Supplement to Lauri Lahti’s conference article “Educational exploration along the shortest paths in conceptual networks based on co-occurrence, language ability levels and frequency ranking”
Lauri Lahti, 20150922
Department of Computer Science, Aalto University School of Science, Finland.

Conference article:

This supplement is available online in Aaltodoc publication archive at http://aaltodoc.aalto.fi.

3018 highest-ranking nouns of British National Corpus available from lemmatized word lists of an online database (Leech et al. 2001), shown in decreasing order of the number of occurrences per million words (indicated in the parenthesis):
time (1833); year (1639); people (1256); way (1108); man (1003); day (940); thing (776); child (710); nr (673); government (670); work (653); life (645); woman (631); system (610); case (613); part (612); group (607); number (608); world (600); house (598); area (585); company (579); problem (565); service (549); place (534); hand (532); party (529); school (529); country (486); point (484); week (476); member (471); end (458); state (440); word (438); family (428); fact (426); head (402); month (398); side (398); business (394); night (393); eye (392); home (390); question (390); information (387); power (385); change (384); per cent (384); interest (376); development (375); money (375); book (374); water (372); other (367); form (365); room (364); level (360); car (353); council (348); policy (348); market (346); court (344); effect (338); result (334); idea (328); use (328); study (327); job (326); name (326); body (325); report (325); line (323); law (318); face (315); friend (315); authority (313); road (313); minister (306); rate (303); door (302); hour (302); office (300); right (299); war (297); mother (295); person (290); reason (289); view (289); term (288); period (283); centre (282); figure (282); society (282); police (278); city (275); need (273); community (272); million (272); kid (271); price (271); control (270); action (269); cost (269); issue (269); process (269); position (268); course (267); minute (266); education (260); type (259); research (258); subject (256); programme (255); girl (254); moment (254); age (252); father (252); force (250); order (250); value (250); act (248); matter (248); health (246); lot (246); decision (243); street (243); industry (242); patient (242); class (241); mind (241); church (238); condition (237); paper (237); bank (234); century (233); section (232); activity (231); hundred (231); table (231); death (230); building (229); sense (229); sort (229); staff (227); team (226); experience (223); student (222); language (221); mrs (221); town (221); plan (220); department (219); management (219); morning (219); committee (217); product (217); practice (216); evidence (215); ground (215); letter (215); meeting (215); foot (214); boy (213); back (212); game (212); food (211); union (211); role (210); event (208); land (208); art (206); support (204); range (203); stage (203); teacher (203); trade (203); voice (203); arm (202); club (202); field (201); history (201); parent (201); account (200); material (199); care (198); situation (198); manager (197); project (197); record (197); example (196); training (194); window (194); sale (192); air (191); difference (191); light (191); university (191); wife (190); relationship (189); sir (189); quality (188); rule (187); pound (184); story (184); tax (184); worker (184); data/datum* (183); model (183); nature (182); officer (181); structure (181); bed (180); hospital (180); method (180); unit (180); movement (179); detail (178); date (177); wall (175); computer (174); amount (171); approach (171); 201 (171); authority (170); president (170); scheme (170); chapter (169); theory (168); property (167); son (166); director (165); leadership (165); application (165); application (165); computer (164); board (162); king (162); production (162); secretary (162); finance (161); operation (161); opportunity (161); share (161); agreement (160); lord (160); contract (159); picture (159); test (159); security (159); charge (149); heart (149); basis (148); design (148); p (148); piece (148); population (147); tree (147); knowledge (146); performance (146); plant (146); pressure (146); fire (145); news (145); environment (144); garden (144); size (144); analysis (143); rest (143); success (143); thought (142); region (141); attention (140); list (140); relation (140); set (140); space (140); analysis (143); rest (143); success (143); thought (142); region (141); attention (140); list (140); relation (140); set (140); space (140); analysis (143); rest (143); success (143); thought (142); region (141); attention (140); list (140); relation (140); set (140); space (140); analysis (143); rest (143); success (143); thought (142); region (141); attention (140); list (140); relation (140); set (140); space (140); analysis (143); rest (143); success (143); thought (142); region (141); attention (140); list (140); relation (140); set (140); space (140); analysis (143); rest (143); success (143); thought (142); region (141); attention (140); list (140); relation (140); set (140); space (140);
(11); broken (11); burial (11); butterfly (11); calf (11); cassette (11); cast (11); catch (11); chorus (11); clarity (11); close (11); copyright (11); courtsey (11); cult (11); cure (11); daylight (11); deadline (11); decree (11); deficiency (11); descent (11); dominance (11); donor (11); dressing (11); drill (11); electoract (11); ending (11); englishman (11); express (11); fever (11); fisherman (11); fit (11); flash (11); flock (11); flour (11); fog (11); folly (11); formulation (11); fringe (11); glimpse (11); greenhouse (11); harvest (11); hatred (11); honey (11); hostage (11); hunger (11); hunt (11); ignorance (11); immigration (11); impulse (11); inability (11); insider (11); investigator (11); japense (11); jungle (11); keyboard (11); landowner (11); liaison (11); liberal* (11); listener (11); lobby (11); locality (11); lodge (11); luxury (11); madame (11); mainframe (11); mainland (11); majesty (11); memorandum (11); methodology (11); ml (11); monitor (11); monkey (11); mucosa (11); newcomer (11); notebook (11); nuisance (11); opposite (11); outfit (11); pad (11); passport (11); patrol (11); pepper (11); peryter (11); photography (11); pie (11); pillow (11); placement (11); pony (11); preservation (11); privacy (11); proced (11); programming (11); projection (11); prosperity (11); pudding (11); punch (11); pylorus (11); query (11); quotation (11); racism (11); rebellion (11); recipient (11); refuge (11); renaissance (11); renewal (11); repayment (11); repetition (11); resentment (11); reservoir (11); retreat (11); rev (11); revenge (11); rifle (11); romance (11); ruin (11); sacrifice (11); salvation (11); scent (11); scholarship (11); sediment (11); sentiment (11); solidarity (11); stadium (11); staircase (11); stool (11); subscription (11); suburb (11); superintendent (11); supervisor (11); survivor (11); triangle (11); unionist (11); variant (11); wardrobe (11); warrant (11); warrior (11); welcome (11); adjective (10); agony (10); aluminium (10); ant (10); asylum (10); balloon (10); bass (10); biography (10); blast (10); boost (10); bowler (10); bracket (10); brewery (10); bulb (10); cargo (10); chord (10); colitis (10); commentary (10); contemporary (10); coral (10); dairy (10); default (10); dirt (10); ed (10); ego (10); elder (10); enjoyment (10); enthuisiast (10); extreme (10); fax (10); fine (10); fork (10); gail (10); gowm (10); guild (10); heal (10); hemisphere (10); herd (10); hire (10); holy (10); housewife (10); infrastructure (10); injunction (10); irony (10); jam (10); jar (10); jazz (10); knitting (10); lb (10); lieutenant (10); listing (10); magnitude (10); monarch (10); monarchy (10); morale (10); mould (10); mug (10); nationalism (10); nationalist (10); nationality (10); north-west (10); para (10); pillar (10); poison (10); prescription (10); primary (10); rainbow (10); ram (10); reasoning (10); rehearsal (10); reluctance (10); residue (10); ribbon (10); scrap (10); semi-final (10); shilling (10); soviet (10); spelling (10); spider (10); t-shirt (10); toast (10); vegetation (10); velocity (10); voltage (10); waiter (10); waiting (10); wartime (10); wheat (10)

