Appendix - Algorithm for expansion policy

Algorithm 1: Search candidates for named entities and expansion policy

---

Data: $N = \{N_1, N_2 \ldots N_n\}$ sorted in ascending order of their string length, trigram similarity threshold σ

Result: set of candidates C

begin

    heuristicExpansion $\longleftarrow \emptyset$, **C** $\longleftarrow \emptyset$;

    for $N_i \in N$ **do**

        label $\longleftarrow$ **string**(N_i);

        tmp $\longleftarrow$ **label**;

        expansion $\longleftarrow$ **false**;

        for key $\in$ **heuristicExpansion do**

            if key contains label then

                if tmp.length $>$ **key.length** $\&\&$ **tmp** $! =$ **label then**

                    tmp $\longleftarrow$ **key**;

                    expansion $\longleftarrow$ **true**;

                if tmp.length $<$ **key.length** $\&\&$ **tmp** $==$ **label then**

                    tmp $\longleftarrow$ **key**;

                    expansion $\longleftarrow$ **true**;

        label $\longleftarrow$ **tmp**;

        if $\neg$**expansion then**

            heuristicExpansion $\longleftarrow$ **label** $\cup$ **heuristicExpansion**

        $C \longleftarrow C\cup$ **searchCandidates**(**label**,σ);

end

Table 1: Test corpora specifications

Corpus	Language	#Doc.	#Ent.	Avg.(Ent./Doc.)	Annotation
Reuters-21578	English	145	769	5.30	voter agreement
RSS 500	English	500	1,000	2.00	domain expert
news.de	German	53	627	11.83	domain expert
AIDA-YAGO2	English	1,393	34,956	25.07	voter agreement

Results on other corpora with different KBs

Table 2: Evaluation of AGDISTIS. Bold indicates best F-measure.

Corpus	AGDISTIS					
Knowledge Base	DBpedia			YAGO2		
	F-measure	σ	d	F-measure	σ	d
Reuters-21578	**0.78**	0.87	2	0.60	0.29	3
RSS 500	**0.75**	0.76	2	0.53	0.82	2
news.de	**0.87**	0.71	2	—	—	—-
AIDA-YAGO2	0.73	0.89	2	0.58	0.76	2

Table 3: Evaluation of AGDISTIS against AIDA and DBpedia Spotlight. Bold indicates best F-measure.

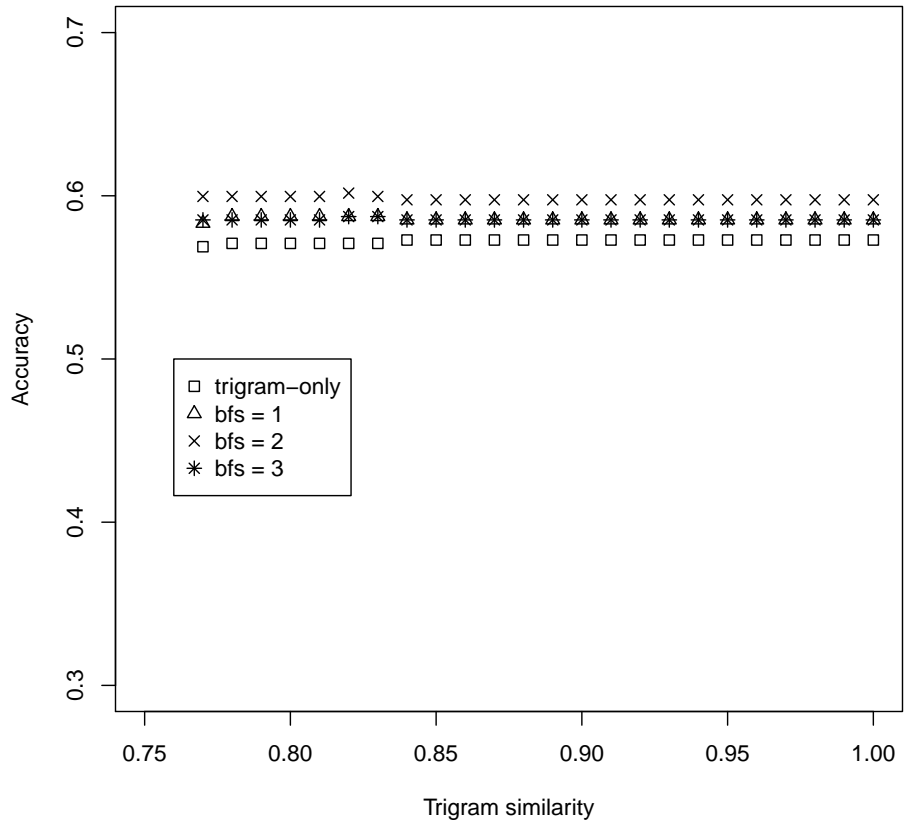| Corpus	AIDA	Spotlight
Knowledge Base	YAGO2	DBpedia
	F-measure	F-measure
Reuters-21578	0.62	0.56
RSS 500	0.60	0.40
news.de	—-	0.84
AIDA-YAGO2	**0.83**	0.57

Appendix - Figures of experimental results

Figure 1: Accuracy of AGDISTIS on the **RSS 500** corpus.

