Evaluation

A big problem for ontology learning efforts is performance evaluation, as performance in an open set is to be measured. How would one want to measure something that has been learned, but which is not yet known? Because, if it had been known in the first place, there would not have been the need for a learning effort.

As of now, various promising proposals have been made for comparing ontologies on the lexical as well as on the taxonomic level [20], which could be used in order to evaluate against a *gold standard*. Still, what happens when something learned shows up that is not covered by the gold standard? Certainly, these are problems that have to be addressed as well.

### 4.1 Benchmark Construction

Nevertheless, we did measure the performance of our system against a gold standard that we constructed to benchmark different parameters. For this purpose, we split up the corpus into 4 equally sized sub-corpora of 300 documents, from which we used one sample for benchmark construction.

From this sample (consisting of 101536 tokens) we generated 192 triples and presented these to 3 domain experts. It was their task to determine whether a given triple was an appropriate one for the football domain, or not. We also allowed an “in-between” rating, to be used if the annotator could not make up his mind. The annotation process was



Fig. 4. WWW evaluation interface

performed via a web-based interface, as shown in Fig. 4. The Kappa-statistic [21], which was computed in order to determine agreement among the 3 annotators, was found to be at roughly 27%. As this value was rather low, we additionally considered the per-class agreement [22] of the annotators, as presented in Table 6. Given reasonable agreement among the annotators' judgement on appropriate and not appropriate generated relations, we constructed two benchmarks. The strict benchmark  $GS_{strict}$  consists of 26 elements, containing only those triples which all annotators agreed upon to be appropriate. The relaxed benchmark  $GS_{relaxed}$  excluded all those triples, which were rated as not appropriate by at least one of the domain experts, leading to a set of 38 triples.

Table 6. Per-class agreement among the 3 annotators for evaluation of 192 triples

Judgement	Assigned	Agreed	Disagreed	Agr. Ratio
appropriate	202	264	140	65%
undecided	112	50	174	22%
not appropriate	262	300	224	57%

## 4.2 Experiment

As it was our aim to find relations for the extension of the SmartWeb SportEventOntology, we put the following restrictions on generated triples: Firstly, we considered only those triples, where the terms for domain and range could already be mapped to ontology classes. Secondly, we further reduced the sets of triples to be evaluated by removing items containing auxiliary and modal verbs.

**Setup.** Introducing a parametrizable setup option for triple construction, we distinguished between two setups which controlled the amount of triples to be generated. Setup I generated only one triple per verb (namely the best, according to the selectional preferences for SUBJs, OBJs and NP-heads of PP\_ADJUNCTs), while setup II generated all possible triples (by considering less preferred SUBJs and OBJs for triple composition), resulting in a set of triples with larger size.

We therefore generated 6 sets of triples, 2 for each of the 3 test corpora. Furthermore, we introduced a linear order based on the sum of the selectional preferences that were used in order to compose a given triple.<sup>10</sup> This enabled us to rank the triples within a set.

According to the ranking, we constructed 2 samples of different size. The first sample (A) was scaled exactly to the size of the Gold Standard by considering only the first best triples, and the second sample (B) simply contained all triples of the set.

As a result, we now obtained highly relevant relations connecting two highly relevant terms. Some examples are given in Table 5.

**Metrics and Results.** Various contributions in recent and ongoing work (i.e. [23], [24] or [25]) are concerned with establishing metrics for quantitative evaluation for ontology learning. However, as pointed out briefly in the introduction of this section, this effort is rather difficult. Sabou [23] proposes an evaluation strategy to be carried out over different stages, addressing issues like *extraction performance*, *ontology building support*, *domain coverage* and *fitness for the task at hand*. Therefore, the well-established metrics *recall* and *precision* are employed, and from them, new derived metrics like *Lexical Overlap* or *Ontological Improvement* are proposed to operate on the ontological level. Still, many of the evaluation stages rely on a manual inspection or consultation of domain experts.

As we are evaluating triples against a gold standard, we decided to use only the classic metrics *recall* and *precision* which are given below, measuring the system's performance on the 3 test corpora which were not used for benchmark creation.

$\mathbf{T}_s$  reflects the set of triples for a given sample, while  $\mathbf{GS}$  denotes the set of triples contained in the benchmark.

$$recall = \frac{|\mathbf{T}_s \cap \mathbf{GS}|}{|\mathbf{GS}|} \quad (3)$$

$$precision = \frac{|\mathbf{T}_s \cap \mathbf{GS}|}{|\mathbf{T}_s \cup \mathbf{GS}|} \quad (4)$$

Table 7 and Table 8 display the evaluation results for the 3 test corpora with different samples, as described in Sect. 4.2.