For each of 3018 highest-ranking nouns of British National Corpus available from lemmatized word lists of an online database (Leech et al. 2001) we queried an online database for Google Web 1T 5-gram database (FAU Erlangen-Nürnberg 2015) to identify all other nouns belonging to the same set of 3018 nouns that co-occur at the distance of at most four words left or right, with the association measure of t-score and considering at most 50 highest-ranking nouns having a frequency of at least 40. When identifying nouns we relied on just matching spelling and thus some non-nouns may have possibly become unintentionally considered as nouns and some nouns may have become unintentionally excluded. For all 3018 unique nouns we gained together a set of 54 610 unique pairs of nouns (i.e. each pair corresponding to a co-occurrence of two nouns, both belonging to the set of 3018 nouns). There were 2994 unique nouns in 54 610 unique pairs of nouns.

24 unique nouns that belonged to the set of 3018 unique nouns but did not belong to the set of 2994 unique nouns:

bit*; co-operation; conservative*; data/*datum*; decision-making; elite*; fig*; gen.; good*; labour*; lead*; liberal*; lot*; make-up; miss*; north-east; north-west; per cent; pole/pole; semi-final; south-east; t-shirt; va*; va/*; vice-president

All 54 610 unique pairs of nouns so that each pair corresponds to a co-occurrence of two nouns, both belonging to the set of 3018 unique nouns:

abby=bath; abbey=church; abbey=court; abbey=hotel; abbey=house; abbey=manor; abbey=museum; abbey=park; abbey=plc; abbey=road; abbey=saint; abbey=st; abbey=street; abbey=theatre; abbey=wood; ability=handle; ability=increase; ability=limit; ability=offer; ability=pay; ability=williness; ability=work; abolition=act; abolition=agriculture; abolition=capital; aboltion=child; abolition=clause; abolition=conversion; abolition=death; abolition=duty; abolition=man; abolition=monarchy; abolition=movement; abolition=penalty; abolition=prison; abolition=property; abolition=punishment; abolition=resolution; abolition=slave; abolition=tax; abolition=trade; abolition=war; abortion=adoption; abortion=ban; abortion=birth; abortion=bush; abortion=choice; abortion=court; abortion=debate; abortion=decision; abortion=desire; abortion=disability; abortion=disease; absence=employee; absence=exidence; absence=illness; absence=judgment; absence=leave; absence=pay; absence=period; absence=presence; absence=president; absence=proof; absence=reason; absence=sickness; abuse=alcohol; abuse=child; abuse=childhood; abuse=discretion; abuse=drug; abuse=editor; abuse=elder; abuse=fraud; abuse=health; abuse=insitute; abuse=nursing; abuse=potential; abuse=prevention; abuse=prisoner; abuse=rape; abuse=report; abuse=reporting; abuse=show; abuse=subject; abuse=treatment; abuse=violence; ac=battery; ac=cable; ac=heating; ac=input; ac=notebook; ac=outlet; ac=power; ac=supply; ac=voltagae; academ=air; academy=army; academy=art; academy=award; academy=charter; academy=christian; academy=education; academy=engineering; academy=euopean; academy=face; academy=force; academy=insitute; academy=knighet; academy=leadership; academy=marketing; academy=medicine; academy=music; academy=police; academy=press; academy=russian; academy=school; academy=science; academy=sport; academy=training; academy=winner; accent=american; accent=chair; accent=desk; accent=diamond; accent=english; accent=french; accent=furniture; accent=gold; accent=grave; accent=high; accent=lighting; accent=pillow; accent=plate; accent=recovey; accent=round; accent=rug; accent=table; acceptance=agreement; acceptance=approval; acceptance=gain; acceptance=inspection; acceptance=letter; acceptance=modification; acceptance=privacy; acceptance=reject; acceptance=site; acceptance=speech; acceptance=testing; acceptance=user; access=control; access=database; access=ferry; access=gain; access=information; access=help; access=ms; access=network; access=permission; access=point; access=public; access=spread; access=bus; accident=car; accident=claim; accident=compensation; accident=death; accident=emergency; accident=illness; accident=incident; accident=injury; accident
None
References: