Self-Tuned Descriptive Document Clustering using a Predictive Network

Supplemental Material

Austin J. Brockmeier, *Member, IEEE*, Tingting Mu, *Member, IEEE*,
Sophia Ananiadou, and John Y. Goulermas, *Senior Member, IEEE*

This supplement contains additional results on the datasets used in the main paper (Appendix 1) and results for other datasets not included in the main manuscript in Appendix 2.

Appendix 1 Additional Results

We detail the descriptive clustering solutions and running time for the proposed predictive-descriptive clustering (PDC) algorithm across all datasets. The running times are given in Table 1 and the features selected for each cluster are given in the tables on the subsequent pages, followed by the contingency tables showing the correspondence between the clusters and the ground truth topics.

The running time includes generating candidate clusterings, cluster size selection, generating feature subsets for each cluster of the selected clustering, and selecting the best feature subset. The clusterings ranging from 2 to 26 clusters are generated using Ng, Jordan, and Weiss's spectral clustering algorithm [1] on an implicit similarity matrix formed via cosine similarity between normalized term-frequency inverse document frequency weighted vectors [2]. The algorithm firstly uses a truncated eigendecomposition of the normalized similarity matrix. As the eigendecomposition only requires matrix-vector multiplication, it can be performed as a series of multiplications between sparse or diagonal matrices and a vector—avoiding the explicit formation of the dense $n \times n$ similarity matrix. After the eigendecomposition, the first C eigenvectors are used as input (after normalizing the vector for each instance) to the standard k-means algorithm with $k = C$. This is repeated for $C \in \{2, 3, \ldots, 26\}$.

For cluster size selection, we use Akaike information criterion (AIC) to select the number of clusters based on a multinomial model of features that occur in at least 5 instances. The multinomial distribution is estimated with additive smoothing by adding 1 to each count.

For feature selection, we use the CMIM+ (conditional mutual information maximization) algorithm [3] with additional positivity constraints on binarized feature vectors (that is, feature occurrence is used instead of feature count). The number of features per cluster is chosen by the Bayesian information criterion (BIC) where the likelihood is based on a ℓ_2-regularized logistic regression model, which is fit in LIBLINEAR [4] with input arguments `-s 0 -c 10 -B 1` (primal optimization, $\tau = \frac{1}{10}$, and included bias).

References

[1] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Adv. Neural Inf. Process. Syst.*, vol. 14, 2002, pp. 849–856.

[2] G. Salton, A. Wong, and C.-S. Yang, "A vector space model for automatic indexing," *Comm. ACM*, vol. 18, no. 11, pp. 613–620, 1975.

[3] F. Fleuret, "Fast binary feature selection with conditional mutual information," *J. Mach. Learn. Res.*, vol. 5, no. 11, pp. 1531–1555, 2004.

[4] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "LIBLINEAR: A library for large linear classification," *J. Mach. Learn. Res.*, vol. 9, no. 8, pp. 1871–1874, 2008.

[5] M. Lichman, "UCI machine learning repository," 2013. [Online]. Available: http://archive.ics.uci.edu/ml

[6] USDA NRCS, "The plants database," 2008, 31 December 2008. [Online]. Available: http://plants.usda.gov/

[7] G. R. Kiss, C. Armstrong, R. Milroy, and J. Piper, "An associative thesaurus of English and its computer analysis," in *The Computer and Literary Studies*, A. J. Aitken, R. W. Bailey, and N. Hamilton-Smith, Eds. Edinburgh: University Press, 1973, pp. 153–165.

[8] T. A. Davis and Y. Hu, "The University of Florida sparse matrix collection," *ACM Transactions on Mathematical Software (TOMS)*, vol. 38, no. 1, p. 1, 2011.

TABLE 1
Dataset profiles, selected cluster size, and total running time as logged in MATLAB on Mac OS X with a 2.8 GHz Intel Core i7 and 16 GB RAM.

	Classes	Instances n	Features N	Clusters C	Time (s)
Reviews (+/-):					
movie kitchen	2	4000	25543	2	4.22 s
movie books	2	4000	27436	2	4.70 s
movie electronics	2	4000	25778	2	3.23 s
kitchen electronics	2	4000	7786	4	3.24 s
dvd electronics	2	4000	13048	4	3.23 s
movie books electronics kitchen	4	8000	30079	4	5.49 s
dvd movie electronics kitchen	4	8000	29552	3	5.95 s
dvd books movie kitchen	4	8000	31101	3	7.32 s
dvd books electronics movie	4	8000	31299	3	6.83 s
dvd books electronics movie kitchen	5	10000	32375	4	8.25 s
20-News (subsets):					
med guns	2	1896	61188	4	2.37 s
autos space hardware	3	2951	61188	4	3.53 s
comp	5	4852	61188	10	7.96 s
rec	4	3968	61188	7	6.10 s
sci	4	3945	61188	5	5.34 s
mix2	5	4913	61188	6	6.64 s
mix	5	4677	61188	5	5.99 s
sci comp	9	8797	61188	10	15.20 s
sci comp rec	13	12765	61188	11	20.81 s
20-News (all)	20	18820	41747	12	31.36 s
Reuters	8	7674	10250	10	9.65 s
Recipe	20	39774	10506	21	118.53 s
NSF	35	125730	119773	23	387.60 s
AG's news	4	127600	39763	26	268.67 s

TABLE 2
Reviews: movie kitchen. Descriptive features selected for each cluster sorted by coefficient value.

1: film, movie, story, films, life, characters, movies, director, character, scene, man, comedy, plot, young, action, world, scenes, funny, role, john, plays, script, audience, screen, played, acting, actors, end, people
2: bought, product, buy, clean, kitchen, pan, plastic, purchased, coffee, amazon, item, size, unit, price, dishwasher, ordered, sheets, kitchenaid, oven, sturdy, knives, lid, blender, pot, heat, cups, stainless, purchasing, purchase, cooking, toaster, dishes, rust, cuisinart, oxo, shipping, counter, cook, durable, warranty, grill, products, baking, cookware, pots, heating, calphalon

TABLE 3
Reviews: movie books. Descriptive features selected for each cluster sorted by coefficient value.

1: film, movie, films, director, comedy, scene, movies, script, cast, screen, watch, audience, scenes, plays, man, black, performance, actors, role, action, young, wife, character, played, performances, acting, john, moments, hollywood, actor, minutes, comic, head, big, star, funny, plot, watching, bad
2: book, read, books, reading, author, reader, quot, authors, recipes, pages, volume, page, readers

TABLE 4
Reviews: movie electronics. Descriptive features selected for each cluster sorted by coefficient value.

1: film, movie, story, films, man, character, director, characters, scene, comedy, plot, action, life, young, role, scenes, john, movies, cast, funny, script, world, audience, plays, acting, girl, sex, played, actors, hollywood, interesting, people, watch
2: bought, product, quality, buy, unit, price, ipod, purchased, headphones, battery, card, amazon, software, sony, purchase, button, cord, pc, volume, batteries, printer, cable, customer, cd, wireless, keyboard, item, laptop, products, speakers, usb, mp, plug, headset, router, install, speaker, hp, signal, gb, panasonic, cables, plugged, logitech, xp, adapter, defective, sandisk, warranty, charger

TABLE 5
Reviews: kitchen electronics. Descriptive features selected for each cluster sorted by coefficient value.

1: pan, knife, sheets, kitchen, oven, dishes, cooking, set, dishwasher, cutting, knives, cake, handle, thin, grill, rust, clean, sturdy, handles, pans, bowls, wash, food, beautiful, family, plates, love, pieces, baking, pizza, stick, cook, toaster, skillet, surface, cookware, heat, bread, evenly, chicken, steel, meat, heavy, cleans, utensils, vegetables, calphalon, bake, cooks
2: water, coffee, maker, ice, cup, machine, tea, filter, blender, grinder, grind, cooker, motor, broke, drink, liquid, plastic, fill, years, drinks, pour, milk, time, cuisinart, lid, hot, pot, cups
3: card, camera, printer, work, windows, hp, drive, amazon, memory, item, canon, data, garmin, information, failed, days, gb, product, software, computer, cards, service, network, router, photos, worked, support, link, sandisk, customer, firmware, gps, pictures, problems, connection, wireless, site, laptop, refund, usb, pc, device, linksys, company, working, drivers, mb, install, download, ethernet
4: sound, ipod, phone, headphones, ear, dvd, quality, speakers, logitech, reception, speaker, radio, cable, noise, headset, bass, hear, remote, audio, screen, low, cables, price, ears, batteries, stations, monster, volume, picture, music, cord, system, phones, player, fm, battery, listen, static, component, interference, microphone, station, power, tv, headphone, earphones, stereo, panasonic, listening, sony

TABLE 6
Reviews: dvd electronics. Descriptive features selected for each cluster sorted by coefficient value.

1: movie, film, story, acting, man, characters, people, guy, script, comedy, men, plot, role, character, films, funny, young, movies, life, ending, girl, book, women, real, watch, cast, scene, interesting, sex, director, scenes, actor, actors, boring, predictable
2: dvd, season, video, collection, show, videos, episodes, music, elmo, workout, seasons, series, disc, dvds, song
3: ipod, price, sound, quality, headphones, speakers, printer, cable, remote, ink, noise, ears, bass, ear, monster, player, radio, product, difference, cd, stations, cables, cds, nano, bought, portable, expensive, cheaper, fm, mp, earphones, cost, sony, speaker, stereo, bose, listen
4: phone, card, mouse, keyboard, wireless, memory, router, worked, sandisk, work, data, customer, unit, support, laptop, works, software, batteries, item, computer, gps, windows, linksys, hardware, garmin, problem, pda, device, usb, buttons, months, products, computers, pc, netgear, cordless, working, battery, install, download, contacted, month, problems, product, bought

TABLE 7
Reviews: movie books electronics kitchen. Descriptive features selected for each cluster sorted by coefficient value.

1: film, movie, films, director, scene, comedy, movies, man, script, character, action, john, role, young, scenes, audience, plays, cast, funny, actors, acting, watch, played, hollywood, actor, plot, characters, story
2: book, read, books, reading, author, history, writing, authors, pages, text, information, readers, page, reader, quot, language, analysis, chapter, novels, library, bias, enjoyed, essay, chapters, illustrations
3: ipod, card, sound, phone, battery, headphones, product, printer, cable, worked, dvd, software, laptop, usb, gps, pc, sony, keyboard, wireless, mp, batteries, amazon, speakers, hp, sandisk, signal, router, logitech, unit, bought, price, fm, cables, adapter, gb, cord, quality, headset, button, speaker, customer, refund, plugged, xp, plug, charger, buy, purchased, purchase, warranty
4: clean, water, pan, kitchen, coffee, dishwasher, bowl, knife, sheets, handle, maker, pot, dishes, blender, heat, oven, toaster, stainless, knives, plastic, counter, cup, sturdy, lid, size, kitchenaid, cuisinart, pans, cups, cook, oxo, cooker, metal, cooking, mixer, grill, cleaning, cookware, bread, wash, container, pour, rust, baking, calphalon, evenly, cooks, pots, heating, bake

TABLE 8
Reviews: dvd movie electronics kitchen. Descriptive features selected for each cluster sorted by coefficient value.

1: movie, film, story, characters, films, funny, watch, character, comedy, man, director, young, scene, plot, show, scenes, acting, movies, world, life, guy, boring, role, series, shes, script, humor, people, action, interesting, men, girl, cast, woman, fun, played, plays, actors, audience, hollywood, actor, sex, watching, end
2: sound, ipod, card, software, battery, product, phone, headphones, printer, cable, worked, cd, laptop, usb, batteries, keyboard, pc, sony, amazon, gps, wireless, bought, mp, hp, quality, speakers, gb, router, signal, buy, sandisk, unit, logitech, price, cables, adapter, cord, fm, headset, volume, customer, speaker, xp, refund, purchased, plugged, purchase, plug, charger, item
3: water, pan, coffee, kitchen, clean, sheets, dishwasher, bowl, handle, knife, dishes, maker, pot, oven, blender, kitchenaid, easy, toaster, plastic, cook, stainless, heat, food, cooking, knives, size, lid, sturdy, cup, oxo, counter, metal, cuisinart, pans, bread, rust, cooker, cups, baking, cleaning, grill, cookware, pour, container, calphalon, evenly, pots, cooks, heating, bake

TABLE 9
Reviews: dvd books movie kitchen. Descriptive features selected for each cluster sorted by coefficient value.

1: book, read, books, reading, author, reader, authors, page, pages, text, writing, information, research, readers, quot, israel, analysis, chapter, published, bias, library, data, novels, illustrations, chapters
2: film, movie, films, director, movies, comedy, watch, dvd, scene, music, script, show, video, scenes, plays, acting, young, funny, cast, man, actors, performance, played, watched, camera, action, performances, role, guy, hollywood, screen, plot, hilarious, tv, special, star, character, actor, watching, play, story, characters
3: product, clean, kitchen, plastic, pan, water, item, handle, coffee, size, sheets, unit, knives, bowl, dishwasher, bought, purchased, maker, sturdy, toaster, kitchenaid, pot, oven, blender, dishes, price, amazon, ordered, lid, rust, cups, stainless, counter, pans, heat, replacement, cleaning, cuisinart, shipping, products, replace, cooking, purchase, cook, warranty, cookware, buy, baking, heating

TABLE 10
Reviews: dvd books electronics movie. Descriptive features selected for each cluster sorted by coefficient value.

1: book, read, books, author, reading, recipes, authors, writing, research, pages, text, readers, reader, practical, page, israel, analysis, quot, concepts, library, bias, novels, chapter, resource, illustrations, chapters
2: film, movie, films, director, comedy, cast, scene, watch, movies, script, acting, scenes, show, young, performances, plays, watched, man, funny, audience, action, played, actors, guy, role, hollywood, girl, hilarious, plot, actor, character, star, love, watching, characters, story
3: product, quality, unit, ipod, bought, price, card, sound, headphones, printer, battery, phone, cable, sony, worked, purchased, button, item, batteries, keyboard, laptop, software, purchase, buy, pc, wireless, usb, cord, customer, amazon, signal, brand, plug, mp, logitech, router, speakers, gb, install, sandisk, fm, cd, plugged, headset, adapter, panasonic, warranty, speaker, xp, charger

TABLE 11
Reviews: dvd books electronics movie kitchen. Descriptive features selected for each cluster sorted by coefficient value.

