

Unsupervised POS-Tagging Employing Efficient Graph Clustering

Chris Biemann, University of Leipzig, Coling/ACL SRW 2006

biem@informatik.uni-leipzig.de

Tagset for GERMAN

This tagset was automatically induced by merely presenting large amounts of raw text to the algorithm. Numbers following hashmarks # correspond to tag labels, numbers in brackets give the number of words per tag in the lexicon. Maximally, four randomly chosen entries per tag are given. Special characters are transformed for technical reasons, e.g. `_ESENTS_` is the full stop, `_CLOSEBRAC_` is a closing bracket etc.

#1(34797):Kapitalbeteiligung, Eingriffsmöglichkeiten, Wahlverhalten, Straßenkinder #2(9454):mordeten, einheimsen, forciere, druckten #3(9867):Rockfans, Umweltprojekte, Richterstellen, Veterinärämter
#4(860):Helene, Henk, Hans-Peter, Daniel #5(5324):menschenunwürdig, dämlich, unüberwindbar, gewichtig #6(15154):Ohnmachtsgefühl, Endverbraucher, Theaterereignis, Posten
#7(4477):erfolgreichsten, indisch-pakistanischen, psychotherapeutischen, ständigen #8(2614):Gegenspieler, Schuldezernent, Oberstudiendirektor, Schwiegersohn #9(4576):intensivere, praktikable,
unmittelbare, gefestigte #10(1606):Altstädter, Weiskircher, geladene, syrischen #11(1644):härterer, freischaffender, verärgerter, volatiler #12(1559):Erlensee, Bonn, Schleswig-Holstein, Katar
#13(1246):Dell, Linotype-Hell, Solvay, Allweiler #14(4186):gesessen, hereingenommen, erreicht, entfacht #15(7473):Jang, Köstner, Brandt, Stratmann #16(114):einhundertfünfzig, sechshundert,
Dutzend, fünfundfünfzig #17(168):geschäftlich, betriebswirtschaftlich, therapeutisch, kaufmännisch #18(323):totalitärer, qualitativer, universitärer, gerichtlicher #19(669):Strafverteidiger, Baumeister,
Tischler, TV-Journalist #20(8):Euch, Dich, euch, dich #21(588):DFJW, Sozialgericht, Staatsbibliothek, Fremdenverkehrsverband #22(632):angolanische, türkische, aufgebrauchte, faschistoide
#23(492):abtrünnigen, dahinterliegenden, nächstgrößeren, aufgegebenen #24(649):Village, Writing, Tourism, Girl #25(691):Jod, Stunk, Windkraft, Gänseleber #26(286):VHC, Bundesausschuß, Super-
Jumbo, Bundesrat #27(107):vielbeschworenen, abgesegnete, angepeilte, avisierte #28(129):Trifft, Wären, Solange, Hat #29(259):Tunesiens, Nepals, Afghanistans, Mailands #30(61):rief, denn, wieviel,
wobei #31(98):Dienstagvormittag, Vortag, Samstagmittag, Donnerstagabend #32(39):studierende, wohnenden, deportierten, abgeschobene #33(250):Won, kW, Westmark, kWh #34(19):vielfach,
unverbrüchlich, wochenlang, stundenlang #35(35):dank, oberhalb, mangels, entlang #36(341):moslemischer, zurückgetretener, italienischer, marokkanischer #37(111):reduzierte, drohende, gedrückte,
nachgebende #38(50):Se, hebbt, weer, he, #39(375):Hallenserin, Magdeburgerin, Serbin, Weltranglisten-Fünfte #40(2):gehörende, gehörenden #41(27):dummerweise, bloß, allerorts, erstens
#42(13):herunter, entlang, südwärts, westwärts #43(33):dreißigste, wichtigste, berühmte, schönste #44(326):Nixons, Jansens, Kloses, Havels #45(11):sämtlich, zumeist, ausnahmslos, hauptsächlich
#46(8):oftmals, mitunter, streckenweise, zuweilen #47(23):Nov, Mi, Zi, di #48(2):van, Van #49(2):längerem, langem #50(11):diversen, zahlreichen, manchen, derselbe #51(18):Donau, Weser, Neiße,
Spree #52(16):voreinander, blitzblank, dazwischen, dahinter #53(3):Ach, Oh, Na #54(24):Südzipfel, Foyer, Wohnzimmer, Südwesten #55(13):siebziger, zwanziger, fünfziger, vierziger
#56(71):Ungarisch, Russisch, Hebräisch, Chinesisch #57(14):zahlloser, einschlägiger, zweier, mancher #58(115):Damaszener, Schweizer, Kapstädter, Hanauer #59(19):ungezählte, hunderttausende,
Hunderttausende, unzählige #60(9):unseres, meines, Ihres, deines #61(64):Zaubertrank, Wein, Spoiler, Meerrettich #62(10):XII, VI, IX, VII #63(45):vorhergehenden, vorigen, vierten, elften
#64(10):nordöstlich, südöstlich, südlich, Südlich #65(35):doppelt, vierzehnmal, fünfzehnmal, neunmal #66(19):ungefähr, schätzungsweise, maximal, mindestens #67(4):rechtsaußen, links, rechts,
linksaußen #68(96):rosafarbenen, hellen, feinen, orangefarbenen #69(8):neulich, jüngst, sicherheitshalber, kürzlich #70(21):BR, RB, RIAS, Rias #71(2):derjenigen, derer #72(202):Handwerksverband,
VAMV, Hartmannbund, Blindenverband #73(15):brauner, grauer, gelber, weißer #74(5):gleichsam, gewissermaßen, handelnd, sozusagen #75(2):Ergebnis, Resultat #76(2):Millionen, Milliarden
#77(2):zudem, überdies #78(2):Beispiel, erstermal #79(2):folgt, folgte #80(1):ihm #81(1):andere #82(2):einige, etliche #83(4):Sommer, Herbst, Winter, Frühjahr #84(2):verantwortlich, zuständig
#85(1):immer #86(2):jetzt, nun #87(2):Opfern, Bewohnern #88(1):Als #89(1):am #90(1):Wir #91(2):kosten, geschätzt #92(2):seiner, ihrer #93(3):Angeklagte, Kunde, Pilot #94(5):sie, er, man, wir, ich
#95(3):weiß, wußte, glaubte #96(2):Hände, Arme #97(2):bisher, bislang #98(1):ihr #99(1):auch #100(2):Vielleicht, Wahrscheinlich #101(1):_ESENTF_ #102(1):ließen, zeigten, hielten, setzten
#103(2):Programm, Papier #104(5):hat, hatte, will, macht, nennt #105(3):Umgang, Zusammen, Gemeinsam #106(2):Diskussion, Debatte #107(2):Doch, Denn #108(3):Ansicht, Darstellung,
Einschätzung #109(1):im #110(1):Aber #111(2):handelt, handelte #112(5):ist, war, sei, wäre, bleibe #113(1):einmal #114(2):ebenfalls, gleichfalls #115(3):Für, Über, Durch #116(2):Quartal, Halbjahr
#117(1):dieser #118(1):_ESENTS_ #119(4):In, Bei, An, Entgegen #120(1):erst #121(2):Dr, Mr #122(1):dem #123(2):erhöhte, verbesserte #124(3):Werk, Volk, Ensemble #125(2):kündigte, deutete
#126(1):bereits #127(2):welcher, welchen #128(2):Mitarbeiter, MitarbeiterInnen #129(3):Ausland, Mittelalter, Ruhrgebiet #130(2):Manager, Verteidiger #131(2):Bewohner, BewohnerInnen
#132(2):Regen, Schnee #133(2):solche, derartige #134(12):DM, Dollar, Pfund, Franc #135(1):seine #136(2):laufenden, vorigen #137(2):vorstellen, einig #138(1):Jahr #139(2):Oft, Meist #140(1):Das
#141(2):die, das #142(3):Seine, Deren, Dessen #143(2):Kopf, Rücken #144(3):Film, Text, Apparat #145(2):Viertel, Fünftel #146(2):Wedding, Stade #147(2):dahin, dato #148(2):unsere, unser
#149(2):Mio, Mrd #150(3):Nationen, Metall, Khmer #151(3):Landtag, Stadtteil, Bundesstaat #152(2):Angst, Furcht #153(4):aufeinander, nebeneinander, treu, aneinander #154(1):eine #155(2):Theater,
Schauspiel #156(48):So, Dabei, Damit, Außerdem #157(2):mehr, länger #158(1):an #159(2):Ideen, Visionen #160(2):bestätigte, beweist #161(1):vom #162(2):Präsident, Intendant

see next page

#163(2):Natürlich, Selbstverständlich #164(4):meine, mein, deine, dein #165(2):grundsätzlich, generell #166(2):Ein, Eine #167(2):Computer, Rechner #168(3):spielt, spielte, schuf #169(1):uns
 #170(1):unter #171(2):Komponisten, Prinzen #172(1):_KOM_ #173(4):aber, doch, also, nämlich #174(18):erklärt, berichtet, behauptet, glaubt #175(1):sowie #176(1):neuen #177(3):Zugleich,
 Anschließend, Stattdessen #178(2):liegt, liege #179(2):Zunächst, Zuerst #180(1):wieder #181(2):stieg, sank #182(1):sich #183(3):fällt, läuft, lief #184(1):Wie #185(2):steht, stand #186(1):ab
 #187(1):Jahren #188(2):Verzicht, Zugriff #189(2):Rande, Vorabend #190(6):wird, wurde, werde, würde, erscheint, erschien #191(2):mancher, manches #192(1):Wenn #193(2):bleibt, blieb
 #194(4):Unter, Neben, Hinter, Außer #195(2):Wachstum, Wirtschaftswachstum #196(1):diese #197(2):Pläne, Vorschläge #198(3):dafür, daran, darin #199(2):Mal, Male #200(2):Bürgermeister, Bischof
 #201(3):oben, unten, vorne #202(2):Befragten, befragten #203(1):Es #204(2):Kind, Baby #205(2):niedrigen, steigenden #206(2):Abgeordneten, Delegierten #207(2):per, via #208(2):Verantwortung,
 Qualifikation #209(4):für, über, durch, gegen #210(2):habe, hätte #211(2):scheint, schien #212(3):TV, SV, SC #213(12):Land, Gesetz, Produkt, Gerät #214(2):wies, weist #215(2):Daten, Unterlagen
 #216(2):Gemeinden, Roma #217(1):damit #218(2):Freundin, Kollegin #219(2):gutem, schwarzem #220(2):sitzt, saß #221(2):allgemeinen, schweizerischen #222(2):bedeutet, bedeute
 #223(2):wichtigsten, schönsten #224(5):letzten, vergangenen, nächsten, kommenden #225(1):ihre #226(1):hier #227(2):Menschen, Autos #228(1):Mit #229(3):König, Kaiser, Doktor #230(2):Ordnung,
 Erscheinung #231(2):Karriere, Laufbahn #232(1):Ende #233(2):Wasser, Gras #234(2):zwischen, Zwischen #235(2):Bisher, Bislang #236(2):Projekt, Vorhaben #237(1):alle #238(2):sucht, suchte
 #239(14):gute, riesige, unabhängige, leichte #240(2):geht, ging #241(2):Wirklichkeit, Wahrheit #242(2):Geschäfte, Züge #243(1):Von #244(2):Die, Der #245(2):glaube, denke #246(1):nicht
 #247(9):drei, vier, fünf, sechs #248(1):Sie #249(2):Bach, Mozart #250(1):Auf #251(2):Moment, Augenblick #252(12):Juni, Januar, Mai, März #253(2):Hinblick, Rückblick #254(1):nach #255(2):wirkt,
 wirkte #256(2):bezeichnete, wertete #257(3):miteinander, voneinander, gegeneinander #258(3):Nein, Ja, Gewiß #259(2):Bau, Aufbau #260(2):inzwischen, mittlerweile #261(3):forderte, fordert,
 verlangte #262(3):Mark, Kubikmeter, Einwohnern #263(2):daß, dass #264(2):dennoch, trotzdem #265(2):Ministerpräsidenten, Staatspräsidenten #266(13):Gesellschaft, Verfassung, Rezession, Fabrik
 #267(3):größere, höhere, geringere #268(4):Rund, Etwa, Knapp, Mindestens #269(2):Nachbarn, Konkurrenten #270(2):Zahl, Mehrzahl #271(2):Verluste, Gewinne #272(4):dazu, darauf, darüber, Na
 #273(3):Preise, Zinsen, Mieten #274(1):nichts #275(1):Zeit #276(2):Windows, Schalke #277(1):als #278(5):kann, könnte, sollte, könne, darf #279(2):besteht, bestand #280(3):ebenso, genauso,
 ebensowenig #281(2):diejenigen, denjenigen #282(2):teuren, teure #283(2):endgültig, vollends #284(2):Publikum, Feuer #285(1):seit #286(2):mitten, Mitten #287(1):zwei #288(3):neuer, sogenannter,
 zusätzlicher #289(1):denen #290(2):Plus, Volumen #291(5):jedoch, allerdings, freilich, indes, indessen #292(4):muß, müsse, solle, müßte #293(2):Punkte, Zähler #294(1):allem #295(9):zu, ins, aufs,
 ans, fürs, durchs, gen, unterm, vorm #296(10):York, Angeles, Francisco, Salvador #297(2):Roboter, Taxifahrer #298(2):Gas, Brot #299(19):Behörden, Banken, Eltern, Gewerkschaften #300(1):von
 #301(3):stehen, liegen, sitzen #302(3):möglicherweise, vermutlich, womöglich #303(1):einen #304(6):West, Grüne, Contra, Naturschutz, Handelskammer, Sandoz #305(17):wissen, sagen, zeigen,
 glauben #306(2):oft, häufig #307(1):soll #308(2):dagegen, hingegen #309(2):schön, nett #310(1):ohne #311(3):einzelnen, armen, einschlägigen #312(2):braucht, brauchte #313(2):Unsere, Meine
 #314(2):Gruppe, Sammlung #315(1):selbst #316(2):bezug, Anspielung #317(1):ihren #318(2):höher, niedriger #319(1):vor #320(2):Waffen, Emotionen #321(1):bis #322(1):heute #323(1):so
 #324(2):Gefühl, Gerücht #325(1):Jahre #326(2):Ist, War #327(1):mit #328(2):scheinen, schienen #329(1):auf #330(3):St, bzw, dpa #331(1):Ich #332(6):versucht, versuchen, aufgefordert, davor
 #333(2):solchen, derartigen #334(2):fehlen, fehlten #335(1):ganz #336(2):Auch, Insbesondere #337(2):erneut, abermals #338(1):anderen #339(2):Ebene, Ebenen #340(1):seinen #341(3):Bilder,
 Texte, Träume #342(2):positiv, negativ #343(1):Er #344(2):Genuß, Köpfen #345(2):Telefon, Telephon #346(4):sind, wurden, waren, seien #347(2):künftig, zukünftig #348(1):ihn #349(2):Kein, Keine
 #350(2):wichtig, schwierig #351(1):ein #352(2):Weizsäcker, vornherein #353(5):Inzwischen, Seitdem, Mittlerweile, Seither, Längst #354(2):gerne, gern #355(2):ständig, regelmäßig #356(1):noch
 #357(7):Zwei, Drei, Vier, Fünf, Sechs, Zehn, Acht #358(1):beim #359(1):Und #360(1):diesem #361(2):letztlich, letztendlich #362(1):zur #363(2):derzeit, gegenwärtig #364(3):kommt, kam, komme
 #365(2):Verkauf, Erwerb #366(2):tut, tat #367(3):de, di, del #368(2):Dagegen, Demgegenüber #369(2):Zusammenarbeit, Kooperation #370(2):weltweit, bundesweit #371(1):Nach #372(1):nur
 #373(2):und, oder #374(2):Sache, Höhle #375(3):außen, hinten, innen #376(1):der #377(1):Im #378(2):relativ, vergleichsweise #379(2):falsch, unerträglich #380(5):meist, zumeist, vielfach, meistens,
 oftmals #381(2):Million, Milliarde #382(1):zum #383(2):Trainer, Bundestrainer #384(2):Touristen, Investoren #385(1):dieses #386(2):keineswegs, keinesfalls #387(2):teuer, niedrig #388(2):du, Du
 #389(1):%N% #390(2):Umsatz, Jahresüberschuß #391(2):Konzept, Modell #392(2):lautet, lautete #393(9):New, San, Bad, Los, Buenos, Prenzlauer, North, Las, Santa #394(2):heißt, hieß
 #395(2):üblich, gewohnt #396(122):Geschichte, Entwicklung, Einführung, Studie #397(2):Städte, Löhne #398(1):ersten #399(22):entsprechende, wichtige, bessere, besondere #400(1):um
 #401(2):bekommt, erhält #402(2):deshalb, deswegen #403(2):aus, gegenüber #404(2):Sogar, Einzig #405(2):Dokumente, Gegenstände #406(2):verbundenen, verbundene #407(3):Sein, Mein, Unser
 #408(2):führt, führte #409(2):Tür, Kamera #410(1):schon #411(2):Kroatien, Sparkassen #412(2):warnte, warnt #413(2):kennt, kannte #414(4):usw, etc, Nr, Str #415(5):konnte, wollte, mußte, möchte,
 durfte #416(2):Veränderungen, Änderungen #417(1):es #418(2):schlägt, schlug #419(2):Angaben, Schätzungen #420(2):Mütter, Busse #421(2):hoch, groß #422(4):keine, kein, keinen, keinerlei
 #423(3):begleitet, herausgegeben, angeführt #424(4):Vorstand, Verband, Beirat, Landesverband #425(2):spricht, träumt #426(4):USA, Vereinigten, umliegenden, verschiedensten
 #427(13):Frankreichs, Italiens, Rußlands, Japans #428(2):standen, saßen #429(1):etwas #430(2):Internationalen, Nationalen #431(4):einer, einem, seinem, ihrem #432(2):System, Medium #433(2):in,
 bei #434(2):anstelle, unweit #435(2):Parteien, Gesellschaften #436(2):den, diesen #437(2):gleichzeitig, zugleich #438(2):Viele, Einige #439(3):La, El, Les #440(6):_ESENTE_, _QUOT_,
 OPENBRAC, _CLOSEBRAC_, _COL_, _SEMICOL_