

Appendix - Effect of cumulatively removing properties

Table 1: Effect of cumulatively removing properties, trigram similarity threshold $\sigma = 0.835$, Reuters-21578, $d = 2$, Prefix: dbpedia-owl.

Property	Accuracy
owner	0.675
isPartOf	0.672
countySeat	0.670
city	0.666
owningCompany, party, hometown, religion, memberOfParliament, occupation, influencedBy, locationCity, type, deputy, aircraftHelicopter, spouse, usingCountry, architect, knownFor, place, residence, governor, industry, timeZone, influenced, league, division, deathPlace, team, populationPlace, keyPerson, bandMember, operatedBy, officialLanguage, broadcastNetwork, march, education, season, regionalLanguage, battle, aircraftRecon, stylisticOrigin, militaryUnit, predecessor, wikiPageDisambiguates, leaderFunction, youthWing, derivative, fourthCommander, nationality, leaderParty, languageRegulator, leftTributary, parentCompany, language, country, militaryBranch, mayor, athletics, languageFamily, formerBandMember, largestCity, programmeFormat, state, location, territory, musicSubgenre, affiliation, aircraftTransport, associatedBand, garrison, primeMinister, part, otherParty, restingPlace, locationCountry, currency, aircraftElectronic, foundationPlace, related, neighboringMunicipality, award, relation, chairman, foundedBy, distributingLabel, isPartOfMilitaryConflict, colour, secondCommander, commander, ideology, lieutenant, formerTeam, broadcastArea, product, commandStructure, child, ground, routeStart, headquarter, spokenIn, profession, musicFusionGenre, recordLabel, anthem, successor, notableCommander, regionServed, governingBody, operator, service, hubAirport, militaryRank, aircraftTrainer, president	0.664
parent, jurisdiction, vicePresident, district, genre, pictureFormat, capital, birthPlace, region, governmentType, leaderName, leader, ethnicGroup, targetAirport, campus, monarch, thirdCommander, subsidiary, twinCity, almaMater, instrument, internationalAffiliation	0.661
managerClub, manager	0.663
sisterStation	0.664

Appendix - Algorithm for expansion policy

Algorithm 1: Search candidates for named entities and expansion policy

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

Result: set of candidates C

begin

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

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

label \leftarrow **string**(N_i);

tmp \leftarrow **label**;

expansion \leftarrow **false**;

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

if **key** **contains** **label** **then**

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

tmp \leftarrow **key**;

expansion \leftarrow **true**;

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

then

tmp \leftarrow **key**;

expansion \leftarrow **true**;

label \leftarrow **tmp**;

if \neg **expansion** **then**

heuristicExpansion \leftarrow **label** \cup **heuristicExpansion**

C \leftarrow **C** \cup **searchCandidates**(**label**, σ);

end

Appendix - Figures of experimental results

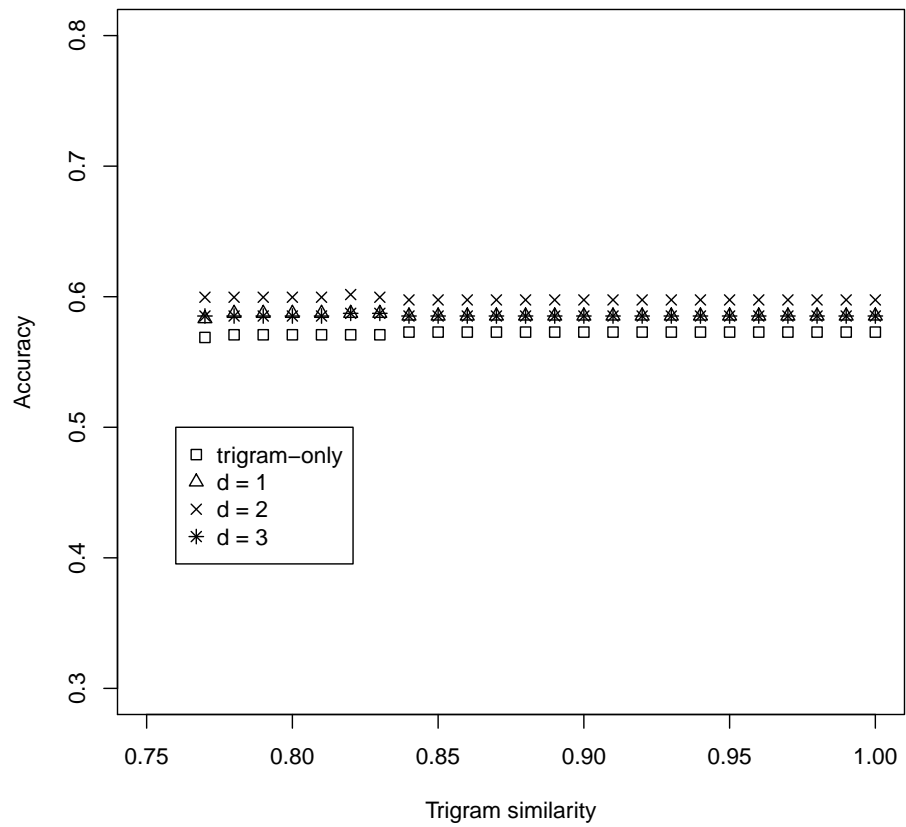


Figure 1: Accuracy of AGDISTIS on the **RSS 500** corpus.

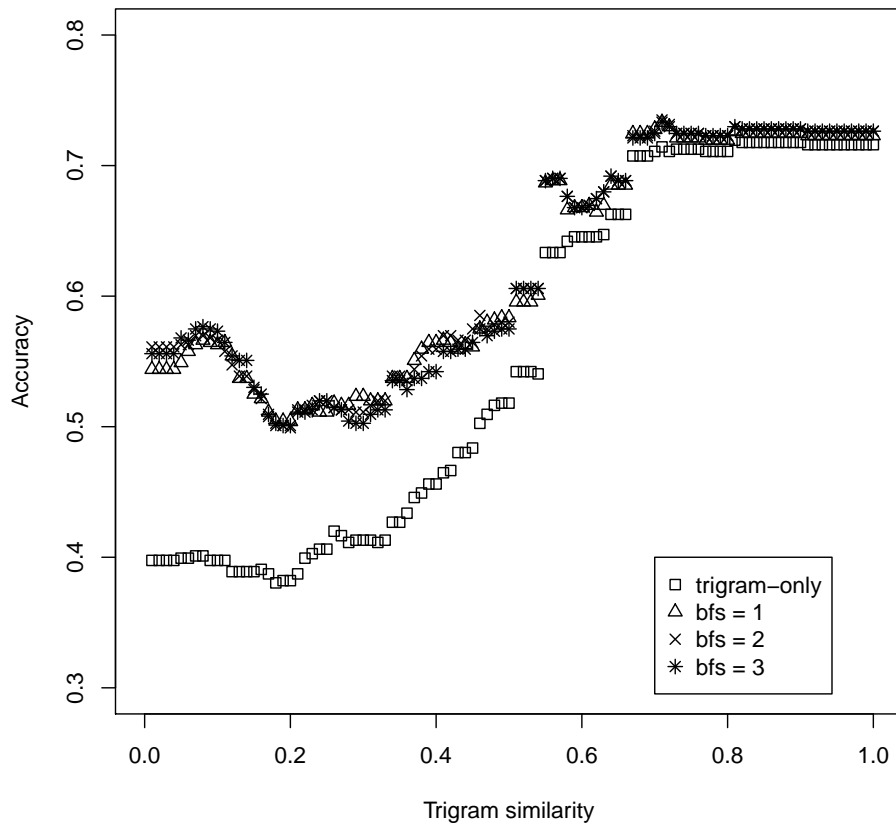


Figure 2: Accuracy of AGDISTIS on the **news.de** corpus.

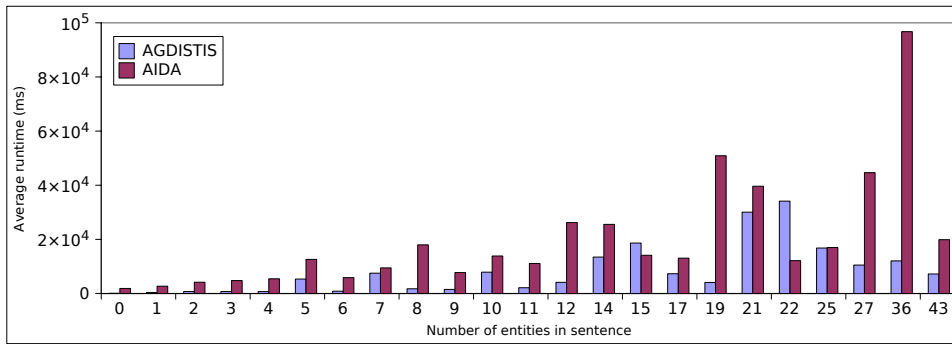


Figure 3: Comparison of the average runtime of AGDISTIS and AIDA on Reuters-21578 corpus with respect to the number of entities per sentence. AGDISTIS is clearly more time-efficient than AIDA.