1: book, read, books, reading, author, history, writing, authors, pages, page, readers, research, literature, text, reader, analysis, israel, quot, chapter, novels, published, historical, library, bias, illustrations, chapters
2: film, movie, films, director, comedy, scene, watch, movies, acting, script, cast, show, scenes, performances, young, watched, funny, plays, man, audience, played, role, actors, action, guy, girl, hollywood, actor, woman, guys, humor, plot, hilarious, moments, watching, character, effects, end, fun, series, star, story, dialogue, people, characters
3: sound, ipod, card, battery, phone, product, cable, headphones, printer, software, worked, laptop, sony, usb, amazon, quality, gps, batteries, wireless, pc, keyboard, speakers, bought, mp, router, hp, signal, price, unit, cd, buttons, sandisk, cord, buy, gb, adapter, fm, logitech, headset, speaker, purchased, customer, plugged, xp, plug, purchase, refund, item, charger, warranty
4: water, pan, kitchen, coffee, clean, handle, knife, bowl, maker, dishwasher, pot, kitchenaid, dishes, sheets, blender, oven, stainless, toaster, plastic, heat, easy, knives, sturdy, size, lid, cup, cuisinart, counter, metal, pans, oxo, cook, cooker, cups, cleaning, grill, cookware, food, cooking, wash, rust, container, pour, baking, bread, calphalon, evenly, cooks, heating, bake

TABLE 12
20-News: med guns. Descriptive features selected for each cluster sorted by coefficient value.

1: msg, food, dyer, photography, stuff, uga, sauce, foods, amateur, leung, kirlian, urinals, allergic, taste, unlv, corona, linus, sol, aura, plates, processing
2: banks, medicine, treatment, thanks, disease, health, doctor, patients, information, infection, physician, pain, dr, surgery, patient, weight, russell, medical, symptoms, sci, condition, chronic, cancer, medication, diagnosed, treatments, hospital, doctors, diagnosis
3: gun, guns, killed, weapons, police, government, crime, waco, senate, ufl, federal, firearms, military, rights, veal, nra, state, laws, firearm, against, law, citizens, amendment
4: cdt, fbi, brent, fire, gas, handheld, koresh, survivors, chopin, inside, bastard, bd, atf

TABLE 13
20-News: autos space hardware. Descriptive features selected for each cluster sorted by coefficient value.

1: thanks, card, pc, please, dos, monitor, bios, modem, port, windows, dx, disk, gateway, cpu, drive, motherboard, computer, board, scsi, chip, mb, hi, machine, cards, video, drives, pin, cache, ide, floppy, controller, mhz, bus, meg, ram, vlb, isa, os
2: car, cars, engine, ford, oil, insurance, honda, magnus, auto, bmw, eliot, police, wheel, rear, nissan, front, tires, driving, toyota, vw, transmission, mustang, wagon, clutch, autos, gear, drew
3: jbh, mccall, nick, sky, calpoly, bby, wats, rights, pgf, world, theory, war, steinn, energy, known, human, pat, rays, oort, idea, miner, universe, nature, we, writes, they, given, things, bursts, see, earth, space
4: nasa, space, henry, shuttle, spacecraft, science, orbit, jupiter, solar, nsmca, satellite, astronomy, propulsion, project, zoo, pat, russian, mission, earth, moon, launch, sun, flight, military, launched, development, lunar, funding, high, mars, missions, orbital, satellites

TABLE 14
20-News: comp. Descriptive features selected for each cluster sorted by coefficient value.

1: ac, navy, group, uk, graphics, comp, xv, subscribe, proposed, chris
2: windows, andre, zip, cica, file, files, biochemistry, ftp, directory, alec, bmp, shareware, dos
3: windows, dos, win, it, mouse, crash, ms, installed, when, writes, manager, apps, microsoft, ndw, os, load, emm, autoexec, ini, my, run, exe, rmohns, problem, disk, loaded, sys
4: thanks, please, firstclass, mail, email, where, information, looking, utas, advance, hpgl, dxf, anybody, hi, pcx, gif, skiing, rpicas
5: card, bus, diamond, dx, video, eisa, vesa, ati, drivers, mode, driver, cards, vlb, isa, board, svga, mb, local, bios
6: image, points, objects, sphere, science, images, sun, source, motif, we, public, package, simple, code, convex, create, bezier, books, color, algorithm, surface, hidden, human, looking, vector, convert, tar, postscript, form, tool, graphs, header, integrate, graphics, lines, ray, information, such
7: apple, they, edu, life, power, cnn, hacker, modem, just, modems, price, months, computers, re, macs, mac, powerbook, money, battery, dealer, years, day, that, what, duo, he, selling, buy, pluses, heat, electricity, cable
8: drive, scsi, drives, disk, hd, hard, controller, ide, slave, seagate, mb, quantum
9: monitor, simms, cpu, iisi, clock, nubus, mhz, pentium, processor, centris, fpu, performance, simm, lc, board, slot, chips, chip, motherboard, price, vram, lciii, video, quadra, apple, slots
10: window, server, xterm, widget, lib, int, usr, openwindows, motif, colormap, event, application, return, set, default, displayed, events, ld, function, xdm, following, foreground, cursor, null, button, sun, xlib, gc, error, endif, compile, string, pointer, false, pixmap, olwm, sunos, display, mit, trying, xt, problem, clients, undefined, override, running, code, when, using

TABLE 15
20-News: rec. Descriptive features selected for each cluster sorted by coefficient value.

1: car, cars, thanks, engine, ford, eliot, engines, callison, dealer, geico, price, toyota, nissan, autos, vw, auto, models, trunk, turbo, door, mustang, wagon, model, sedan, transmission, audi
2: my, shop, fit, bmw, police, interstate, gently, roads, vehicle, cop, lane, killed, tossing, bikes, lock, tank, driving, gear, gas, kawasaki, required, your, rear, suggestions, bike, miles, mph, filters, oil, cops, truck, motorcycle, com, dod, ride
3: dod, com, uk, ranck, winona, shaft, nec, riding, infante, bc, bike, dog, org, her, ride, biker, rider, countersteering, cookson, ed, hydro, mbeaving, cb, helmet, cage, gsxr, chain, riders, writes, harley, motorcycle, sun, dogs, bikers, egreen, motorcycles, cbr, bmw, zen, bikes
4: he, baseball, runs, ball, pitching, sherri, jewish, pitcher, braves, players, field, ted, career, year, roger, pitchers, mets, inning, hitter, batting, bat, pitch, player, batter, pitched, baseman, alomar, catcher, valentine, hitting, his, clemens, balls, mss, bonds, games, hit, bases, outs, rbi, pitches, batters, league, majors, base, hitters, team, obp, season, game
5: gm, hockey, rochester, mailing, baltimore, tickets, sweden, april, list, please, ericsson, stamber, finland, maine, sports, ahl, champs, teams, slutsky, sport, attend, czech, munich, cup, italy, tournament, nhl, dates, farenebt, providence, shi, contact, briza
6: det, bos, la, nyi, pit, stl, nyr, sj
7: buffalo, detroit, montreal, go, pittsburgh, rangers, puck, play, goals, louis, flyers, blues, pens, game, leafs, winnipeg, espn, devils, bruins, penguins, canucks, quebec, penalties, goal, caps, wings, capitals, playoffs, nhl, kings, islanders, team, cup, ice, overtime, season, playoff, stanley, patrick, calgary, win, games, hockey

TABLE 16
20-News: sci. Descriptive features selected for each cluster sorted by coefficient value.

1: key, clipper, encryption, government, nsa, security, sternlight, crypto, keys, pgp, law, des, privacy, com, eff, secret, chip, police, pmetzger, encrypted, trust, cryptography, court, netcom, algorithm, escrow, fbi, warrant, classified, rsa, secure, wiretap
2: radio, battery, car, michael, hp, plastic, georgia, tone, heat, cooling, wire, batteries, ai, use, bulb, wiring, metal, wd, electronics, amps, wires, covington, capacitor, phone, you, circuit, voltage, concrete
3: thanks, anyone, mb, mail, appreciated, please, am, info, looking, copy, pc, email, appreciate, video, circuit, dsp, hi, ic, motorola, screen, interface, signal, hello, mhz, inputs, advance, eprom, circuits, diode, sehari, suggestions, voltage, pin, audio, games, amp, display, rs, hc, controller, thanx, schematic, help, dos
4: geb, food, she, doctor, medicine, disease, treatment, body, skin, patients, my, blood, medical, dyer, diet, drugs, her, patient, he, eat, studies, yeast, health, symptoms, pain, cause, physician, cancer, infection, foods, effects, med, chronic, surgery, syndrome, brain, infections, doctors, severe, anecdotal, clinical, hospital, diagnosed
5: space, nasa, henry, launch, orbit, earth, flight, sky, moon, spacecraft, pat, gamma, shuttle, dseg, nsmca, cso, atmosphere, prb, rocket, solar, jupiter, man, vehicle, satellite, funding, lunar, mars, observatory, apollo, launched, hst, propulsion, landing, mission, planetary, pluto, payload, manned, orbital, missions

TABLE 17
20-News: mix2. Descriptive features selected for each cluster sorted by coefficient value.

1: windows, file, dos, os, printer, files, font, pc, card, ms, program, using, video, win, drivers, microsoft, driver, tried, disk, mouse, version, window, cica, screen, bmp, progman, machine, ftp, exe, icon, desktop, mb, ini, shareware, meg, vga, click
2: car, cars, engine, dealer, oil, ford, magnus, police, eliot, driving, rear, wheel, sc, price, toyota, auto, engines, honda, wagon, tires, bmw, gear, jeep, nissan, vw, turbo, miles, mustang, transmission, drive, autos, shifting, taurus, door, mazda, dodge, porsche, chevy, valve, drew, wheels, tranny
3: york, mail, philadelphia, thanks, thank, mailing, francisco, please, fax, appreciate, email, address, info, joseph, hello, interested, contact, rangers, angeles, regal, jtchern, ja
4: baseball, he, game, players, team, ball, season, pitcher, year, him, runs, player, his, hitter, roger, career, games, hit, pitching, teams, ted, pitchers, batting, mets, alomar, play, stats, batter, pitch, league, braves, hits, hitting, rbi, hitters, mss, sox, fans, obp, playing, pitches
5: geb, doctor, medicine, medical, treatment, effect, skin, blood, of, is, disease, food, she, body, drugs, patient, med, diet, msg, cause, patients, pain, effects, dyer, infection, drug, symptoms, health, studies, physician, allergic, foods, evidence, chronic, her, yeast, cancer, oral, severe, doctors, treat, therapy, brain, sugar, infections, diagnosed, medication, syndrome, clinical, anecdotal
6: space, pat, nasa, sky, earth, orbit, launch, henry, moon, flight, gamma, shuttle, spacecraft, atmosphere, nsmca, project, srl, satellite, solar, jupiter, government, cost, military, billion, mission, funding, propulsion, landing, zoo, lunar, mars, planetary, observatory, launched, payload, ssto, telescope, pluto, satellites, apollo, orbital, missions, altitude

TABLE 18
20-News: mix. Descriptive features selected for each cluster sorted by coefficient value.

1: windows, file, dos, files, driver, using, thanks, font, ftp, program, drivers, problem, microsoft, advance, win, anyone, tried, running, version, ms, cica, pc, window, ati, progman, utility, hi, exe, printer, ini, shareware, download
2: sale, shipping, offer, condition, asking, interested, sell, mail, email, please, price, brand, obo, contact, monitor, forsale, cd, tapes, stereo, cassette, mint, send, original, selling, external, manuals, offers, drive, sony, keyboard, controller, floppy, manual, warranty
3: baseball, he, players, team, ball, pitcher, season, mets, league, sox, cubs, game, phillies, his, fan, teams, hit, braves, jays, year, giants, career, stats, runs, fans, pitchers, hitter, pitching, yankees, last, alomar, pitch, batting, player, hitting, rbi, games, play
4: space, nasa, pat, earth, orbit, launch, shuttle, moon, henry, atmosphere, spacecraft, energy, srl, gamma, solar, flight, nsmca, sky, science, cost, jupiter, satellite, jbh, project, dseg, military, billion, astronomy, high, of, rockets, mission, lunar, gravity, vehicle, observatory, spencer, orbital, launched, landing, funding, pluto, payload, mars, planetary, satellites, propulsion, manned, altitude, that
5: people, kaldis, com, fbi, clinton, drugs, cramer, health, government, children, writes, rights, trial, you, state, men, koresh, tax, laws, fact, law, country, opinions, gay, their, sexual, not, crime, american, that, americans, themselves, population, we, liberal, by

TABLE 19
20-News: sci comp. Descriptive features selected for each cluster sorted by coefficient value.

1: windows, dos, win, ms, file, os, exe, microsoft, print, ini, cview, cica, version, nt, apps, emm, files, install, ndw, mouse, driver, program, directory, sys, louray, autoexec, unix, config, msdos, disk, drivers
2: ftp, fax, mail, de, domain, visualization, looking, email, format, beck, map, gif, mailing, hi, graphics, ac, pd, archie, binaries, formats, contact, convert, code, program, netherlands, dxf, package, comp, motif, unix, thanks, tiff, hello, jpeg, files, shareware, fortran, pov, conversion, pcx, platforms, image, please, software, address, mcsun, images, amiga, file
3: card, mhz, dx, video, board, cards, lc, simms, eisa, fpu, slot, monitor, cpu, vesa, diamond, ns, price, performance, svga, processor, nubus, iisi, motherboard, upgrade, coprocessor, bit, dram, memory, centris, vram, bus, pin, vlb, slots, cache, ram, mb, isa
4: drive, scsi, drives, disk, controller, hd, ide, mb
5: window, widget, server, xterm, int, lib, display, application, colormap, sun, xlib, drawing, using, problem, usr, event, compile, set, openwindows, motif, error, default, sunos, image, button, menu, width, following, client, pixmap, xdm, program, gc, cursor, return, code, string, compiled, pixel, screen, displayed, endif, mit, xview, font, running, sparc
6: thanks, anyone, please, newsgroup, mail, information, schematic, arcade, pinouts, eprom, gandler, malaysia, thank, ecst, tdawson, email, info, mixer, looking, netters, stalnaker, heikki, appreciated, thanx, tsa, advance, abhin, appreciate, hi
7: tube, solder, power, led, pulse, ink, channel, apple, but, signal, circuit, noise, voltage, equipment, turn, modem, signals, tv, heat, low, monitor, amp, battery, lines, khz, rootstown, cmkrnl, volt, just, at, wires, ground, electronics, when, output, volts, on, circuits, wire, connected, mac, line, that, outlets, you, vdc, is, cable, use
8: clipper, key, encryption, sternlight, government, nsa, crypto, security, law, pgp, privacy, des, cryptography, chip, keys, court, eff, secret, com, encrypted, pmetzger, trust, police, escrow, secure, they, phones, classified, algorithm, wiretap, rsa
9: geb, doctor, food, covington, article, she, medicine, body, blood, skin, treatment, patient, dyer, medical, patients, eat, water, was, diet, med, drugs, cause, disease, he, health, not, years, surgery, people, her, chronic, studies, infection, physician, doctors, effects, his, foods, yeast, symptoms, syndrome, that, cancer, infections, who, been, therapy, severe, illness, diagnosed
10: space, nasa, henry, orbit, earth, launch, pat, spacecraft, shuttle, atmosphere, moon, sky, flight, prb, nsmca, writes, dseg, fraering, solar, observatory, satellite, vehicle, jupiter, fuel, military, uxa, lunar, man, rocket, billion, russian, mars, allen, propulsion, rockets, high, mission, funding, sherzer, landing, payload, telescope, pluto, launched, satellites, manned, orbital, missions, altitude

TABLE 20
20-News: sci comp rec. Descriptive features selected for each cluster sorted by coefficient value.