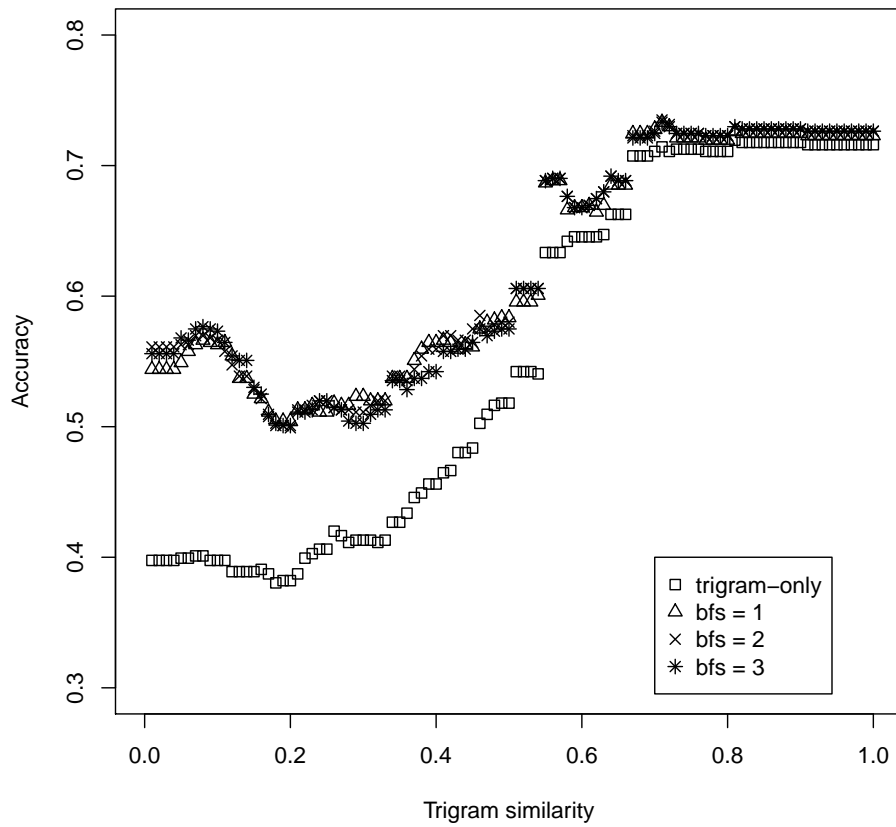# Appendix - Quality Rater Tool

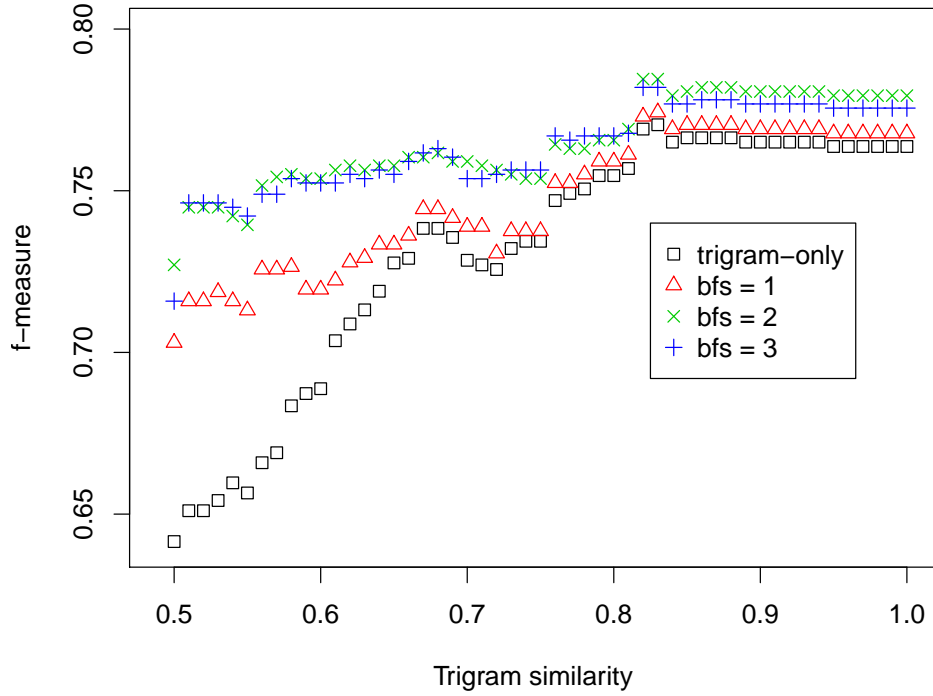Figure 2: Accuracy of AGDISTIS on the **news.de** corpus.

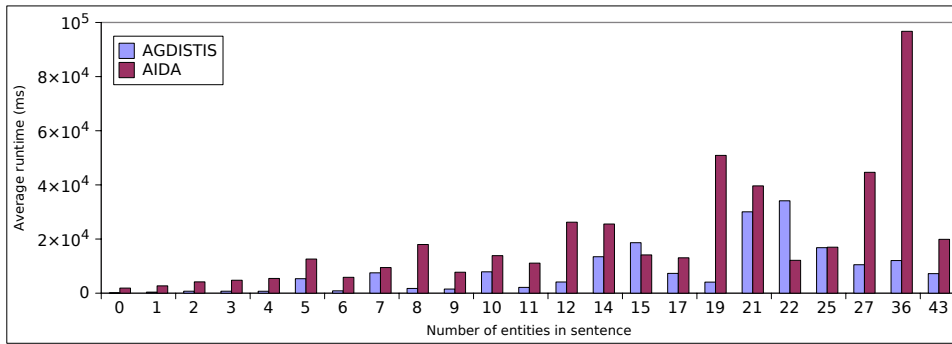Figure 3: Accuracy of AGDISTIS on the **reuters.de** corpus.

Figure 4: Comparison of the average runtime of AGDISTIS and AIDA on Reuters-21578 corpus with respect to the number of entities per sentence. AGDISTIS is cleary more time-efficient than AIDA.
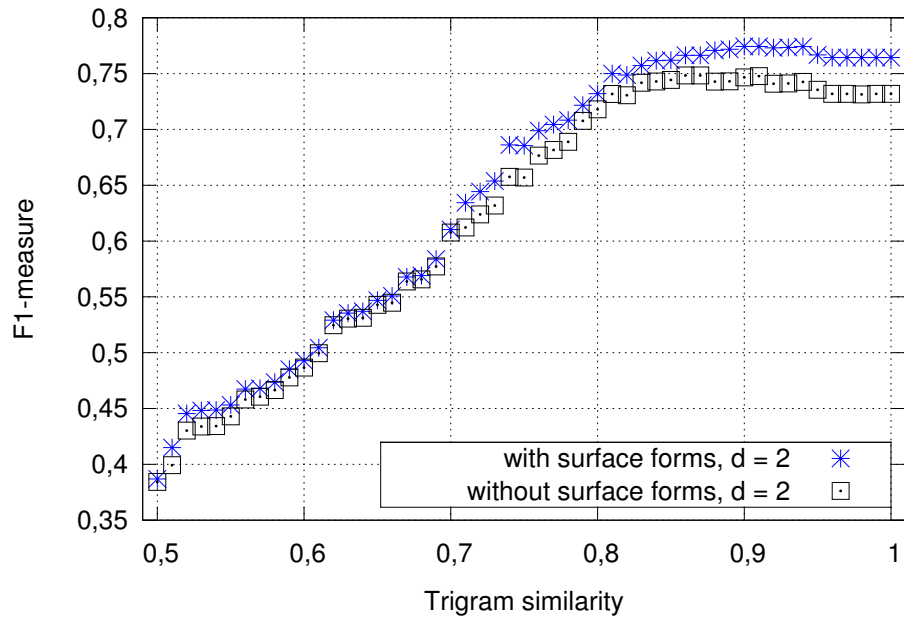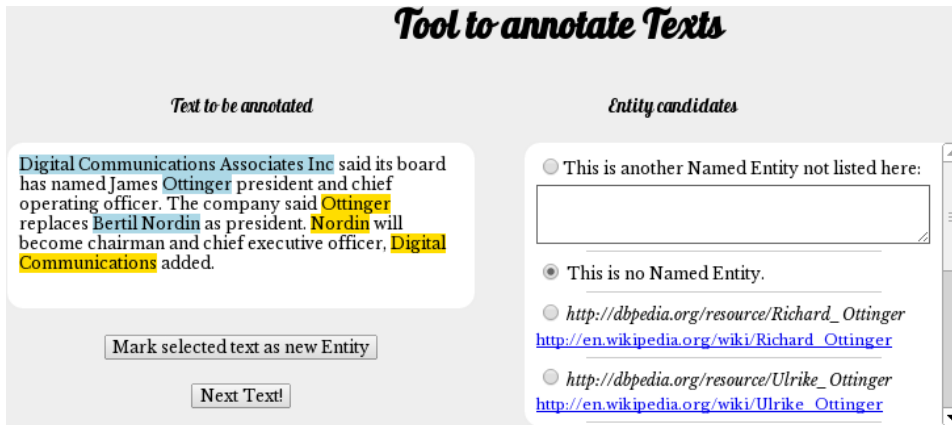
Figure 5: Influence of surface forms.

Figure 6: Quality Rater Tool.