

Unsupervised POS-Tagging Employing Efficient Graph Clustering

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

biem@informatik.uni-leipzig.de

Tagset for FINNISH

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(14849):valmisteluihin, uhkista, nykytasolla, juureksista #2(7512):opiaattien, aikakausien, henkilöstöresurssien, lähtöjen #3(1741):autamme, dialogissa, yhdensuuntaiset, pyrittävä #4(7496):keskustelutalaisuus, sopeuttaminen, työttömyyspäiväraha, rikos #5(2145):ilahduttava, puettu, passiivinen, saavutettava #6(4796):käynnit, näkemyserot, laiturit, tutkimustulokset #7(817):pelitilillä, Suunnitteilla, Työpaja, Esite #8(1335):istutettuja, suoranaisia, suuriakin, menestyneitä #9(1113):viestii, opetetaan, punnitaan, ilmoitellaan #10(841):toteutuneista, mahdollisilla, normaaleista, sähköisestä #11(711):kotimaiselle, moderniin, kielelliseen, asianomaisille #12(92):pientaloalueilla, arkisin, virastoaikana, päiväaikaan #13(500):tavoittaa, toivovat, onnittelee, inhoaa #14(851):pyrittävä, harrasta, johdu, kykenemättömiä #15(337):kaipa, tarkoittaako, joilla, keitetään #16(899):moitteetonta, arvokasta, koneellista, kiinteää #17(2094):perheasian, toimivallan, työntekijäkoulutuksen, henkilöstörakenteen #18(212):k, mm2, y, kW, b, GWh #19(624):tilaamasi, tärkeimmät, tekemät, keskimääräiset #20(474):rajaamaan, ymmärtämään, ennaltaehkäisemään, näkemään #21(439):materiaalisten, kirurgisten, suurimpien, sosiaalisten #22(582):Leppävaaran, Hämeenlinnan, Taivassalon, Länsi-Uudenmaan #23(756):luettelointiin, virheettömyyttä, ylläpitämiselle, virkistyskäytölle #24(99):melko, perustavan, turhan, silmiinpistävän #25(3853):viinejä, kunnostustöitä, tuttavuuksia, kommunikaatiotaitoja #26(2553):sijaintia, sielunhoitoa, tietoisuutta, Seinäjoelle #27(568):palvelisivat, kilpailivat, tulkitsevat, hahmottivat #28(61):ehdotettua, vieläkin, oletettua, keskimääräistä #29(165):rakoilla, tapahtua, karttua, sulivat #30(337):hermostuneita, tasaisemmin, keskeisemmäksi, vapaammin #31(346):jäissä, rikollista, vastenmielistä, siistiä #32(30):maanantaina, maanantai-iltana, tiistaina, viikonvaihteessa #33(81):matkamme, lokakuun, tuhatluvun, 300-luvun #34(367):Valtsu, Ahmat, lukio, Gee, #35(58):romahtanut, pienenee, kasvane, lisäänty #36(557):puoluesihteri, kansleri, kunnaninsinööri, QFD #37(119):muistamme, koetamme, päätinkin, unohdat #38(120):Tehtävä, Fax, Katuosoite, Jakeluosoite #39(33):kerättiin, koulutetaan, tallennettu, kokoamme #40(25):1980-luvulla, 30-luvulla, 1700-luvulla, 60-luvulla #41(41):viidesosalla, murto-osan, kuudennes, neljäsosalla #42(32):Opetussuunnitelman, Kuntoutuksen, Arvioinnin, Tutkinon #43(115):enään, milloinkaan, tulevaisuudessakaan, turvallisuussyistä #44(44):OR, oranssi, hamaa, sirkus #45(309):interaktiivisen, sosiaalisten, yrityskohtaisen, valtakunnallisen #46(30):tapahtumapaikkana, alkuaikoina, keskivaiheilla, loppupuhetillä #47(61):osoitteeseen, korjattavaksi, toimistollemme, postikortilla #48(19):heinäkuuta, toukokuuta, lokakuuta, § #49(3):Syksyn, Kevään, Kesän #50(11):miljoonia, miljardeja, tuhansia, kymmeniä #51(26):kasvamaan, lähiaikoina, heikkenemään, paranemaan #52(5):orkesterin, yhtyeen, kaksikon, trion #53(242):Erik, Joonas, Ernst, Hannele #54(10):näyttävät, tuntuivat, kuulostavat, tuntuvat #55(22):dramaattisesti, väillisesti, miksikään, erilaiseksi #56(9):laskutetaan, veloitetaan, tilitetään, perittäisiin #57(6):Neljä, Kolmella, Kolme, Kuusi #58(20):päättäväisesti, ranteella, etukulmasta, näitisti #59(24):Opiskellaan, Perehdyttää, Perehdytään, Laajennetaan #60(28):kahdella, perinteisellä, kolmellakymmenellä, luotettavalla #61(12):Viranhaltijan, Opettajien, Näyttelijän, Hakemus #62(589):psykologiasta, tekstinkäsittelyyn, polkupyörä-, kilpailu- #63(32):osaisimme, säilyttämme, saatte, ostimme #64(387):katsoja, kansalaiseksi, keksijä, velkoja #65(18):nopeuttavat, korjaavat, värjätään, parantavat #66(4):Paras, Tekokkain, Kaikkein, Oikea #67(40):kahteenkymmeneen, miljoonaan, miljardiin, kuhunkin #68(2909):Korpisaari, Lång, Mäntyranta, Kaunismäki #69(21):Uusi, Seuraava, Tilan, Eri #70(9):tämäkaltainen, tämäkin, tällainen, kyseinen #71(21):Tanskan, Makedonian, Liettuan, Kiinan #72(61):kehottivat, kohtelevat, tervehtivät, säättelevät #73(10):tunnissa, yössä, lukuvuodessa, sekunnissa #74(325):Kuopio, Rautaruukki, Kuortane, Åbo #75(15):Eenok, Lilli, Gideon, pelle #76(280):heitän, syön, valitan, tunnistan #77(39):viidentuhannen, parinkymmenen, viidensadan, kolmenkymmenen #78(198):henkilökunnalta, päättäjiltä, osakkailta, joukkueelta #79(39):Sosiaalilautakunta, Yhteistyötoimikunta, Johtokunta, Tarkastuslautakunta #80(19):aamuvarhaisella, iltapäivällä, käyttövalmis, lauantaiamuna #81(3):pääosa, valtaosa #82(107):kiellettiin, lakkasi, käynnistyvät, keskeytyi #83(180):aidon, voimassaolevan, periaatteellisen, rusteellisen #84(23):kouluvuosien, kertojen, öiden, vuosisatojen #85(10):aikaiset, tarjoamat, toimittamat, tuomat #86(16):keväänä, syksynä, kesänä, talvena #87(2):jollain, jollakin #88(52):AA, 8, HP, alkumatkan, 2-3 #89(8):saattaisimme, lajinsa, alueensa, alansa #90(10):neljäs, kymmenes, yhdeksäs, kahdeskymmenes #91(45):upeasti, lujaa, rehellisesti, väsyneinä #92(303):verottaja, turvallisuusneuvosto, virastotalo, Stakesista #93(2):käytettävät, käytetyt #94(30):parissakymmenessä, erillisessä, viidessätoista, kuudessa #95(11):Yhden, Kolmen, Yhteiskunnan, Toisen #96(203):kehittynyt, pienentyneet, heijastunut, viisinkertaistunut #97(6):pituinen, suuruinen, etäisyydellä, laajuinen #98(5):tarjoavat, luovat, antavat, loivat #99(61):milj., askelman, millisievertin, miljoonaan #100(25):vajaat, pelattava, jokunen, arviolta #101(56):paneutua, tottua, taipua, edisti #102(1):ajan #103(2):niitä, niistä #104(2):tehokasta, miellyttävä #105(2):uusien, hyvien #106(3):aikoina, kädessä, vuosikymmeniä #107(1):yksin #108(1):_ESENTS_ #109(3):tulevat, tulivat, käyvät #110(2):Aikaisemmin, Aiemmin #111(2):nuoret, pojat #112(2):Teemme, Järjestämme #113(2):miehen, tytön #114(3):edelleen, parhaillaan, edelleenkin #115(1):oli #116(3):usein, useimmiten, monesti #117(1):lähes #118(1):tehdä #119(2):vuosia, vuosikymmeniä #120(3):vähintään, enintään, korkeintaan #121(2):haluavat, kannattavat #122(2):lähellä, kaukana #123(3):vuosina, viikolla, kaudella #124(2):suuri, pieni #125(1):eli #126(2):sopii, soveltuu #127(2):sama, samanlainen #128(1):Mutta #129(2):palkkaa, pisteitä #130(2):tilanteen, valinnan #131(4):opettaja, lääkäri, ohjaaja, esimies #132(20):noin, n., Vuoden, Noin, Yli, ISO #133(16):lapset, yritykset, kunnat, naiset #134(4):tehdään, annetaan, asetetaan, täytetään #135(2):johonkin, jollekin #136(2):kunnossa, kohdallaan #137(1):mukana #138(2):siihen, siksi #139(2):Suomi, komissio #140(15):ohjelma, koulu, ryhmä, kirkko #141(1):kaikki #142(3):tulee, tulisi, pitäisi #143(2):asiat, työt #144(2):sai, teki #145(2):pisteen, päivään #146(5):Monet, Useimmat, Useat, Suurimmat, Tärkeimmät #147(3):numerosta, numeroon, Urheilijat #148(2):§, n, momentin #149(1):olevan #150(5):tekevät, tarvitsevat, käyttävät, pitävät, hallitsevat #151(1):eivät #152(3):yliopiston, Energian, Sanomien #153(6):terveysalan, terveysministeriön, terveysministeriö, viestintätekniikan #154(1):tässä #155(3):merkitys, vaikutus, rooli #156(3):suuria, merkittäviä, vakavia #157(2):nykyisin, nykyään #158(1):_KOM_ #159(2):taloutta, talouteen #160(2):kasvavat, muuttuvat #161(1):Ja #162(3):maan, maapallon, ihmiskunnan #163(1):näin #164(2):osallistui, kertyi #165(2):sisältää, sisälsi #166(1):vuonna, klo, vuodesta, vuoteen #168(2):Koulutuksen, Opintojen #169(1):olisi #170(2):toiselle, kahdesti #171(5):perustuva, sisältävä, muodostama, koostuva, käsittelevä #172(1):enemmän #173(2):ja, tai #174(3):tänä, tänäkin, kulvana #175(2):merkitystä, tarvetta #176(2):näyttää, näytti #177(2):Jokaisella, Kaikilla #178(1):tämä #179(2):siirtyi, muutti #180(2):koskevaa, tarvittavaa

see next page

#181(3):mukainen, mukaista, mukaisia #182(2):Hankkeessa, Projektissa #183(1):samalla #184(1):sitten #185(2):On, Näyttää #186(3):meitä, minua, sinua #187(8):kaksi, kolme, neljä, viisi, kuusi, seitsemän, kahdeksan, yhdeksän #188(2):maata, kaupunkia #189(2):minut, sinut #190(3):Kilpailun, Seminaarin, Tapahtuman #191(2):opettajan, ohjaajan #192(2):aivan, ihan #193(1):joko #194(2):tärkein, keskeisin #195(2):ainetta, jätettä #196(2):pituuden, lukumäärän #197(2):voivat, saattavat #198(4):häntä, hänet, heitä, heidät #199(1):_CLOSEBRAC_ #200(1):_OPENBRAC_ #201(21):liittyviä, liittyvää, liittyvistä, koskevia #202(2):hänelle, heille #203(1):siinä #204(2):annettiin, lähetettiin #205(2):Oy_n, Oyj_n #206(2):tulevaa, tulevia #207(2):lähelle, kauas #208(3):näihin, tälle, näille #209(2):tilata, tiedustella #210(2):muuten, muutoin #211(1):ilman #212(2):salasana, salasanan #213(2):fax, faksi #214(3):näma, useimmat, tällaiset #215(1):jälkeen #216(2):oleva, antama #217(1):se #218(3):ymmärtää, uskoa, toivoa #219(15):lasta, paikkaa, miestä, yritystä #220(2):pystyy, kykenee #221(1):viime #222(1):tietoa #223(3):Vuonna, Vuodesta, Hippo-viikkokisat #224(2):Päivän, Illan #225(1):_NUMBER_ #226(2):ohjaa, koordinoi #227(2):riittävästi, tarpeeksi #228(2):muissakin, muillakin #229(2):_ESENTQ_, _ESENTE_ #230(1):muiden #231(3):kuuluu, sisältyy, kuului #232(2):vaikutuksia, edellytyksiä #233(2):Oy, Oyj #234(3):meidän, minun, teidän #235(1):aina #236(2):toteutuksessa, terveydenhuollossa #237(13):Tampereella, Turussa, Oulussa, Kuopiossa #238(3):Tänä, Seuraavana, Samana #239(1):uuden #240(3):ensimmäisessä, ensimmäisellä, ensimmäisestä #241(2):ko_, em_ #242(2):joskus, toisinaan #243(1):koko #244(1):mitään #245(2):II, III #246(2):Niissä, Niillä #247(2):lähtökohta, ydin #248(2):Kunnan, Hallituksen #249(2):toimitetaan, lähetetään #250(2):opiskelijat, oppilaat #251(2):vammaisten, henkisen #252(17):uusi, yhteinen, suomalainen, yleinen #253(1):esim_ #254(2):tiukka, tasaista #255(1):_COL_ #256(7):muita, muista, muihin, muille #257(5):valinnassa, suunniteltaessa, laadittaessa, valmistelussa #258(2):lääninhallituksen, TE-keskuksen #259(2):paljastaa, todistaa #260(2):aineiden, tekijöiden #261(1):Kun #262(2):tuloa, menoa #263(2):suojaa, turvaa #264(1):itse #265(2):yksityisen, yksityisten #266(4):jne, ym, yms, tms #267(2):yleensä, tavallisesti #268(1):lisää #269(4):markan, euron, asukkaan, mk_n #270(1):sekä #271(2):Hyvät, Erilaiset #272(2):mitkä, ketkä #273(2):tarkoitettua, per #274(4):opiskelija, asiakas, oppilas, lukija #275(4):meillä, hänellä, heillä, minulla #276(2):valinta, teko #277(1):- #278(4):puolilla, alojen, puolilta, puolille #279(1):siis #280(2):teknisen, ammatillisen #281(3):sanoo, muistelee, kuvailee #282(1):hän #283(2):voit, voitte #284(1):muun #285(1):osa #286(2):lähinnä, etupäässä #287(2):muutenkin, muutoinkin #288(1):yhdessä #289(1):mukaan #290(2):pidetään, pelataan #291(1):ne #292(2):saa, saisi #293(2):laitos, tiedekunta #294(2):toisessa, kolmannessa #295(2):ihminen, Valtio #296(2):tuntuu, tuntui #297(9):J_, E_, P_, M_, K_, R_, S_, H_, T_ #298(3):saavat, saivat, tekivät #299(1):ennen #300(1):Tämän #301(3):erityisesti, varsinkin, etenkin #302(24):Suomen, Euroopan, Lapin, Keski-Suomen #303(4):Kuntaliiton, Akatemian, Pankin, luonnonsuojeluliiton #304(2):pyritään, pyrittiin #305(2):Nykyisin, Nykyään #306(3):me, mä, te #307(2):kyseessä, kysymyksessä #308(2):nuorten, terveydenhuollon #309(18):alle, syksyllä, keväällä, kesällä #310(5):selvää, muistettava, luonnollista, huomannut #311(2):yhteyksissä, paikkakunnilla #312(1):Tämä #313(3):jne_, yms_, tms_ #314(1):tämän #315(3):Myöskään, Kaikkia, Pelkkä #316(2):ilmenee, selviää #317(2):kaikilla, jokaisessa #318(2):jotain, jotakin #319(2):esim, ks #320(3):tutkimus, historia, kieli #321(1):Lisäksi #322(2):monimuotoisuuden, edistämisen #323(2):näkyä, näkyi #324(1):kertaa #325(1):mm_ #326(2):tuli, saatiin #327(3):varustettu, ehdottomasti, erikoistunut #328(2):esityksen, ehdotuksen #329(1):voi #330(1):jopa #331(2):Tällaiset, Nykyiset #332(1):vuoden #333(1):antaa #334(1):Hän #335(2):silti, onneksi #336(2):suuret, pienet #337(5):Jumalan, Jeesuksen, Kristuksen, Herran #338(3):ilmeisesti, todennäköisesti, luultavasti #339(2):ylläpitää, opastaa #340(2):esille, esiin #341(2):viikon, kuukauden #342(1):nyt #343(4):sellaisia, sellaiset, sellaiseen, sellaisiin #344(1):_QUOT_ #345(3):muuhun, muusta, muulle #346(2):tärkeää, tärkeätä #347(3):kuluttua, väleין, iässä #348(2):pohtii, miettii #349(4):viiden, edellisen, kuluvan, Kuluvan #350(2):lasku, palkka #351(1):jo #352(1):lisäksi #353(13):Helsinki, Turku, Lahti, Oulu #354(29):prosenttia, mennessä, alkaen, lähtien #355(2):silloin, silloinkin #356(1):kuten #357(2):hieman, hiukan #358(2):Suurin, Suuri #359(12):syyskuun, elokuun, toukokuun, kesäkuun #360(1):käyttää #361(1):he #362(6):välinen, välistä, välisiä, väliseen, välisessä, välisestä #363(2):Suomessa, EU_ssa #364(23):asia, joukkue, kirja, malli #365(2):puheenjohtajana, vetäjänä #366(1):Sen #367(1):niin #368(2):Hinta, Kesto #369(4):kesän, syksyn, kevään, lukuvuoden #370(3):verkon, Internetin, internetin #371(2):haluat, käytät #372(2):enää, juurikaan #373(5):sen, niiden, hänen, heidän, muut #374(1):ei #375(2):kaupungin, seudun #376(2):perusterveydenhuollon, köyhyyden #377(1):voidaan #378(3):kaikissa, kaikkiin, jokaiseen #379(2):Kurssin, Projektin #380(2):nähtävissä, luettavissa #381(1):saada #382(1):oman #383(4):Jos, Mikäli, Vaikka, Koska #384(1):vain #385(1):vielä #386(1):paljon #387(2):en, emme #388(1):kuitenkin #389(2):Kuntien, Yliopistojen #390(2):poistaa, korjata #391(1):kuin #392(2):henkilö, järjestö #393(2):sallittua, rajoitettu #394(3):paras, todellinen, ehdoton #395(3):jäävät, jäivät, poikkeavat #396(2):käytössä, käytettävissä #397(3):vuodessa, viikossa, päivässä #398(2):tutkitaan, testataan #399(2):lainkaan, ollenkaan #400(4):samaa, samalle, samoilla, samoihin #401(2):näissä, näillä #402(2):aiheuttama, jälkeinen #403(3):pääasiassa, pääosin, enimmäkseen #404(2):Venäjän, Viron #405(2):tuttu, tuttua #406(2):matkan, illan #407(1):taas #408(3):Uudet, Vanhat, Suuret #409(1):_ #410(2):tietysti, tietenkin #411(3):Suomea, maailmaa, Eurooppaa #412(1):juuri #413(3):Sopimuksen, Esityksen, Päätöksen #414(2):tärkeä, merkittävä #415(2):Kunnanjohtajan, Kj_n #416(3):x, luvun, EUR #417(1):on #418(2):hoidetaan, kirjataan #419(1):Myös #420(25):Hankkeen, Yhdistyksen, Yhtiön, Tällainen #421(2):miettiä, miettimään #422(7):toteaa, kertoi, muistuttaa, arvioi #423(1):olla #424(3):yhteensä, kaikkiaan, arviolta #425(2):tietojen, havaintojen #426(2):antamaan, pitämään #427(2):erikseen, merkittävästi #428(1):vuotta #429(2):harvinaisia, vähäisiä #430(3):seuraavia, seuraavista, seuraaviin #431(1):esimerkiksi #432(15):henkilöä, ihmistä, kappaletta, jäsentä #433(2):monet, useat #434(1):ainakin #435(2):kasvoi, laski #436(3):varmistetaan, uskomme, toivomme #437(2):jossain, jossakin #438(1):huomioon #439(2):puh_, p_ #440(4):hyvin, eniten, selkeästi, tarkoin #441(3):useita, monia, lukuisia #442(2):Osa, Tontin #443(4):liittyvät, koskevat, kohdistuvat, tapahtuvat #444(18):hyvä, laaja, vahva, avoin #445(2):kovin, kovinkaan #446(1):kannalta #447(3):Toinen, Yksi, Eräs #448(3):kahden, kolmen, neljän #449(1):ottaa #450(2):muualta, muualle #451(2):ns_, nk_ #452(2):aseen, luodin #453(3):muutamia, joitakin, joitain #454(2):sitä, sille #455(3):katon, postin, väylän #456(3):myös, nimenomaan, myöskin #457(1):pitää #458(2):todeta, olettaa #459(2):jättää, laittaa #460(2):ks_, vrt_ #461(4):valitse, kirjoita, vieläpä, pyydä #462(5):ovat, olivat, sijaitsevat, perustuvat #463(3):suurempi, pienempi, korkeampi #464(3):Niiden, Ohjelman, Palvelun