1: ftp, fax, format, graphics, package, program, image, convert, motif, create, code, comp, help, version, sparc, widget, print, thanks, tiff, domain, email, wireframe, pd, files, shareware, advance, packages, source, application, xlib, formats, pc, hi, postscript, viewer, software, looking, wuarchive, gif, images, unix, display, file, windows
2: windows, window, abpsoft, file, default, manager, lib, server, using, dos, sunos, ini, ms, font, exe, mouse, directory, menu, screen, works, problem, xterm, exit, use, app, os, button, program, error, config, cursor, text, apps, set, icon, microsoft, command, run, files, client, create, shell, unix, display, tried, motif, running, application, code, user
3: drive, mb, scsi, mhz, card, bus, cpu, drives, dx, mac, board, apple, disk, bios, controller, rom, cards, lc, motherboard, port, simms, external, quadra, processor, pin, eisa, pc, video, centris, cable, isa, monitor, cache, ide, system, adapter, ram, fpu, iisi, vlb, meg, jumpers, hd, memory
4: thanks, anyone, please, mail, looking, info, firstclass, email, information, appreciated, gandler, interested, thank, thanx, appreciate, anybody, hello, schematic, hi, post, arcade, advance, reply, address, am, pointers, fabled, newman, erini, any, bobc, stalnaker, responses, multisync, replies, help
5: key, clipper, encryption, sternlight, nsa, government, security, chip, keys, amendment, law, des, netcom, gtoal, crypto, eff, cryptography, privacy, pmetzger, algorithm, encrypted, pgp, speech, denning, secret, trust, clinton, phones, escrow, court, warrant, secure, classified, police, rsa, cryptographic, wiretap, feds, fbi, terrorists, public, enforcement, private, crypt
6: canada, sweden, usa, april, det, uk, germany, phone, pt, que, france, subscribe, se, university, europe, austria, nerone, fax, energieanwendung, de, internet, switzerland, munich, ga, pp, dortmund, contact, tone, list, sj, finland, la, min, czech, peter, email, ericsson, hc, mail, mtl, smail, address, petr
7: geb, amp, medicine, food, cause, that, skin, not, patient, heat, dyer, doctor, reaction, are, symptoms, usually, she, water, patients, years, treatment, about, physician, yeast, of, few, diet, people, they, doctors, blood, was, very, infection, pain, medical, effects, were, syndrome, health, as, disease, most, because
8: nasa, space, orbit, henry, launch, pat, shuttle, earth, spacecraft, energy, flight, moon, nsmca, nuclear, atmosphere, prb, sky, high, solar, fraering, project, lunar, satellite, jupiter, military, funding, rocket, propulsion, planets, billion, observatory, landing, gamma, zoo, mission, rockets, spencer, pluto, telescope, development, payload, ssto, mars, manned, orbital, satellites, planetary, missions, altitude
9: dod, car, bike, article, engine, cars, com, bmw, ride, eliot, bgardner, rear, wheel, rider, yamaha, oil, ranck, road, honda, my, gear, bikes, shaft, infante, tires, toyota, egreen, motorcycles, brake, driving, seat, front, gas, riding, gt, riders, motorcycle, gs, steering, miles, you, passenger, tire, wheels, mph, brakes, clutch
10: he, his, baseball, him, pitcher, ball, year, player, batting, career, hitting, baseman, players, pitch, pitchers, ted, hit, defensive, pitches, bonds, inning, braves, pitching, sherri, alomar, hitter, last, bat, singer, season, batters, rbi, obp, roger, stats, league, game, team, games
11: philadelphia, go, montreal, detroit, hockey, rangers, boston, division, game, blues, teams, buffalo, games, bruins, nhl, devils, play, pittsburgh, league, espn, wings, stanley, team, cup, leafs, jets, capitals, caps, goals, playoffs, fans, playoff, win, islanders, canucks, fan, pens, season

TABLE 21
20-News (all). Descriptive features selected for each cluster sorted by coefficient value.

1: windows, file, window, files, motif, program, server, font, xterm, widget, microsoft, comp, advance, ftp, postscript, application, openwindows, compile, code, xv, format, client, sunos, manager, gif, xview, create, xlib, directory, bitmap, unix, default, ms, version, shareware, libraries, bmp, graphics, sparc, package, image, user, menu, dos, display, running, app, screen, software
2: card, monitor, video, mhz, monitors, cpu, diamond, apple, board, dx, centris, upgrade, lc, vesa, cards, lciii, simms, vga, ati, irq, port, motherboard, memory, color, simm, mac, coprocessor, processor, iisi, mouse, mode, fpu, ram, pin, connector, pc, nubus, serial, quadra, svga, isa, hardware, bios, ports, computer, vlb, vram, slots, bus, mb
3: drive, scsi, drives, disk, hd, ide, rom, mb, controller, bios, harddisk, disks, hard, jumper, quantum, jumpers, pc, formatted, maxtor, seagate, floppy, mfm
4: sale, mail, email, price, offer, vhs, ink, condition, address, contact, interested, phone, catalog, shack, list, impedance, cassette, shipping, circuit, forsale, joysticks, wanted, sell, audio, analog, stereo, buy, modem, electronic, info, send, cd, obo, fax, oscilloscope, printer, mint, box, voltage, output, selling, cable, package, amp, software, cod, amplifier, reply, brand
5: dod, car, cars, bike, engine, bmw, riding, motorcycle, ford, bgardner, ride, road, oil, bikes, vw, gear, dealer, motorcycles, mph, nissan, driving, cb, eliot, exhaust, yamaha, wheel, porsche, brake, rider, egreen, mustang, zx, rear, countersteering, passenger, toyota, insurance, tire, seat, tires, riders, front, torque, honda, gs, clutch, wheels, steering, brakes, miles
6: team, season, game, baseball, nhl, hockey, players, year, braves, montreal, hitter, teams, ball, pitcher, league, play, mets, rangers, playoff, batting, wings, goalie, pitchers, detroit, jays, coach, playoffs, cup, phillies, pitching, player, leafs, stats, roger, sox, win, bruins, devils, games, fan, fans, scored, boston, espn, pittsburgh, penguins, scoring, pens, islanders, played
7: clipper, key, chip, algorithm, encryption, tapped, nsa, phone, security, crypto, cryptography, sternlight, keys, des, pgp, denning, secret, encrypted, wiretap, privacy, eff, cryptographic, rsa, government, secure, escrow, crypt
8: god, atheists, jesus, bible, christian, verse, morality, lord, church, religion, christians, athos, christ, christianity, argument, catholic, atheism, faith, sin, atheist, benedikt, heaven, moral, beliefs, rutgers, holy, clh, truth, doctrine, verses, sandvik, gospel, biblical, scripture, belief, testament, religions, religious
9: space, nasa, orbit, moon, distance, shuttle, project, launch, henry, research, spacecraft, high, flight, temperature, earth, nuclear, gamma, solar, sci, mission, sky, energy, pat, lunar, nsmca, planets, prb, images, science, propulsion, zoo, spencer, atmosphere, satellite, observatory, ray, mars, satellites, landing, funding, orbital, ssto, jbh, payload, fraering, pluto, planetary, orbiting, missions, altitude
10: geb, cramer, article, doctor, study, kaldis, food, symptoms, health, medicine, patients, foods, sexual, pregnancy, doctors, med, treatment, skin, infection, gay, patient, medication, sex, physician, chronic, yeast, dyer, effects, studies, disease, medical, women, disorder, infections, percentage, syndrome, homosexuals, people, drugs, diet, drug, cancer
11: israel, turkish, israeli, turks, muslim, jews, arab, bosnia, war, occupied, iran, jewish, policy, countries, soldiers, nazi, peace, anti, troops, muslims, propaganda, military, genocide, civilians, israelis, political, land, palestinian, arabs, army, forces, villages, terrorism, country, occupation, population, killed, inhabitants, ethnic, government, people
12: gun, atf, waco, cdt, guns, firearms, fire, clinton, weapons, batf, koresh, fbi, gas, tax, constitution, state, country, compound, crime, article, armed, people, police, utkvm, government, death, dscomsa, federal, murder, assault, reno, rights, bd, killed, jmd, deaths, amendment, children, tanks, davidians, citizens, raid, law, arms

TABLE 22
Reuters. Descriptive features selected for each cluster sorted by coefficient value.

1: split, declared, date, dividend, shareholders, distribution, sets, payable, board, stock, paid, holders, payout, quarterly, share, record
2: div, record, pay, payout, april, qtly, prior, cts
3: loss, shr, net, qtr, revs, note, cts, oper, jan, avg, profit, mths
4: sell, unit, company, terms, undisclosed, expects, completed, buys, division, acquisition, purchase, acquire, disclosed, acquired, earnings, completes, sale, agreed, acquires, sells, subsidiary, intent, expected, equipment, business, revenues, earned, buy, announced, transaction, privately, principle, units, makes, closing, agreement, services, based, products, manufacturing, signed, systems, cash, previously, corp, owned, subject, assets, definitive, dlrs
5: offer, bid, stake, takeover, shares, merger, control, airlines, court, commission, usair, tender, received, proposal, group, buyout, caesars, filing, buy, cyclops, taft, filed, comment, voting, board, rights, investor, owns, proposed, acquire, advisor, seeking, securities, tendered, directors, outstanding, unsolicited, bought, holds, investment, stock, common, management, shareholders, shareholder, exchange, share, led, pct, company
6: trade, grain, dispute, agriculture, japanese, japan, officials, foreign, states, told, senate, ships, work, congress, house, union, tariffs, reagan, official
7: stg, billion, fed, turnover, francs, profits, compared, marks, guilders, income, loans, parent, improvement, earnings, reported, customer, pretax, forecast, year, tax, earlier, system, january, pre, ag, total, results, reserves, deposits, unchanged, extraordinary, rose, offset, repurchase, fell, profit, mln, rise, net
8: rate, bank, pct, banks, rates, prime
9: oil, crude, prices, sea, barrels, refinery, shipping, bpd, port, opec, iranian, barrel, petroleum, gulf, minister, production, shell, kuwait, iran, output
10: monetary, baker, dollar, currency, currencies, finance, exchange, rates, germany, economic, accord, nations, paris, stability

TABLE 23
Recipe. Descriptive features selected for each cluster sorted by coefficient value.