With respect to the benchmark, recall improves with a larger sample, but precision remains low around 10%. However, an inspection of the false positives showed that some triples were in fact appropriate<sup>11</sup>, although they were not contained in the gold standard. Clearly, those cases affect particularly the precision score of the evaluation in a negative manner. In order to account for this situation, Kavalec and Svátek [25] have proposed an additional notion of *posterior precision*, to be assessed after inspection and

<sup>10</sup> The selectional preferences for verb-OBJ (range) and verb-SUBJ (domain), intuitively.

<sup>11</sup> After re-consultation of the annotators.

**Table 7.** Performance for samples generated with setup I from 3 different test-corpora

Corpus			Recall	Precision			
# of Evaluated Triples				<i>a priori</i>		<i>a posteriori</i>	
				percentage	true positives	percentage	true positives
# of	1	38	15.8%	8.6%	6	20.0%	14
Triples	2	38	23.7%	13.4%	9	23.9%	16
=  GS	3	38	15.8%	8.6%	6	20.0%	14
Average over Samples			18.4%	10.2%		21.3%	
# of	1	95	39.5%	12.7%	15	24.6%	29
Triples	2	84	34.2%	11.9%	13	23.9%	26
= ALL	3	92	34.2%	11.1%	13	23.1%	27
Average over Samples			36.0%	11.9%		23.9%	

**Table 8.** Performance for samples generated with setup II from 3 different test-corpora

Corpus			Recall	Precision			
# of Evaluated Triples				<i>a priori</i>		<i>a posteriori</i>	
				percentage	true positives	percentage	true positives
# of	1	38	13.2%	7.0%	5	18.3%	13
Triples	2	38	21.1%	11.8%	8	19.1%	13
=  GS	3	38	15.8%	8.6%	6	15.7%	11
Average over Samples			16.7%	9.1%		17.7%	
# of	1	148	44.7%	10.1%	17	20.7%	35
Triples	2	136	42.1%	10.1%	16	20.3%	32
= ALL	3	146	42.1%	9.5%	16	19.6%	33
Average over Samples			43.0%	9.9%		20.2%	

re-consultation of a domain expert or ontology engineer. Following their line of research, triples from the set of false positives which were found to be relevant, were treated as such, and a recomputation of precision (*a posteriori*) was performed, leading to a significant improvement of the value reported as *a priori* precision. The difference between *prior precision* and *posterior precision* would be a possible way of measuring the amount of learning.

## 5 Conclusions and Future Work

In this paper we described an approach for extracting and evaluating highly relevant relations holding between ontology classes in the football domain. In contrast to the majority of the work carried out in ontology learning, we are concerned with the extraction of domain specific verbal relations other than *is-a*. As our approach is directed towards ontology extension, we rely on an already existing ontology for some domain, in order to map highly relevant headnouns to concept labels. Given that, we claim our approach to be robust and easily adjustable to different domains, as the main steps rely on statistical processing of formerly extracted linguistic information.

We are not (yet) concerned with clustering of extracted relations, which would bring together different predicates as synonyms for a single more abstract relation label. In this way, the relation will be defined as an abstraction over individual English or German verb forms (i.e. predicates).

The *RelExt*-system is implemented as a modular system, which contributes methods for the extraction procedure, the various statistical filtering steps and the triple generation.

Its modular structure allows for easy integration of new methods and composition of processing steps at will, which we think is very beneficial for tuning efforts.

The evaluation procedure we pursued supplies us with insights into overall system performance, while the different setups allow conclusions to be drawn about the performance of subcomponents of the system. As one further step, we propose the incremental extension of the gold standard.

An aspect that certainly has to be focussed on is the generation of higher quality triples, in order to improve precision without lowering recall. This can be done by taking external linguistic resources into account, i.e. interfacing with WordNet [26] or accessing information from subcategorization frame lexica for a given verb.

As SCHUG (the linguistic analysis) introduces a lot of ambiguity in grammatical function assignment (i.e. specifying multiple subjects and/or direct objects per clause), the work carried out here has a very nice side effect. The computed *selectional preferences* can be used in order to support SCHUG when it cannot decide which grammatical function to assign to a given phrase, which will in turn produce linguistic annotations of higher quality.

## Acknowledgements

This research has been supported by BMB+F (German Ministry of Education and Research) grant 01 IMD01 A for the SmartWeb project.

## References

1. Ding, L., Finin, T., Joshi, A., Pan, R., Cost, R.S., Peng, Y., Reddivari, P., Doshi, V.C., Sachs, J.: Swoogle: A Search and Metadata Engine for the Semantic Web. In: Proceedings of the Thirteenth ACM Conference on Information and Knowledge Management, ACM Press (2004)
2. Gangemi, A., Guarino, N., Oltramari, A., Schneider, L.: Sweetening ontologies with dolce. In: Proceedings of EKAW 2002, Siguenza, Spain (2002)
3. Niles, I., Pease, A.: Towards a standard upper ontology. In: FOIS '01: Proceedings of the international conference on Formal Ontology in Information Systems, ACM Press (2001) 2–9
4. Gomez-Perez, A., Manzano-Macho, D.: A survey of ontology learning methods and techniques. deliverable 1.5, ontoweb project (2003)
5. Rindflesch, T., Tanabe, L., Weinstein, J., Hunter, L.: Edgar: Extraction of drugs, genes, and relations from biomedical literature. In: Pacific Symposium on Biocomputing. (2000)
6. Pustejovsky, J., Castano, J., Zhang, J., Cochran, B., Kotecki, M.: Robust relational parsing over biomedical literature: Extracting inhibit relations. In: Pacific Symposium on Biocomputing. (2002)
7. Vintar, S., Todorovski, L., Sonntag, D., Buitelaar, P.: Evaluating context features for medical relation mining. In: ECML/PKDD Workshop on Data Mining and Text Mining for Bioinformatics. (2003)
8. Buitelaar, P., Olejnik, D., Sintek, M.: A protégé plug-in for ontology extraction from text based on linguistic analysis. In: Proceedings of the 1st European Semantic Web Symposium (ESWS). (2004)

9. Ciramita, M., Gangemi, A., Ratsch, E., Saric, J., Rojas, I.: Unsupervised learning of semantic relations between concepts of a molecular biology ontology. In: Proceedings of the 19th International Joint Conference on Artificial Intelligence. (2005) accepted for publication.
10. Gamallo, P., Gonzalez, M., Agustini, A., Lopes, G., de Lima, V.S.: Mapping syntactic dependencies onto semantic relations. In: Proceedings of the ECAI Workshop on Machine Learning and Natural Language Processing for Ontology Engineering. (2002)
11. Resnik, P.: Selection and information: A class-based approach to lexical relationships (1993)
12. Faure, D., Nedellec, C.: A corpus-based conceptual clustering method for verb frames and ontology. In Velardi, P., ed.: Proceedings of the LREC Workshop on Adapting lexical and corpus resources to sublanguages and applications. (1998) 5–12
13. Maedche, A., Staab, S.: Discovering conceptual relations from text. In Horn, W., ed.: Proceedings of the 14th European Conference on Artificial Intelligence (ECAI'2000). (2000)
14. Reinberger, M.L., Spyns, P.: Discovering knowledge in texts for the learning of DOGMA-inspired ontologies. In: Proceedings of the ECAI 2004 Workshop on Ontology Learning and Population. (2004) 19–24
15. Sabou, M.: Extracting ontologies from software documentation: a semi-automatic method and its evaluation. In: Proceedings of the ECAI-2004 Workshop on Ontology Learning and Population (ECAI-OLP). (2004)
16. Declerck, T.: A set of tools for integrating linguistic and non-linguistic information. In: Proceedings of SAAKM (ECAI Workshop). (2002)
17. Buitelaar, P., Declerck, T., Sacaleanu, B., Vintar, S., Raileanu, D., Crispi, C.: A multi-layered, xml-based approach to the integration of linguistic and semantic annotations. In: Proceedings of EACL 2003 Workshop on Language Technology and the Semantic Web, Budapest, Hungary (2003)
18. Manning, C.D., Schütze, H.: Foundations of statistical natural language processing. MIT Press (1999)
19. Faure, D., N'edellec, C.: Asium: Learning subcategorization frames and restrictions of selection. In Kodratoff, Y., ed.: 10th Conference on Machine Learning (ECML 98) – Workshop on Text Mining. (1998)
20. Maedche, A., Staab, S.: Measuring similarity between ontologies. In: EKAW '02: Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web, Springer-Verlag (2002) 251–263
21. Carletta, J.: Assessing agreement on classification tasks: the kappa statistic. *Comput. Linguist.* **22** (1996) 249–254
22. Poesio, M., Vieira, R.: A corpus-based investigation of definite description use. *Comput. Linguist.* **24** (1998) 183–216
23. Sabou M., Wroe C., G.C., G., M.: Learning domain ontologies for web service descriptions: an experiment in bioinformatics. In: Proceedings of the 14th International World Wide Web Conference WWW2005. (2005)
24. Spyns, P., Reinberger, M.L.: Evaluating ontology triples generated automatically from texts. In: Proceedings of the second European Conference on the Semantic Web, LNCS, Springer Verlag (2005)
25. Kavalec, M., Svaték, V.: A study on automated relation labelling in ontology learning. In Buitelaar, P., Cimiano, P., Magnini, B., eds.: *Ontology Learning from Text: Methods, Evaluation and Applications*. IOS Press (2005) 44–58
26. Fellbaum, C.: *WordNet: An Electronic Lexical Database*. MIT Press (1998)