1: seed(s), masala, dal, cumin seed(s), turmeric, red chili, cashew(s), green chilies, curry leaf(ves), cardamom, chili(es), curry, coriander, mustard seeds, basmati, clove, cinnamon sticks, gram, chili powder, paneer, tumeric, coconut, curds, ginger, amchur, ginger paste, chickpea(s), lentil(s), asafoetida, ghee, cumin, garlic paste, leaves, cilantro leaves, coriander powder, yoghurt, fenugreek, ground cumin, cinnamon, salt, red chili powder, powder, cilantro
2: part skim, mozzarella, cheese, italian, pizza, basil, lasagna, tomato(es), bread, ricotta, pasta sauce, marinara sauce, parmesan, pasta, pesto, dough(s), frozen chopped, spinach, cottage, ground beef, oregano, tomato, tomato sauce, sauce
3: self, vegetable oil, rising, eggs, garlic herb, large eggs, boudin, plain flour, chop(s), crumb(s), pink salmon, tapioca flour, breadcrumbs, grits, sharp, bread, catfish fillets, matzo, potatoes, pan drippings, black treacle, mustard, cream style corn, pickle(s), bread crumb(s), mayonaise, freshly ground, saltine(s), egg(s), ham, sausages, milk, seasoning salt, meal, bacon, green tomatoes, prepared mustard, russet, cornmeal, buttermilk, butter, crumbs, flour
4: all purpose, large eggs, baking powder, soda, flour, eggs, baking, chocolate candy, buttermilk, color food green, corn meal, biscuits, bran, unsalted butter, baking soda, drippings, butter, lard, shortening, extract, cornmeal, vanilla, melted, milk, white sugar, sugar, salt, large, large egg(s)
5: cheese, parmesan, bread slices, grated, grated parmesan, mushrooms, half & half, butter, sage, broth, heavy, potato gnocchi, prosciutto, gorgonzola, bacon, fresh asparagus, fresh basil, basil leaves, whipping, polenta, fontina, grated parmesan cheese, fettucine, pasta, asparagus, grated nutmeg, penne, cream cheese, pinenuts, basil, mozzarella cheese
6: orange(s), brown sugar, sugar(s), tea(s), pie, syrup, ice, condensed milk, apples, whiskey, gelatin, bananas, vanilla, cardamom, cachaca, fruit, marshmallow(s), pecan(s), peach(es), strawberri(es), cinnamon, whipped, sweetened, rum, raspberries, coffee, firmly, cake, almond(s), blackberry(ies), cherries, liqueur, chocolate, packed, cookie(s), graham cracker(s), espresso, evaporated milk, mint sprigs, light, cream, brandy, powdered, light brown, cocoa, nutmeg, granulated sugar, whipping, milk, extract
7: red pepper(s), flakes, vinegar, red wine, tomatoes, dried tomatoes, crushed, collard(s), feta, onion, cherry, olives, dried, penne, crushed red, crushed red pepper flakes, caper(s), balsamic, kalamata, oregano, basil, olive(s), dried basil, linguine, crushed tomatoes, garlic cloves, pasta, red, pepper flakes, red pepper flakes
8: tortilla(s), cheddar cheese, beans, taco, old el, soup, mix, jack, cream, shredded, ground beef, pico de, sour cream, green chilies, chips, kernel, salsa, black olives, enchilada sauce, shredded monterey, sharp, diced green chilies, cheese, refried, cooked chicken, ground turkey, dressing, mexican cheese, mexican, chili powder, ripe olives, corn, green chile(s), colby, kernel(s), picante sauce, lettuce, ranch, shredded cheddar cheese, tomato(es), green, shredded lettuce, diced, corn tortilla(s), black bean(s), corn tortillas, guacamole, avocado(s), onion(s), chili(es)
9: jalapeno(s), best, lime, lime juice, cilantro, chipotle(s), avocado(s), tomatillos, crema, tequila, chile(s), cilantro fresh, ancho, onion, purple, queso, hominy, white onion, kernel(s), serrano, corn tortillas, epazote, pork shoulder, tomatoes, cotija, jicama, chopped cilantro, poblano, salt, mayonaise, tortilla(s), chili(es), coarse salt, mexican, corn, chiles, black beans, salsa, fresh, jack, cumin, cilantro leaves, onion(s)
10: extra virgin olive oil, olive oil, garlic cloves, olives, tuna, sea salt, sherry vinegar, pecorino, balsamic, blanched almonds, vinegar, ground pepper, arugula, freshly, basil leaves, cured, baguette, dijon mustard, pinenuts, fine sea salt, fine(s), basil, cheese, fresh lemon, lemon, pepper(s), white wine, pepper
11: lemon, lemon juice, yogurt, horseradish, mayonaise, feta cheese, cucumber, fillet(s), fresh, fresh chives, greek, pita(s), dijon, creole mustard, fresh lemon, dill, mint, salmon, fresh lemon juice, chopped fresh, grated lemon, olives, lemon zest, capers, parsley, fresh parsley, fresh mint, kalamata, chopped fresh mint
12: skinless, chicken, boneless, breast(s), skinless chicken, chicken breast(s), boneless skinless chicken, skinless chicken breast(s), breast, boneless skinless chicken breast(s), soup, chicken breasts, breasts, dressing, breast halves, chicken breast, chicken breast halves, biscuit(s), boneless chicken, poultry seasoning, cream of chicken soup, onions, onion(s), broth, skinless chicken breast, chicken broth, boneless chicken breast, shredded, green, sodium, pepper
13: black pepper, pork, ground, cider, cayenne, guajillo, dried, barbecue sauce, ground black pepper, beef, roast, garbanzo beans, garlic powder, lamb, paprika, allspice, chickpeas, stew, ketchup, cinnamon, chops, garlic, ground ginger, lentil(s), oregano, thyme, tomato, onion, chili powder, salt, shoulder, couscous, kosher, ground cumin, tomato(es), ground cinnamon, ground cloves, onion powder, diced tomatoes, dried thyme, cumin, ground coriander, pepper, dried oregano, onion(s)
14: fat, less, fat free, low, fat free less sodium, sodium, chicken broth, sodium chicken broth, broth, reduced, cooking, %, chopped onion, cooking spray, garlic cloves, chopped, % low, % low fat milk, low sodium
15: bell pepper(s), green(s), celery, long, red beans, kidney, okra, green, cooked, sausage(s), seasoning, hot, cajun, smoked sausage, onions, shrimp, andouille, onion(s), green bell, creole
16: yeast, warm water, water, active dry yeast, unbleached flour, bread flour, masa, flour, barley, ghee, wheat flour, dry, dry yeast, semolina, instant, shortening, eggs
17: white, grain, wine, leaf parsley, broth, collard green(s), low salt chicken, thyme, dry, black peppercorns, red wine, snails, truffle(s), saffron, shallots, bay leaf(ves), heavy, arborio rice, leaf, tarragon, pearl onions, bacon, mushrooms, dry white, ham, veal, stock, porcini mushrooms, rosemary, vinegar, sprig(s), chopped fresh, clam(s), grain white, white wine vinegar, garlic cloves, leek(s), butter, parsley, white wine, celery, wine vinegar, white onion, dried thyme, olive oil, white pepper, white sugar
18: potato(es), half & half, bay, butter, carrot(s), peas, ground pepper, pastry, turnip(s), beef, ham, grits, thyme, bacon, onions, cabbage, stock, sausage(s), rib(s), meat, leek(s), worcestershire sauce, dill, celery, half(ves), smoked, yellow onion, parsley, salt, sweet potato(es), beef stock, pepper
19: egg, yolk(s), puff, egg(s), large egg(s), chocolate, large egg, almond(s), mascarpone, cake, vanilla, cream, large egg yolks, of tartar, espresso, preserves, jam, large eggs, large egg whites, bittersweet chocolate, hazelnut(s), marnier, confectioners sugar, semisweet chocolate, whole milk, pistachio(s), cocoa, vanilla beans, ladyfingers, coffee beans, gelatin, powdered, heavy, whipping cream, raspberry, extract, strawberri(es), dark rum, sugar, unsalted butter, liqueur, whole, coffee, almond, milk, granulated sugar, butter, flour
20: lime, fish sauce, coconut milk, fish, thai, leaves, mint, coconut, chili pepper(s), prawn(s), rice stick(s), vermicelli, sauce(s), chili sauce, curry paste, ginger, basil, peanut(s), peanuts, shallots, spring, cucumber(s), cilantro, tofu, shrimp paste, brown, sriracha, shrimp, sugar, beansprouts, rice, noodle(s), rice paper, rice noodles, chile(s), jasmine, lime leaves, paste, lemongrass, peanut, chili, cilantro leaves, milk, soy sauce
21: soy sauce, sesame, five, gochujang, bean, chinese, wrappers, szechwan, sauce(s), scallion(s), oyster sauce, nori, pork, shiitake, ramen, sushi, miso, shaoxing, mirin, water chestnuts, japanese, sake, tofu, cabbage, spring, seaweed, napa cabbage, rice wine, ginger, hoisin, green onions, udon, mushrooms, dashi, white pepper, wasabi, rice vinegar, starch, daikon, kimchi, noodles, rice, sesame oil, vinegar, seed(s), carrots, ground pork

TABLE 24
NSF. Descriptive features selected for each cluster sorted by coefficient value.

1: theory, conjecture, geometry, spaces, algebraic, algebras, numbers, manifolds, theorem, investigator, groups, questions, invariants, space, objects, polynomials, curves, functions, oldest, proved, representation, topology, lie, arithmetic, classification, topological, analytic, cohomology, finite, geometric, algebra, study, representations, euclidean, infinite, conjectures, operators, invariant, abstract, symmetries, classical, operator, combinatorics, polynomial, mathematics, dimensional, hilbert, mathematical, problems, physics
2: mantle, rocks, magma, tectonic, magmatic, terranes, crustal, constrain, crust, seismic, km, plate, deformation, orogenic, belt, geochemical, ridge, volcanic, subduction, spreading, uplift, dating, deposits, basin, faulting, continental, isotopic, zone, magmas, margin, tectonics, petrologic, faults, minerals, zones, arc, rock, evolution, east, volcanism, geologic, fault, mineral, magmatism, lithosphere, hydrothermal, lithospheric, geological, rift, geophysical
3: laboratory, courses, analysis, students, introductory, exercises, geology, experiments, lecture, instrument, lab, laboratories, equal, software, principles, modern, acquisition, macintosh, biology, spectrometer, equipment, undergraduate, concepts, majors, instructor, curriculum, instruments, instrumental, computers, taught, computer, student, instruction, modules, learning, equipped, instrumentation, spectrophotometer, purchase, workstations, teaching, instructional, teach, chemistry, hands-on, college, department, projects, experience, faculty
4: equations, solutions, problems, nonlinear, models, numerical, asymptotic, methods, equation, method, boundary, fluid, partial, flows, random, investigator, flow, mathematical, approximations, convergence, dynamical, stability, problem, stochastic, finite, phenomena, parameter, arise, behavior, singular, inverse, variational, linear, singularities, solution, dynamics, chaotic, applied, approximation, analysis, differential, arising, computational, complicated, mechanics, applications, physical, systems, algorithms, theory
5: reactions, molecules, professor, complexes, catalysts, compounds, chemistry, synthetic, reaction, reactivity, dr, studies, bond, metal, ions, radicals, catalytic, solvent, macromolecular, spectroscopic, chiral, intermediates, synthesis, ligands, organometallic, program, molecule, organic, bonds, products, chemical, supporting, university, reagents, spectroscopy, oxidation, reactive, catalysis, resonance, intramolecular, dynamics, electron, hydrogen, award, molecular, nmr, atoms, transition, department, inorganic
6: dissertation, political, interviews, people, social, cultural, empirical, economic, examines, archaeological, legal, literature, examine, labor, conflict, firms, outcomes, trade, historical, beliefs, capital, societies, policy, prices, data, individuals, survey, factors, institutional, goods, economy, household, public, decisions, income, implications, ethnographic, markets, households, case, census, risk, welfare, market, decision, organizational, government, organizations, policies, research
7: carbon, water, waters, concentrations, dissolved, ocean, biogeochemical, nitrogen, phytoplankton, atlantic, circulation, production, rates, quantify, sediments, isotopes, ecosystems, biomass, fluxes, anthropogenic, soil, microbial, nutrients, uptake, seasonal, co2, tracer, cycling, composition, sediment, benthic, flux, trace, matter, particulate, seawater, organic, measurements, global, coastal, processes, nutrient, variability, atmosphere, marine, ecosystem, atmospheric, productivity, climate, sea
8: feasibility, phase, conventional, high, demonstrated, cost, sbir, small, process, commercial, laser, business, low, compact, innovation, fiber, technique, efficiency, imaging, fabrication, device, detection, proposed, operating, ii, technology, sensor, demonstrate, fabricated, beam, operation, surface, instrument, prototype, sensitivity, material, fabricate, capability, sensors, optical, silicon, power, monitoring, manufacturing, processing, applications, devices, performance, semiconductor, design
9: species, populations, predators, fossil, diversity, habitat, habitats, dispersal, offspring, taxa, morphological, predation, evolutionary, traits, patterns, plants, insects, variation, history, females, phylogenetic, genus, genetic, test, taxonomic, relationships, population, ecological, birds, genera, reproductive, natural, plant, male, hypotheses, animals, fitness, breeding, characters, insect, reproduction, organisms, mating, phylogeny, selection, biodiversity, conservation, evolution, ecology, dna
10: brain, cells, neurons, nerve, perception, mechanisms, visual, humans, hormones, cell, stimuli, stimulus, muscle, nervous, animals, synaptic, normal, movements, secretion, cortex, stimulation, hormone, receptors, sensory, role, motor, electrophysiological, activity, receptor, tissue, neural, body, signals, physiological, neuronal, vertebrates, movement, neurotransmitters, responses, vertebrate, neurotransmitter, function, system, mammalian, animal, cellular, understanding, developmental, regulation, neuroscience
11: support, cooperative, university, dr, equipment, visit, oceanographic, operated, scientific, purchase, national, vessel, fellowship, expertise, funding, ship, award, months, funds, facilities, nsf, sciences, facility, center, director, installation, research, marine, recommends, acquisition, joint, oceanography, scientists, institute, shipboard, researchers, conduct, institution, qualified, contributing, postdoctoral, dedicated, instrumentation, calendar, awards, maximal, program
12: materials, properties, films, alloys, microstructure, microscopy, devices, polymer, temperature, microstructural, epitaxial, film, mechanical, interfacial, crystals, polymers, fundamental, nanoscale, diffraction, crystal, composites, defects, structures, optical, magnetic, crystalline, interfaces, thin, silicon, oxide, coatings, bulk, atomic, scanning, semiconductors, surface, technologically, liquid, thermal, electrical, surfaces, processing, electron, scattering, characterization, fabrication, semiconductor, metal, spectroscopy, electronic
13: protein, mutants, proteins, gene, mutant, genes, expression, pathway, dna, yeast, coli, binding, mutagenesis, rna, mechanism, vitro, cloned, mutations, enzymes, regulatory, biochemical, acid, sequence, identified, transcription, vivo, arabidopsis, isolated, specific, sequences, bacterial, mrna, plants, enzyme, metabolism, eukaryotic, genome, regulated, involved, amino, plant, function, regulation, cell, functional, cellular, molecular, genetic, mechanisms, cells
14: model, measurements, flow, ionosphere, wave, wind, particle, effects, waves, weather, ionospheric, thermosphere, magnetosphere, turbulence, observed, velocity, plasma, convective, precipitation, transport, dynamics, observations, storms, turbulent, radar, cloud, vorticity, numerical, convection, models, variability, heating, simulations, mesoscale, vertical, scale, heat, horizontal, atmosphere, layer, momentum, atmospheric, doppler, parameterizations, flows, boundary, winds, processes, forcing, circulation
15: workshop, conference, held, meeting, international, travel, researchers, forum, papers, future, bring, symposium, scientists, discuss, june, committee, speakers, organizers, published, society, proceedings, directions, leading, august, american, october, support, organized, countries, attendees, july, invited, september, attend, experts, participants, sessions, discussion, participation, exchange, topics, attendance, young, poster, presentations, discussions, session, plenary, ideas, talks
16: earthquake, design, control, loads, buildings, process, manufacturing, product, damage, objective, geotechnical, reinforced, develop, steel, loading, load, dynamic, vibration, planning, industry, concrete, methodology, failure, power, uncertainties, analytical, system, safety, fault, reliability, stiffness, actuators, civil, developed, controllers, industrial, modeling, performance, cost, intelligent, engineering, controller, engineers, models, feedback, robust, mechanical, systems, optimal, optimization
17: quantum, theoretical, physics, atoms, quark, states, violation, atomic, experimental, experiment, interactions, nuclei, nucleon, nuclear, quarks, particle, electrons, excitations, decay, collisions, photon, meson, fermilab, collider, magnetic, particles, matter, spin, electron, scattering, accelerator, polarized, superconductors, nucleons, energy, detector, excitation, charge, experiments, qcd, condensed, calculations, relativistic, theory, energies, proton, heavy, superconducting, excited
18: ice, climate, records, record, glacier, cores, climatic, antarctic, sea, paleoclimatic, glacial, late, dating, stratigraphic, paleoclimate, holocene, arctic, antarctica, chronology, lake, years, deposits, age, data, pleistocene, past, proxy, variability, high-resolution, stratigraphy, radiocarbon, paleoenvironmental, region, glaciation, northern, atmospheric, alaska, western, pollen, southern, variations, award, sediments, history, sediment, global, circulation
19: algorithms, techniques, languages, efficient, user, language, query, reasoning, complexity, automatic, tasks, framework, specification, algorithm, execution, parallel, code, automatically, software, compiler, architectures, logic, semantics, programming, performance, tools, architecture, processor, memory, application, processing, machine, scheduling, distributed, implementation, parallelism, computation, correctness, processors, optimization, implementations, machines, systems, applications, computational, hardware, computing, computer, problems, design
20: program, reu, graduate, students, university, engineering, faculty, undergraduates, scholars, minority, academic, entering, women, scholarship, summer, talented, institutions, career, scholarships, programs, ethics, participate, participation, recruitment, internships, undergraduate, mentoring, careers, seminars, recruited, retention, student, weekly, college, colleges, projects, seminar, opportunities, mentor, underrepresented, experiences, activities, recruiting, trips, minorities, mentors, year, education, participants, science
21: teachers, teacher, mathematics, science, school, schools, content, standards, systemic, grades, learning, project, children, middle, math, reform, girls, assessment, staff, parents, instructional, activities, achievement, dissemination, community, district, education, classrooms, public, educators, districts, technology, grade, educational, curriculum, leadership, classroom, workshops, elementary, professional, materials, programs, skills, standards-based, curricula, teaching, instruction, college, student, students
22: galaxies, stars, observations, telescope, star, observatory, ast, solar, stellar, radio, interferometer, detector, galaxy, cosmic, dr, sky, gravitational, mass, galactic, emission, pulsars, supernova, observational, spectra, interstellar, quasars, spectral, angular, wavelengths, universe, infrared, objects, sun, evolution, clouds, observatories, dust, astrophysical, radiation, supernovae, photometric, astronomical, observing, faint, extragalactic, telescopes, luminosity, astrophysics, astronomers, astronomy
23: network, internet, access, services, connection, resources, networks, wireless, bandwidth, users, information, nsfnet, management, service, file, multicast, digital, connect, vbns, connections, packet, remote, protocol, speed, libraries, traffic, routing, atm, servers, user, resource, protocols, supercomputers, infrastructure, performance, requests, qos, testbed, server, support, ip, requirements, backbone, scalable, networking, distributed, communication, multimedia, communications

TABLE 25
AG's news. Descriptive features selected for each cluster sorted by coefficient value.

1: sales, quarter, profit, earnings, percent, research, profits, shares, stores, year, loss, mart, wall, results, outlook, net, maker, quote, update, expectations, chain, fiscal, retailer, forecast, forecasts, rose, increase, growth, increased, reported, revenue, demand, drop, fall, income, largest, posted, sees, lower, rise, estimates, posts, company, fell, strong, higher, corp, quarterly
2: company, drug, executive, motors, financial, group, research, billion, buy, jobs, ceo, accounting, amp, companies, investment, corp, million, stock, quote, buys, firm, announced, plans, kmart, cash, insurance, shareholders, maker, business, bank, plc, nortel, bn, stake, largest, based, holdings, sale, acquire, sell, management, unit, cut, shares, merger, update, biggest, giant, deal, agreed
3: bankruptcy, million, airlines, pay, airways, airline, securities, deal, air, workers, hollinger, accounting, executive, insurance, agreed, plan, contract, sec, executives, fund, enron, agreement, cuts, boeing, employees, company, yesterday, federal, filed, union, lawsuit, settlement, unions, bankrupt, delta, alitalia, financial, debt, fraud, billion, charges, settle, pilots, carrier, attendants, cost, update, pension, cut, restructuring
4: oil, prices, crude, price, fuel, opec, gas, futures, winter, supplies, energy, barrel, supply
5: reuters, dollar, rates, rate, jobless, economy, figures, economic, euro, average, survey, tokyo, treasury, month, august, growth, investors, report, showed, reserve, year, september, july, consumer, index, market, inflation, data, gains, fed, economists, pace, bank, interest, week, nikkei, lower, fell, rise, federal, low, slightly, higher, expected, yen, jobs, rose, stocks, percent, prices
6: oracle, peoplesoft, bid, takeover, offer, abbey, woolworths, software, wmc, ellison, chilcott, battle, enterprise, rival, applications, shareholders, bn, rejects, company, monti, billion, rejected, hostile, psft, orcl
7: wireless, phone, mobile, internet, fi, service, phones, cell, nokia, services, broadband, sprint, communications, network, blackberry, docomo, cellphone, spam, messaging, networks, verizon, vodafone, ip, voip, net, technology, siemens, telecom, access, vonage, customers, cellular, device, provider, users, voice, subscribers, connections, mail, telephone, companies, handsets, fcc, telephony, speed, motorola, based, providers, telecommunications, cable
8: search, google, web, engine, yahoo, ipo, rss, pages, beta, site, click, tool, links, blog, initial, page, sites, online, content, information, amazon, desktop, application, ads, offering, tools, mail, advertising, rsquo, toolbar, browser, popular, searchable, searches, internet, service, users, goog
9: ibm, apple, dvd, hp, pc, computer, format, digital, intel, storage, sony, ipod, power, technology, sun, hardware, music, portable, chip, gb, computing, memory, design, amd, graphics, dell, systems, toshiba, processor, product, announced, processors, notebook, electronics, unveils, servers, devices, linux, based, system, chips, dual, products, computers, version, market, server, desktop, company
10: microsoft, software, windows, security, worm, flaw, vulnerability, xp, code, linux, system, spam, firewall, browsers, application, source, firefox, spyware, browser, novell, virus, systems, applications, mozilla, versions, products, users, version, updates, explorer, tools, mail, os, hackers, sp, enterprise, release, malicious, aol, vulnerabilities, server, features, web, flaws, service, released, operating, open, internet, computer
11: gt, lt, fullquote, http, investor
12: space, nasa, earth, scientists, mars, astronomers, moon, prize, astronaut, genesis, spacecraft, saturn, flight, rocket, stars, cassini, lunar, sky, planet, capsule, surface, rutan, universe, launch, fly, solar, shuttle, science, jupiter, mission, telescope, spaceshipone, canaveral, titan, astronauts, planets, moons, craft, planetary, atmosphere, mojave, propulsion, satellite, orbit, miles, probe, desert, exploration, unmanned, orbiting
13: court, federal, drug, trial, law, pinochet, online, sex, rape, equatorial, years, sexual, industry, movie, case, file, copyright, press, cp, alleged, martha, prison, peer, attorney, murder, jail, guilty, convicted, charges, british, child, canadian, justice, lawyers, legal, sharing, prosecutors, recording, sentence, piracy, filed, charged, accused, judge, illegally, fraud, criminal, internet, music
14: quot, writes, coach, stewart, halo, president, minister
15: china, japan, scientists, researchers, bird, cambodia, chinese, disease, humans, south, study, fossil, animals, afp, sea, north, hong, death, island, children, human, species, beijing, typhoon, arctic, taiwan, submarine, ancient, earthquake, communist, state, decades, explosion, quake, global, climate, found, pacific, years, asian, asia, decade, country, korean, warming, missing, northern, people, toll, korea
16: gold, olympic, red, athens, sox, yankees, pedro, olympics, boston, medal, games, phelps, series, doping, women, marathon, men, meters, greece, gymnastics, schilling, hamm, american, silver, sprinter, marion, athletes, radcliffe, medals, sprinters, team, athletics, bronze, meter, greek, relay, world, night, win
17: season, quarterback, game, sports, yards, touchdown, nl, nba, miami, offense, bowl, coach, points, houston, dodgers, cleveland, field, night, giants, atlanta, kansas, ranked, defensive, college, receiver, career, nfl, left, eagles, bay, mike, pittsburgh, seattle, minnesota, back, redskins, florida, games, manning, football, teams, patriots, inning, play, victory, team, st, win, league, beat
18: cup, open, grand, round, masters, golf, championship, woods, champion, world, schumacher, final, tour, seed, par, race, henman, classic, tennis, title, vijay, sharapova, williams, hewitt, seeded, spain, els, lead, agassi, federer, event, finals, mauresmo, win, top, qualifying, championships, tournament, roddick, wins, atp, shot, pga, wta, sunday, semifinals, davenport, match, stroke, victory
19: striker, manchester, arsenal, league, manager, chelsea, champions, england, club, liverpool, united, premiership, celtic, newcastle, football, side, draw, boss, beckham, real, milan, team, players, coach, sir, wales, barcelona, rooney, english, goal, fa, portsmouth, defender, tottenham, trafford, midfielder, uefa, madrid, blackburn, captain, soccer, bolton, defeat, match, everton, goals, southampton, fulham, owen, cup
20: ap, coach, quarterback, california, helens, bowl, nfl, bay, st, miss, bryant, defensive, night, points, pittsburgh, colorado, receiver, career, touchdown, nba, scored, oklahoma, season, game, democratic, sen
21: india, test, australia, cricket, minister, pakistan, sri, indian, day, delhi, pm, zealand, pervez, africa, bangladesh, wickets, wicket, captain, tendulkar, kashmir, batting, islamabad, ganguly, south, aziz, pakistani, australian, nagpur, manmohan, icc, match, indies, afp, series, trophy
22: president, bush, john, election, presidential, susilo, vote, administration, voters, republican, kerry, ukraine, house, voting, debate, karzai, political, campaign, presidency, afghanistan, electoral, opposition, democratic, senate, democracy, party, war, terror, bill, elections, afghan, leader, polls, poll, candidate, votes, sen, democrats, republicans, parliament, congress, afp, incumbent, country, cheney, candidates
23: sudan, nuclear, darfur, talks, iran, nations, african, government, eu, states, ivory, peace, region, congo, treaty, uranium, union, european, wto, council, sudanese, rebels, libya, crisis, turkey, atomic, united, aid, countries, leaders, africa, secretary, end, resolution, reuters, south, country, negotiations, trade, rebel, foreign, sanctions, security, diplomats, afp, urged, meeting, summit, discuss, arms
24: people, police, hurricane, killed, bomb, car, storm, capital, attack, plane, haiti, crash, dead, school, killing, victims, blast, suspected, muslim, separatist, chechen, afghan, injured, authorities, died, officials, bodies, russia, deadly, suicide, security, blasts, arrested, bombs, attacks, embassy, kills, death, city, bombing, southern, explosion, wounded, exploded, forces, militants, kill, soldiers, wounding, baghdad
25: iraq, iraqi, british, al, hostage, najaf, blair, sadr, hostages, troops, kidnapped, saddam, minister, kidnappers, hassan, prisoners, fallujah, abu, falluja, army, release, bigley, italian, freed, military, muslim, sunni, allawi, shi, shiite, forces, cleric, aid, soldiers, government, insurgents, baghdad, captors, held, rebel, iraqis, militants, stronghold, militant, interim, fighting, zarqawi, insurgent, city, kill
26: palestinian, arafat, gaza, israeli, israel, sharon, palestinians, leader, west, hamas, jewish, israelis, jerusalem, bank, egypt

TABLE 26
Reviews: movie kitchen. Contingency table between the clusters and true topic categories.

	movie	kitchen
1: film, movie, story, films, life,. . .	**2000**	23
2: bought, product, buy, clean, kitchen, pan,. . .	0	**1977**

TABLE 27
Reviews: movie books. Contingency table between the clusters and true topic categories.

	movie	books
1: film, movie, films, director, comedy,. . .	**2000**	156
2: book, read, books, reading, author, reader,. . .	0	**1844**

TABLE 28
Reviews: movie electronics. Contingency table between the clusters and true topic categories.

	movie	electronics
1: film, movie, story, films, man, character,. . .	**2000**	21
2: bought, product, quality, buy, unit, price,. . .	0	**1979**

TABLE 29
Reviews: kitchen electronics. Contingency table between the clusters and true topic categories.

	kitchen	electronics
1: pan, knife, sheets, kitchen, oven, dishes,. . .	**1059**	38
2: water, coffee, maker, ice, cup, machine,. . .	**671**	42
3: card, camera, printer, work, windows, hp,. . .	175	**883**
4: sound, ipod, phone, headphones, ear, dvd,. . .	95	**1037**

TABLE 30
Reviews: dvd electronics. Contingency table between the clusters and true topic categories.

	dvd	electronics
1: movie, film, story, acting, man,. . .	**1416**	7
2: dvd, season, video, collection, show,. . .	**533**	33
3: ipod, price, sound, quality, headphones,. . .	32	**981**
4: phone, card, mouse, keyboard, wireless,. . .	19	**979**

TABLE 31
Reviews: movie books electronics kitchen. Contingency table between the clusters and true topic categories.

	movie	books	electronics	kitchen
1: film, movie, films, director, scene,. . .	**1999**	182	9	15
2: book, read, books, reading, author,. . .	1	**1776**	11	9
3: ipod, card, sound, phone, battery,. . .	0	19	**1923**	313
4: clean, water, pan, kitchen, coffee,. . .	0	23	57	**1663**

TABLE 32
Reviews: dvd movie electronics kitchen. Contingency table between the clusters and true topic categories.

	dvd	movie	electronics	kitchen
1: movie, film, story, characters, films,. . .	1754	**2000**	12	17
2: sound, ipod, card, software, battery,. . .	224	0	**1934**	321
3: water, pan, coffee, kitchen, clean, sheets,. . .	22	0	54	**1662**

TABLE 33
Reviews: dvd books movie kitchen. Contingency table between the clusters and true topic categories.

	dvd	books	movie	kitchen
1: book, read, books, reading, author, reader,. . .	70	**1792**	1	10
2: film, movie, films, director, movies,. . .	1834	178	1999	21
3: product, clean, kitchen, plastic, pan,. . .	96	30	0	**1969**

TABLE 34
Reviews: dvd books electronics movie. Contingency table between the clusters and true topic categories.

	dvd	books	electronics	movie
1: book, read, books, author, reading,. . .	67	**1800**	10	1
2: film, movie, films, director, comedy, cast,. . .	1699	175	17	**1999**
3: product, quality, unit, ipod, bought,. . .	234	25	**1973**	0

TABLE 35
Reviews: dvd books electronics movie kitchen. Contingency table between the clusters and true topic categories.

	dvd	books	electronics	movie	kitchen
1: book, read, books, reading, author,. . .	63	**1775**	11	1	8
2: film, movie, films, director, comedy,. . .	1685	178	11	**1999**	16
3: sound, ipod, card, battery, phone, product,. . .	232	21	**1927**	0	331
4: water, pan, kitchen, coffee, clean, handle,. . .	20	26	51	0	**1645**

TABLE 36
20-News: med guns. Contingency table between the clusters and true topic categories.

	med	guns
1: msg, food, dyer, photography, stuff, uga,. . .	**229**	5
2: banks, medicine, treatment, thanks,. . .	**730**	17
3: gun, guns, killed, weapons, police,. . .	23	**597**
4: cdt, fbi, brent, fire, gas, handheld,. . .	5	**290**

TABLE 37
20-News: autos space hardware. Contingency table between the clusters and true topic categories.

	pc.hardware	autos	space
1: thanks, card, pc, please, dos, monitor,. . .	**941**	38	34
2: car, cars, engine, ford, oil, insurance,. . .	28	**926**	13
3: jbh, mccall, nick, sky, calpoly, bby, wats,. . .	8	22	**419**
4: nasa, space, henry, shuttle, spacecraft,. . .	2	1	**519**

TABLE 38
20-News: comp. Contingency table between the clusters and true topic categories.

	graphics	ms-windows	pc.hardware	mac.hardware	windows.x
1: ac, navy, group, uk, graphics, comp, xv,. . .	**126**	25	12	11	49
2: windows, andre, zip, cica, file, files,. . .	85	**209**	24	6	55
3: windows, dos, win, it, mouse, crash, ms,. . .	20	**386**	102	15	28
4: thanks, please, firstclass, mail, email,. . .	266	72	100	108	95
5: card, bus, diamond, dx, video, eisa, vesa,. . .	65	97	**232**	20	13
6: image, points, objects, sphere, science,. . .	**263**	22	12	13	178
7: apple, they, edu, life, power, cnn, hacker,. . .	96	77	133	**397**	25
8: drive, scsi, drives, disk, hd, hard,. . .	0	23	**237**	126	1
9: monitor, simms, cpu, iisi, clock, nubus,. . .	26	9	118	**253**	7
10: window, server, xterm, widget, lib, int,. . .	23	43	9	9	**531**

TABLE 39
20-News: rec. Contingency table between the clusters and true topic categories.

	autos	motorcycles	baseball	hockey
1: car, cars, thanks, engine, ford, eliot,. . .	**723**	26	41	13
2: my, shop, fit, bmw, police, interstate,. . .	221	**284**	16	13
3: dod, com, uk, ranck, winona, shaft, nec,. . .	14	**668**	7	12
4: he, baseball, runs, ball, pitching, sherri,. . .	9	9	**802**	181
5: gm, hockey, rochester, mailing, baltimore,. . .	18	5	74	**141**
6: det, bos, la, nyi, pit, stl, nyr, sj,. . .	0	0	5	**46**
7: buffalo, detroit, montreal, go, pittsburgh,. . .	2	1	46	**591**

TABLE 40
20-News: sci. Contingency table between the clusters and true topic categories.

	crypto	electronics	med	space
1: key, clipper, encryption, government, nsa,. . .	**834**	16	3	4
2: radio, battery, car, michael, hp, plastic,. . .	10	**371**	46	26
3: thanks, anyone, mb, mail, appreciated,. . .	90	**560**	138	85
4: geb, food, she, doctor, medicine, disease,. . .	50	16	**792**	39
5: space, nasa, henry, launch, orbit, earth,. . .	5	21	8	**831**

TABLE 41
20-News: mix2. Contingency table between the clusters and true topic categories.

	ms-windows	autos	baseball	med	space
1: windows, file, dos, os, printer, files,. . .	**825**	5	1	3	5
2: car, cars, engine, dealer, oil, ford,. . .	18	**878**	5	11	14
3: york, mail, philadelphia, thanks, thank,. . .	82	65	131	107	50
4: baseball, he, game, players, team, ball,. . .	10	12	**841**	18	14
5: geb, doctor, medicine, medical, treatment,. . .	17	21	10	**829**	25
6: space, pat, nasa, sky, earth, orbit,. . .	11	6	3	19	**877**

TABLE 42
20-News: mix. Contingency table between the clusters and true topic categories.

	ms-windows	forsale	baseball	space	politics.misc
1: windows, file, dos, files, driver, using,. . .	**878**	33	17	15	6
2: sale, shipping, offer, condition, asking,. . .	26	**865**	17	14	4
3: baseball, he, players, team, ball, pitcher,. . .	7	16	**931**	13	21
4: space, nasa, pat, earth, orbit, launch,. . .	12	12	0	**876**	44
5: people, kaldis, com, fbi, clinton, drugs,. . .	40	38	26	67	**699**

TABLE 43
20-News: sci comp. Contingency table between the clusters and true topic categories.

	graphics	ms-windows	pc.hardware	mac.hardware	windows.x	crypto	electronics	med	space
1: windows, dos, win, ms, file, os, exe,. . .	62	**564**	112	9	64	11	11	2	1
2: ftp, fax, mail, de, domain, visualization,. . .	380	117	33	24	216	29	34	15	50
3: card, mhz, dx, video, board, cards, lc,. . .	75	77	307	268	16	1	53	1	0
4: drive, scsi, drives, disk, controller, hd,. . .	2	31	**263**	168	1	1	34	0	1
5: window, widget, server, xterm, int, lib,. . .	110	70	15	29	**592**	5	14	3	0
6: thanks, anyone, please, newsgroup, mail,. . .	204	39	121	171	44	23	247	112	19
7: tube, solder, power, led, pulse, ink,. . .	68	21	95	241	12	31	369	11	19
8: clipper, key, encryption, sternlight,. . .	3	3	9	5	7	**817**	33	2	4
9: geb, doctor, food, covington, article, she,. . .	50	34	19	37	15	64	138	**830**	77
10: space, nasa, henry, orbit, earth, launch,. . .	16	7	5	6	15	7	51	11	**814**

TABLE 44
20-News: sci comp rec. Contingency table between the clusters and true topic categories.

	graphics	ms-windows	pc.hardware	mac.hardware	windows.x	crypto	electronics	med	space	autos	motorcycles	baseball	hockey
1: ftp, fax, format, graphics, package,. . .	341	138	47	33	277	18	21	7	10	1	0	1	1
2: windows, window, abpsoft, file, default,. . .	115	565	84	39	491	9	24	3	0	4	1	0	0
3: drive, mb, scsi, mhz, card, bus, cpu,. . .	58	100	586	508	15	4	108	0	2	1	1	0	0
4: thanks, anyone, please, mail, looking,. . .	254	79	169	229	72	27	263	125	39	92	21	61	26
5: key, clipper, encryption, sternlight, nsa,. . .	10	1	10	12	16	**843**	45	3	3	3	0	1	1
6: canada, sweden, usa, april, det, uk,. . .	71	23	23	26	66	18	92	40	18	38	23	14	81
7: geb, amp, medicine, food, cause, that,. . .	65	28	41	68	14	41	260	639	85	168	44	29	8
8: nasa, space, orbit, henry, launch, pat,. . .	42	11	10	13	23	3	92	73	**810**	12	4	1	1
9: dod, car, bike, article, engine, cars, com,. . .	5	10	7	13	4	14	72	46	7	646	**873**	13	8
10: he, his, baseball, him, pitcher, ball,. . .	7	5	2	12	4	12	5	50	10	21	25	**657**	221
11: philadelphia, go, montreal, detroit,. . .	2	3	0	5	0	0	2	1	1	1	1	214	**650**

TABLE 45
20-News (all). Contingency table between the clusters and true topic categories.

	alt.atheism	comp.graphics	comp.os.ms-windows.misc	comp.sys.ibm.pc.hardware	comp.sys.mac.hardware	comp.windows.x	misc.forsale	rec.autos	rec.motorcycles	rec.sport.baseball	rec.sport.hockey	sci.crypt	sci.electronics	sci.med	sci.space	soc.religion.christian	talk.politics.guns	talk.politics.mideast	talk.politics.misc	talk.religion.misc
1: windows, file, window, files, motif,. . .	5	507	738	105	47	828	21	8	4	9	1	30	57	11	11	7	3	0	3	4
2: card, monitor, video, mhz, monitors, cpu,. . .	1	137	137	419	514	21	75	4	1	1	0	15	103	0	0	0	8	1	0	0
3: drive, scsi, drives, disk, hd, ide, rom,. . .	0	6	39	318	182	1	84	10	12	2	0	2	55	0	4	2	0	0	1	1
4: sale, mail, email, price, offer, vhs, ink,. . .	3	103	22	96	156	48	715	49	22	27	5	21	282	54	15	12	11	4	8	6
5: dod, car, cars, bike, engine, bmw, riding,. . .	6	2	7	4	15	4	39	864	918	10	2	0	145	37	6	2	8	4	7	5
6: team, season, game, baseball, nhl, hockey,. . .	2	4	2	3	5	1	14	2	3	900	**980**	0	13	4	1	2	3	1	5	2
7: clipper, key, chip, algorithm, encryption,. . .	1	26	2	11	14	22	3	1	0	1	0	**867**	82	3	5	0	7	6	9	4
8: god, atheists, jesus, bible, christian,. . .	566	2	2	0	0	1	1	1	5	2	2	0	5	13	3	903	2	6	13	397
9: space, nasa, orbit, moon, distance,. . .	9	152	11	17	15	49	9	8	6	5	6	7	200	195	**890**	5	3	3	14	8
10: geb, cramer, article, doctor, study,. . .	114	31	4	7	12	9	12	19	15	23	2	12	35	658	43	35	29	32	333	84
11: israel, turkish, israeli, turks, muslim,. . .	63	3	0	0	1	1	0	0	0	11	0	1	2	1	1	20	11	**868**	32	19
12: gun, atf, waco, cdt, guns, firearms, fire,. . .	29	0	1	2	2	0	2	23	10	3	1	36	5	14	8	8	**824**	15	350	98

TABLE 46
Reuters. Contingency table between the clusters and true topic categories.

	earn	acq	trade	ship	grain	crude	interest	money-fx
1: split, declared, date, dividend,. . .	**222**	7	0	1	24	0	1	1
2: div, record, pay, payout, april, qtly,. . .	**490**	0	0	0	73	0	0	0
3: loss, shr, net, qtr, revs, note, cts, oper,. . .	**1406**	1	0	1	746	0	0	0
4: sell, unit, company, terms, undisclosed,. . .	253	**752**	1	354	43	2	0	0
5: offer, bid, stake, takeover, shares,. . .	40	**687**	0	303	10	1	0	0
6: trade, grain, dispute, agriculture,. . .	71	42	**228**	56	29	5	1	9
7: stg, billion, fed, turnover, francs,. . .	365	37	30	23	176	15	27	80
8: rate, bank, pct, banks, rates, prime,. . .	39	44	0	13	12	11	**212**	32
9: oil, crude, prices, sea, barrels, refinery,. . .	26	26	17	51	9	244	0	120
10: monetary, baker, dollar, currency,. . .	3	10	11	2	2	62	30	85

TABLE 47
Recipe. Contingency table between the clusters and true topic categories.

	greek	southern us	filipino	indian	jamaican	spanish	italian	mexican	chinese	british	thai	vietnamese	cajun creole	brazilian	french	japanese	irish	korean	moroccan	russian
1: seed(s), masala, dal, cumin seed(s),. . .	7	17	4	**1712**	7	2	12	87	23	10	47	12	5	10	8	91	3	8	58	5
2: part skim, mozzarella, cheese, italian,. . .	49	13	1	4	0	8	**1331**	30	1	3	2	0	9	5	21	3	2	0	3	2
3: self, vegetable oil, rising, eggs,. . .	22	418	27	38	15	49	137	75	38	97	9	5	45	20	66	41	31	13	0	41
4: all purpose, large eggs, baking powder,. . .	36	893	21	58	19	14	261	149	31	168	5	6	47	6	183	39	161	15	8	68
5: cheese, parmesan, bread slices, grated,. . .	13	76	1	3	2	6	**835**	16	2	7	1	1	23	12	79	5	1	0	0	2
6: orange(s), brown sugar, sugar(s), tea(s),. . .	45	878	95	121	76	75	266	189	72	132	37	20	57	154	262	61	91	18	48	44
7: red pepper(s), flakes, vinegar, red wine,. . .	83	122	10	35	15	45	612	99	61	3	43	19	51	12	28	42	3	54	11	7
8: tortilla(s), cheddar cheese, beans, taco,. . .	7	160	4	3	2	3	55	**1940**	9	12	0	2	14	2	18	6	15	5	0	7
9: jalapeno(s), best, lime, lime juice,. . .	4	55	5	66	20	27	22	**1787**	5	0	26	21	12	21	6	6	4	5	10	1
10: extra virgin olive oil, olive oil,. . .	164	69	4	35	6	218	**1145**	129	11	13	4	5	35	8	209	24	13	5	94	5
11: lemon, lemon juice, yogurt, horseradish,. . .	362	170	10	155	3	37	402	74	13	29	21	10	96	7	207	45	23	5	143	33
12: skinless, chicken, boneless, breast(s),. . .	61	106	28	163	29	23	361	474	290	2	170	41	96	11	55	47	4	7	41	8
13: black pepper, pork, ground, cider, cayenne,. . .	95	247	83	142	142	53	270	416	35	32	3	30	144	21	80	18	16	7	217	16
14: fat, less, fat free, low,. . .	59	133	3	122	4	51	411	226	60	11	32	14	49	10	155	20	42	4	54	8
15: bell pepper(s), green(s), celery, long,. . .	57	224	50	60	49	140	341	397	97	2	82	16	562	50	58	35	5	14	39	15
16: yeast, warm water, water, active dry yeast,. . .	17	39	19	97	16	9	230	53	36	37	1	9	44	6	67	20	6	4	22	51
17: white, grain, wine, leaf parsley, broth,. . .	45	156	37	24	18	111	619	142	44	36	14	13	73	31	400	41	12	10	19	28
18: potato(es), half & half, bay, butter,. . .	20	334	97	62	65	57	162	56	20	138	4	20	153	26	269	41	205	19	42	103
19: egg, yolk(s), puff, egg(s), large egg(s),. . .	28	167	19	7	3	56	340	64	39	69	5	11	26	20	467	47	29	5	7	45
20: lime, fish sauce, coconut milk, fish, thai,. . .	1	18	99	87	27	4	19	23	58	1	**898**	440	2	33	3	20	1	21	5	0
21: soy sauce, sesame, five, gochujang, bean,. . .	0	25	138	9	8	1	7	12	1728	2	135	130	3	2	5	771	0	611	0	0

TABLE 48
NSF. Contingency table between the clusters and true topic categories.

	DMS	DUE	EAR	DMI	OCE	SES	CHE	IBN	INT	DMR	DEB	BCS	MCB	ATM	CCR	CMS	DBI	CTS	PHY	ECS	OPP	ESI	IIS	BES	AST	EIA	ANI	HRD	EEC	REC	ACI	OIA	DGE	ESR	EPS
1: theory, conjecture, geometry, spaces,. . .	**4044**	12	0	12	0	25	0	1	85	2	0	10	0	2	127	9	8	0	16	10	0	0	25	1	0	3	0	9	0	1	0	0	0	0	0
2: mantle, rocks, magma, tectonic, magmatic,. . .	5	7	**4039**	2	1051	1	3	1	132	7	9	34	0	2	1	35	0	0	0	1	180	1	1	10	0	0	0	9	0	1	0	0	0	0	0
3: laboratory, courses, analysis, students,. . .	28	**5135**	122	50	10	9	43	8	0	50	9	22	20	50	13	19	251	28	27	16	0	31	14	25	5	82	0	25	62	79	0	58	8	0	2
4: equations, solutions, problems, nonlinear,. . .	**2896**	42	56	82	23	67	30	3	146	87	5	5	3	34	180	204	8	197	77	136	0	1	24	33	2	9	4	15	12	2	75	0	0	0	0
5: reactions, molecules, professor, complexes,. . .	5	145	80	45	23	7	**4306**	8	323	179	2	1	196	98	1	10	90	369	31	4	2	3	1	68	8	1	0	26	2	2	5	17	3	1	4
6: dissertation, political, interviews,. . .	184	37	22	82	12	**4897**	0	45	124	0	52	2674	4	20	10	105	84	2	3	9	153	45	143	12	3	89	10	45	4	59	2	8	12	4	3
7: carbon, water, waters, concentrations,. . .	10	23	600	63	2357	18	21	162	116	8	675	82	78	526	3	40	87	62	3	2	493	7	0	218	5	1	0	16	9	0	5	4	3	0	1
8: feasibility, phase, conventional, high,. . .	10	43	64	**3035**	39	6	31	4	13	120	2	4	5	54	58	188	125	359	37	626	6	3	18	154	30	17	6	8	53	0	2	2	0	3	6
9: species, populations, predators, fossil,. . .	52	20	206	29	434	44	5	1019	272	3	**3641**	561	112	11	2	1	382	6	5	1	135	14	3	13	0	5	0	16	0	1	1	9	6	0	3
10: brain, cells, neurons, nerve, perception,. . .	82	28	6	44	35	39	22	**2589**	119	3	19	405	425	0	16	14	222	30	5	28	13	14	74	311	0	36	0	29	3	7	2	5	1	0	1
11: support, cooperative, university, dr,. . .	426	31	448	33	924	51	41	19	1377	52	87	107	11	100	6	151	505	73	51	59	74	46	16	79	8	104	10	11	318	7	11	94	6	18	45
12: materials, properties, films, alloys,. . .	88	82	158	507	4	9	395	2	368	3257	4	18	22	1	6	544	26	637	22	570	2	12	2	71	1	6	0	35	71	0	6	12	3	1	14
13: protein, mutants, proteins, gene, mutant,. . .	44	48	10	106	38	6	71	1205	145	22	150	12	**3258**	0	13	0	648	52	7	7	19	6	15	283	0	32	1	31	7	0	7	3	5	0	1
14: model, measurements, flow, ionosphere,. . .	150	13	266	45	478	3	11	9	60	77	19	31	19	1778	4	183	18	569	76	64	70	2	3	64	38	1	1	8	3	1	5	1	0	0	1
15: workshop, conference, held, meeting,. . .	557	282	189	163	118	231	100	286	749	283	135	159	361	189	191	451	127	350	254	300	112	92	214	210	49	105	91	38	41	40	47	10	2	21	12
16: earthquake, design, control, loads,. . .	97	296	228	1314	9	100	1	7	228	13	1	11	1	8	216	1823	4	222	5	542	4	7	183	221	1	58	11	20	249	15	14	4	8	3	5
17: quantum, theoretical, physics, atoms,. . .	172	85	37	11	3	35	222	7	543	967	1	0	29	76	19	18	22	64	1878	176	4	10	5	3	14	29	0	18	4	1	19	10	3	0	2
18: ice, climate, records, record, glacier,. . .	15	4	444	11	197	6	14	5	76	5	53	158	3	660	1	27	7	10	6	2	1437	1	1	2	6	0	0	7	0	0	0	0	1	0	1
19: algorithms, techniques, languages,. . .	481	174	24	562	17	69	27	7	190	20	14	292	13	22	2909	79	164	27	19	235	3	5	1362	53	13	614	55	41	14	32	450	1	5	0	1
20: program, reu, graduate, students,. . .	261	870	102	34	120	159	328	17	90	170	18	58	20	49	13	60	448	55	206	48	13	345	13	119	35	252	2	628	480	41	4	225	188	26	32
21: teachers, teacher, mathematics, science,. . .	38	1085	41	85	23	71	7	5	13	21	7	31	2	10	2	0	6	5	18	5	7	1986	15	5	6	53	2	406	44	543	5	6	126	238	4
22: galaxies, stars, observations, telescope,. . .	20	58	21	18	3	12	8	1	179	4	0	2	1	480	2	0	15	6	299	9	87	24	6	6	**1620**	0	1	18	2	1	2	2	1	0	0
23: network, internet, access, services,. . .	40	187	38	173	20	65	5	5	41	6	37	19	5	42	355	40	129	7	11	121	20	15	176	7	4	233	1423	15	29	36	53	2	1	6	14

TABLE 49
AG's news. Contingency table between the clusters and true topic categories.

	Business	Sci/Tech	Sports	World
1: sales, quarter, profit, earnings, percent,. . .	**3874**	558	24	102
2: company, drug, executive, motors,. . .	**4236**	657	68	100
3: bankruptcy, million, airlines, pay,. . .	**4121**	358	320	266
4: oil, prices, crude, price, fuel, opec, gas,. . .	**3436**	116	19	429
5: reuters, dollar, rates, rate, jobless,. . .	**3509**	480	35	344
6: oracle, peoplesoft, bid, takeover, offer,. . .	**776**	607	19	17
7: wireless, phone, mobile, internet, fi,. . .	1030	**3832**	35	82
8: search, google, web, engine, yahoo, ipo,. . .	425	**1742**	45	65
9: ibm, apple, dvd, hp, pc, computer, format,. . .	811	**6172**	30	97
10: microsoft, software, windows, security,. . .	414	**4469**	7	33
11: gt, lt, fullquote, http, investor,. . .	2236	1359	581	735
12: space, nasa, earth, scientists, mars,. . .	80	**2775**	11	107
13: court, federal, drug, trial, law, pinochet,. . .	2032	1870	574	1922
14: quot, writes, coach, stewart, halo,. . .	897	1060	1033	936
15: china, japan, scientists, researchers,. . .	785	1330	182	2132
16: gold, olympic, red, athens, sox, yankees,. . .	327	294	**4302**	384
17: season, quarterback, game, sports, yards,. . .	282	635	**10260**	246
18: cup, open, grand, round, masters, golf,. . .	117	320	**5233**	276
19: striker, manchester, arsenal, league,. . .	113	137	**4405**	237
20: ap, coach, quarterback, california, helens,. . .	209	1466	2937	1949
21: india, test, australia, cricket, minister,. . .	436	240	1259	1875
22: president, bush, john, election,. . .	731	576	139	**3883**
23: sudan, nuclear, darfur, talks, iran,. . .	613	297	110	**4145**
24: people, police, hurricane, killed, bomb,. . .	290	470	207	**4280**
25: iraq, iraqi, british, al, hostage, najaf,. . .	103	62	60	**4643**
26: palestinian, arafat, gaza, israeli, israel,. . .	17	18	5	**2615**

Appendix 2 Other Datasets

In order to test the generality of descriptive clustering, we applied the PDC algorithm on three datasets that were not formed from text documents.

Appendix 2.1 Mushrooms

We applied the descriptive clustering algorithm to the Mushroom dataset from the UCI Machine Learning Repository [5]. To do this we converted the categorical features into 112 binary features. There are 8,124 instances in this dataset. When applying the descriptive clustering algorithm, the algorithm selected 26 clusters, which is the upper limit of the range used, so we increased the range of cluster sizes to 50, and the algorithm selected 47 clusters. The selected features for each cluster are in Table 50. Interestingly, the clusters were very pure in terms of edible and poisonous, with only 4 of the 47 clusters having a mixture of both edible and poisonous instances as shown in Table 51. However, using only the selected features, the predictive classifier makes some mistakes; there are poisonous mushrooms that the classifier assigns to an edible cluster. Furthermore, the list of features is not sufficient to make the cluster assignment since the classifier is using a weighted combination of features to make the cluster assignment. Nonetheless, the descriptive clustering is able to succinctly describe many of the clusters using only a few features.

TABLE 50
Mushroom. Descriptive features selected for each cluster sorted by coefficient value.

1: population=numerous, habitat=grasses, odor=almond, ring-number=two, gill-spacing=crowded, spore-print-color=white, cap-shape=bell, cap-shape=knobbed, odor=none, stalk-surface-below-ring=silky, stalk-surface-above-ring=silky, gill-color=black, gill-color=gray, cap-color=gray, gill-color=pink, gill-color=white, cap-color=yellow, stalk-surface-above-ring=smooth, stalk-surface-below-ring=smooth
2: gill-color=black, population=scattered, ring-type=evanescent, habitat=grasses
3: gill-color=chocolate, population=abundant
4: stalk-color-below-ring=gray, stalk-color-above-ring=pink
5: gill-color=purple, stalk-color-below-ring=pink, bruises?=bruises, odor=none, ring-type=pendant, stalk-shape=tapering, stalk-color-above-ring=gray, stalk-surface-below-ring=smooth, gill-size=broad, population=solitary, gill-spacing=close
6: stalk-color-above-ring=orange
7: population=abundant, gill-color=pink, gill-color=brown, gill-color=black
8: stalk-color-above-ring=gray, gill-color=pink, gill-color=white, gill-color=brown, stalk-color-below-ring=pink
9: habitat=waste
10: gill-color=brown, habitat=woods, cap-color=brown
11: gill-color=pink, spore-print-color=black, habitat=woods
12: cap-surface=fibrous, habitat=urban, cap-shape=sunken, gill-size=narrow, bruises?=no, stalk-shape=enlarging
13: gill-color=white, habitat=woods, odor=none, bruises?=bruises, stalk-shape=tapering, stalk-color-above-ring=pink, stalk-color-below-ring=pink, stalk-surface-below-ring=smooth
14: habitat=meadows, odor=anise, odor=almond, spore-print-color=black, spore-print-color=brown, population=scattered, ring-number=one
15: gill-spacing=crowded, odor=almond, odor=anise, gill-size=narrow, spore-print-color=purple, habitat=woods, bruises?=bruises
16: population=scattered, stalk-surface-above-ring=smooth, ring-number=two, gill-spacing=crowded, gill-color=chocolate, gill-color=brown, spore-print-color=white, habitat=grasses, odor=none, gill-color=pink, stalk-surface-below-ring=fibrous, gill-color=gray, bruises?=no, cap-shape=bell, cap-color=white, ring-type=evanescent, gill-size=broad, cap-color=gray, ring-type=pendant, spore-print-color=black, cap-surface=smooth, spore-print-color=brown
17: cap-color=gray, gill-color=brown, habitat=woods
18: cap-color=red, gill-color=brown, gill-color=pink, spore-print-color=brown, bruises?=bruises, habitat=woods, odor=none, ring-type=pendant, stalk-surface-below-ring=smooth
19: gill-color=pink, spore-print-color=brown, habitat=woods
20: habitat=grasses, odor=anise, stalk-surface-below-ring=smooth, cap-shape=bell, population=numerous, stalk-shape=enlarging, gill-spacing=close
21: stalk-color-above-ring=gray, stalk-color-below-ring=gray
22: gill-color=white, stalk-shape=enlarging, spore-print-color=white, cap-color=cinnamon, habitat=leaves, cap-surface=grooves, odor=none, stalk-color-below-ring=brown, stalk-color-above-ring=white, gill-spacing=crowded
23: stalk-surface-below-ring=scaly
24: ring-type=flaring, stalk-surface-above-ring=fibrous, habitat=grasses, population=scattered, stalk-surface-below-ring=fibrous
25: spore-print-color=green, cap-color=pink, bruises?=bruises, odor=almond, habitat=grasses, population=scattered, stalk-surface-below-ring=smooth, stalk-shape=enlarging, cap-shape=bell, gill-size=broad, stalk-color-below-ring=white, stalk-color-above-ring=white
26: odor=spicy, habitat=woods, cap-shape=convex, cap-surface=smooth, cap-shape=flat, stalk-color-below-ring=pink, habitat=paths, cap-color=brown, cap-color=red
27: stalk-color-below-ring=pink, stalk-color-above-ring=brown
28: odor=foul, habitat=leaves
29: stalk-color-below-ring=buff, stalk-color-above-ring=buff, stalk-color-above-ring=pink, gill-color=chocolate
30: habitat=urban, odor=pungent
31: odor=musty
32: cap-shape=knobbed, odor=spicy, habitat=woods, habitat=paths
33: stalk-color-below-ring=brown, gill-color=pink, stalk-color-above-ring=pink, gill-color=chocolate, stalk-color-above-ring=buff, ring-type=large
34: odor=fishy, habitat=woods, cap-shape=convex, cap-shape=flat, habitat=paths
35: odor=foul, habitat=woods, gill-color=buff, habitat=paths
36: habitat=leaves, odor=fishy
37: odor=spicy, habitat=woods, cap-shape=convex, cap-surface=scaly, cap-shape=flat, habitat=paths, cap-color=red, stalk-color-below-ring=white, stalk-color-below-ring=pink
38: habitat=leaves, odor=spicy
39: habitat=grasses, odor=pungent
40: spore-print-color=chocolate, cap-surface=smooth, cap-color=buff, habitat=urban, bruises?=bruises, stalk-shape=tapering, stalk-surface-below-ring=fibrous, stalk-color-below-ring=white, stalk-surface-above-ring=fibrous
41: odor=creosote
42: cap-shape=knobbed, odor=fishy, habitat=woods, habitat=paths
43: stalk-color-below-ring=buff, stalk-color-above-ring=brown, gill-color=chocolate, stalk-color-above-ring=pink, gill-color=pink
44: stalk-color-below-ring=pink, ring-type=large, stalk-color-above-ring=buff, cap-color=yellow
45: gill-color=gray, stalk-color-below-ring=brown, stalk-color-above-ring=pink, stalk-color-above-ring=buff, ring-type=large
46: stalk-color-below-ring=brown, stalk-color-above-ring=brown
47: ring-type=large, cap-color=gray, stalk-color-below-ring=pink, stalk-color-above-ring=pink

TABLE 51
Mushroom. Contingency table between the clusters and true topic categories.

	edible	poisonous
1: population=numerous, habitat=grasses,. . .	**208**	0
2: gill-color=black, population=scattered,. . .	**120**	0
3: gill-color=chocolate, population=abundant	**96**	0
4: stalk-color-below-ring=gray,. . .	**384**	0
5: gill-color=purple,. . .	**288**	0
6: stalk-color-above-ring=orange	**192**	0
7: population=abundant, gill-color=pink,. . .	**288**	0
8: stalk-color-above-ring=gray,. . .	**288**	0
9: habitat=waste	**192**	0
10: gill-color=brown, habitat=woods,. . .	**80**	0
11: gill-color=pink, spore-print-color=black,. . .	**96**	0
12: cap-surface=fibrous, habitat=urban,. . .	**96**	0
13: gill-color=white, habitat=woods, odor=none,. . .	**192**	0
14: habitat=meadows, odor=anise, odor=almond,. . .	**256**	0
15: gill-spacing=crowded, odor=almond,. . .	**96**	0
16: population=scattered,. . .	**288**	0
17: cap-color=gray, gill-color=brown,. . .	**64**	0
18: cap-color=red, gill-color=brown,. . .	**80**	0
19: gill-color=pink, spore-print-color=brown,. . .	**64**	0
20: habitat=grasses, odor=anise,. . .	**120**	0
21: stalk-color-above-ring=gray,. . .	**192**	0
22: gill-color=white, stalk-shape=enlarging,. . .	**72**	8
23: stalk-surface-below-ring=scaly	**208**	40
24: ring-type=flaring,. . .	**192**	48
25: spore-print-color=green, cap-color=pink,. . .	56	**72**
26: odor=spicy, habitat=woods,. . .	0	**128**
27: stalk-color-below-ring=pink,. . .	0	**144**
28: odor=foul, habitat=leaves	0	**192**
29: stalk-color-below-ring=buff,. . .	0	**240**
30: habitat=urban, odor=pungent	0	**128**
31: odor=musty	0	**36**
32: cap-shape=knobbed, odor=spicy,. . .	0	**128**
33: stalk-color-below-ring=brown,. . .	0	**192**
34: odor=fishy, habitat=woods,. . .	0	**256**
35: odor=foul, habitat=woods, gill-color=buff,. . .	0	**384**
36: habitat=leaves, odor=fishy	0	**192**
37: odor=spicy, habitat=woods,. . .	0	**128**
38: habitat=leaves, odor=spicy	0	**192**
39: habitat=grasses, odor=pungent	0	**128**
40: spore-print-color=chocolate,. . .	0	**240**
41: odor=creosote	0	**192**
42: cap-shape=knobbed, odor=fishy,. . .	0	**128**
43: stalk-color-below-ring=buff,. . .	0	**192**
44: stalk-color-below-ring=pink,. . .	0	**216**
45: gill-color=gray,. . .	0	**96**
46: stalk-color-below-ring=brown,. . .	0	**144**
47: ring-type=large, cap-color=gray,. . .	0	**72**

TABLE 52
Mushroom. Contingency table between predicted clusters and true topic categories.

	edible	poisonous
1: population=numerous, habitat=grasses,. . .	**214**	0
2: gill-color=black, population=scattered,. . .	**112**	10
3: gill-color=chocolate, population=abundant	**97**	2
4: stalk-color-below-ring=gray,. . .	**379**	0
5: gill-color=purple,. . .	**283**	0
6: stalk-color-above-ring=orange	**192**	0
7: population=abundant, gill-color=pink,. . .	**293**	0
8: stalk-color-above-ring=gray,. . .	**275**	0
9: habitat=waste	**191**	0
10: gill-color=brown, habitat=woods,. . .	**80**	0
11: gill-color=pink, spore-print-color=black,. . .	**82**	0
12: cap-surface=fibrous, habitat=urban,. . .	**96**	21
13: gill-color=white, habitat=woods, odor=none,. . .	**201**	0
14: habitat=meadows, odor=anise, odor=almond,. . .	**248**	0
15: gill-spacing=crowded, odor=almond,. . .	**93**	0
16: population=scattered,. . .	**270**	0
17: cap-color=gray, gill-color=brown,. . .	**80**	0
18: cap-color=red, gill-color=brown,. . .	**61**	0
19: gill-color=pink, spore-print-color=brown,. . .	**82**	0
20: habitat=grasses, odor=anise,. . .	**143**	0
21: stalk-color-above-ring=gray,. . .	**195**	0
22: gill-color=white, stalk-shape=enlarging,. . .	**60**	8
23: stalk-surface-below-ring=scaly	**208**	33
24: ring-type=flaring,. . .	**184**	24
25: spore-print-color=green, cap-color=pink,. . .	69	**72**
26: odor=spicy, habitat=woods,. . .	0	**149**
27: stalk-color-below-ring=pink,. . .	0	**154**
28: odor=foul, habitat=leaves	0	**189**
29: stalk-color-below-ring=buff,. . .	0	**233**
30: habitat=urban, odor=pungent	0	**126**
31: odor=musty	0	**36**
32: cap-shape=knobbed, odor=spicy,. . .	0	**134**
33: stalk-color-below-ring=brown,. . .	0	**187**
34: odor=fishy, habitat=woods,. . .	0	**268**
35: odor=foul, habitat=woods, gill-color=buff,. . .	0	**370**
36: habitat=leaves, odor=fishy	0	**196**
37: odor=spicy, habitat=woods,. . .	0	**153**
38: habitat=leaves, odor=spicy	0	**196**
39: habitat=grasses, odor=pungent	0	**125**
40: spore-print-color=chocolate,. . .	0	**266**
41: odor=creosote	0	**191**
42: cap-shape=knobbed, odor=fishy,. . .	0	**134**
43: stalk-color-below-ring=buff,. . .	0	**179**
44: stalk-color-below-ring=pink,. . .	0	**217**
45: gill-color=gray,. . .	0	**85**
46: stalk-color-below-ring=brown,. . .	2	**144**
47: ring-type=large, cap-color=gray,. . .	4	**78**

Appendix 2.2 USDA Plant database

The USDA plant database [6] from the UCI Machine Learning Repository [5] contains a list of 34,781 plant varieties that occur across 70 locations in the USA, Canada, and surrounding islands. We applied the descriptive clustering to the locations and use the plants as the features. The algorithm found 3 clusters of locations each described by a single species of plant. The clusters of locations are given in Table 53. Figure 1 shows the layout of most of each cluster's locations: Greenland (DENGL), Puerto Rico (PR), and the Saint Pierre and Miquelon (FRASPM) are not shown. The distributions for each of the plants selected to describe the clusters are shown in Figure 2. Most of the cluster locations match the corresponding plant distributions. Only four US states (IA, MN, WI, MI) and one Canadian province (YK) are not covered by the selected plants.

TABLE 53
Plants. Correspondence between the clusters (with a plant species as a descriptive feature) and locations.

1: *elephantopus*	AL, AR, DC, DE, FL, GA, HI, IA, IL, IN, KS, KY, LA, MD, MO, MS, NC, NJ, OH, OK, PA, PR, SC, TN, TX, VA, VI, WV
2: *crepis occidentalis*	AB, AZ, BC, CA, CO, ID, MT, ND, NE, NM, NV, OR, SD, SK, UT, WA, WY
3: *puccinellia tenella*	AK, CT, DENGL, FRASPM, GL, LB, MA, MB, ME, MI, MN, NB, NF, NH, NS, NT, NU, NY, ON, PE, QC, RI, VT, WI, YT

Fig. 1. Locations colored by cluster membership (some island locations are missing from the map).

Fig. 2. Distribution of plants selected as descriptors of location clusters. (Left to right) *elephantopus*, *crepis occidentalis*, *puccinellia tenella*.

Appendix 2.3 Edinburgh Associative Thesaurus

The Edinburgh Associative Thesaurus (EAT) [7] is a dataset consisting of empirical word associations. When the dataset was collected, participants were given a sheet with stimulus words or phrases and were instructed to write a single response word or phrase for each stimulus. Each participant was given 100 stimuli. See http://vlado.fmf.uni-lj.si/pub/networks/data/dic/eat/Eat.htm for more details. We use the MATLAB version from the SuiteSparse Matrix Collection https://sparse.tamu.ed [8] with a total count of 788,828 associations between the 23,219 entries: 8,210 unique entries were used as stimuli and 22,765 were given as responses.

TABLE 54
Edinburgh Associative Thesaurus: Response Stimulus. Descriptive features (stimuli) selected for each cluster of responses.

1: 40, 22, psalm, 18, route, 30, unready, life-span, million, century, teen, twelve, 10, 1968, decade, year, 13, twenty-one, seconds, thirty, ages, fortnight, fifteen, years, 9, 20th, thousand, month, 19, lifetime, eighteen, hundred, five, ninety, calendar, 24, six, forty, seven, minutes, 100, seventy, age, nine, o'clock, twentieth, twenty, ten
2: david, alan, auntie, martin, hardy, ed, elegy, ronnie, godard, goodbye, saint, walter, dell, henry, doc, dr, inter, russell, william, john, cole, al, charles, emile, george, liz, darling, ken, mick, mister, glen, arthur, washington, mike, bye, orphan, bond, blimey, fie, beadle, wilson, aunt, author, dolittle, dee, elizabeth, dale, philosopher, ned, marx
3: humph, persists, choosers, constance, la, prolongs, hillock, izzy, lope, andrews, mount, louis, jack, anthill, chinaman, doubles, prolong, continuum, ho, caper, continual, twice, conquered, volcano, earthquake, shares, descending, flux, chinese, persevere, follows, semibreves, leonard, ling, remains, cone, encore, dizzy, atoll, larva, non, plurals, shrews, waltzers, hoo, grabs, et, lo, creepers
4: degrade, letting, debts, deficit, troubles, bills, redress, emitted, provider, owned, excluded, obligations, unpaid, exclude, recompense, repay, remit, send out, distribute, reprimanded, fund, repudiate, tax, credit, merit, return, joined, abandon, restore, distributed, payment, abdicate, debt, give out, hired, remittance, admission, more money, collect, refund, pay, allow, owing, replace, dispense, owed, concessive, wage, due, increase
5: proof, teaching, solving, review, effects, unequal, specialize, qualify, clause, plan, deduce, imprecise, selection, method, personal, observations, objective, means, theory, systematic, of a kind, decisions, words, cognition, difference, objects, instructions, diverse, analogy, process, values, determine, scientific, comparison, information, dative, particular, origin, basic, interpolate, progress, analysis, procedure, option, analyse, fact, critical, data, stipulation, arrange
6: classes, catalogue, commerce, posts, meetings, sports, manager, goods, aircraft, players, dealer, bureau, furniture, barrier, posture, theatre, hobbies, card, fare, retail, pose, ticket, warehouse, seats, post office, supermarket, coach, shop, stationery, studio, advertising, crossing, shops, sales, hall, status, production, directors, architects, freight, crowded, airport
7: tower, pier, county, marchioness, busby, jutland, mountains, midlands, halifax, continents, castle, maisonettes, wales, sussex, isles, vale, bristol, castles, arches, dukedom, isle, inland, buildings, suburbs, flagstones, outboard, inca, peninsula, via, newyork, brussels, ville, lakes, moorland, cherubs, mainland, colonies, whales, peasants, hills, chalet, docks, devon, moor, germany, spain, urban, lands, rome, canals
8: contents, diffraction, element, magnetism, submarine, mended, scientist, forceps, apparatus, wad, lens, flexible, jotting, dislocated, mal, lined, ion, valve, synthetic, expansion, spectacles, ganglia, mechanics, chemical, dot, magnetic, tilt, weights, fusion, edges, reactor, micro, connect, physical, muscles, atomic, structure, solder, electricity, fluids, compound, refill, wire, circuit, weld, atom, chloride, organic, plastic, bipolar
9: charlady, cider, corn, apple-tree, hen, prickly, pans, reddish, butter, toot, bud, meadows, roe, sipping, mop, veal, sprout, blueberry, chops, buds, vegetation, sedge, cheese, crab, hacks, olives, slime, plants, pickles, hawthorn, dandelion, fry, geese, fruit, planting, boiled, pigs, onion, roots, insects, lard, nuts, mildew, seeds, rotting, kippers, petals, drainage, buttercups, ponds
10: glycerine, potash, sugars, absorb, salts, acids, gland, choler, spuds, digestion, magnesium, kidneys, cough, stodge, indigestion, antiseptic, enzymes, fattening, vitamin c, jam, sweets, cereal, rice pudding, flour, gravy, vim, gastric, liver, potatoes, slurp, seasoning, oesophagus, essence, vegetables, protein, colic, cooker, tasty, chocolate, pies, pancreas, beef, creamy, flavour, savoury, bitters, sherbet, grapefruit, veg, sauce
11: bird, eery, doves, rake, jokes, witches, bitch, ghosts, smart, mighty, dad, graceful, nanny, dog, jeep, graveyard, dame, unicorn, woods, monster, gorilla, randy, wolf, wail, effeminate, eagle, lass, hairy, tigress, silly bitch, incense, hawk, blonde, fluffy, glade, aspidistra, witch, sexless, shrew, cub, peacock, tawdry, horns, witchcraft, brassy, prey, predator, grandma, feminine, scraggy
12: tribe, terrorist, russia, slavery, immigration, overlord, c. of e., waster, hungary, defender, dictatorship, townsman, foreigners, settlement, protestant, immigrants, immigrant, regime, mayoral, officials, alabama, india, capitalist, heraldry, customs, monarchy, governors, home rule, oppressor, hostage, communists, aggressor, warlike, wog, commandant, islam, national, servility, racialism, bushman, republic, indians, capitalism, warrior, wogs, indian, anarchist, american, liege, colonies
13: pi, variables, lb, 3/4, per, frequency, rounded, quantities, rectangle, sums, triad, circle, shillings, rang, computer, equation, cubic, angle, bipolar, measure, centimetre, geometrical, coordinate, vector, dimension, series, pound, gram, metric, 2001, wireless, decimal, curves, sixpence, longitudinal, gramme, ellipse, rhombus, axis, hexagonal, stereo, geometry, diagonal, division, remainder, percentage, fraction, calculate, per cent, 100
14: realised, freely, persuaded, thoughtlessly, hopeful, apt, accepted, forgave, consider, in need, understood, deliberate, unfounded, ignored, intention, positively, unnecessary, reasonable, asked, forgot, disprove, proposition, obliged, gladly, certainty, repent, assured, surmise, never mind, couldn't, irrelevant, carefully, thanks, ability, unproven, explain, acknowledged, objection, ascertain, decide, doesn't, promised, deny, necessary, please, unable, prove, agree, reason, pertain
15: thugs, gangsters, hideaway, basher, thug, knifing, foundered, swords, hides, cosh, gangster, hideout, aground, lawless, perishers, castaway, violence, asylum, contraband, mob, thieves, chicago, rifleman, knives, cops, crew cuts, rifles, forces, riot, burglary, fences, loot, headquarters, anarchy, gang, rabble, trespass, outlaw, break in, bandit, fist-fight, rampage, detective, robber, criminology, rifle, burglar, raid, armed, crime
16: speaker, infamous, liar, humanism, manners, respect, bigotry, atheist, phoney, fame, nonconformist, frank, egomaniac, haughty, sincerity, honest, unholy, creed, representative, politic, ideals, moral, faithless, liberal, heckler, moderate, realistic, honour, hypocrisy, immoral, decent, spiritual, idealism, truth, righteousness, dogma, morality, polite, preaching, sinner, goodness, honor, religious, pious, righteous, methodist, related, beloved
17: attach, flagpole, pow, orders, about turn, draft, flags, prestige, cadet, convention, baton, welsh, flag, standard, actions, enlist, regiment, onward, colonel, wham, procession, corps, general, coup, enrol, backward, armies, cavalry, lieutenant
18: thereafter, therein, besides, forthwith, whatsoever, otherwise, penultimate, thereupon, as before, anyway, et cetera, whilst, began, whence, wherein, despite, wherefore, to what, however, finally, meanwhile, whatever, were, recently, nevertheless, soon, moreover, than, all the time, thus, unless, was, notwithstanding, though, happened, somehow, unto, perhaps, except, therefore, accordingly, also, eventually, yet, although, but, as, again, especially, since
19: beguilingly, aghast, distraught, quietly, coaxing, relaxing, slowness, paranoia, crying, hysteria, glee, laughter, impatient, fright, tantrum, easiness, grief, boredom, morbid, terribly, heartily, grieving, monstrously, leer, arousal, downcast, sad, shy, fervour, stupidly, old age, confused, peaceful, enthusiasm, insipid, liveliness, restless, annoyed, frown, anguish, morose, frenzy, depressed, woe, furtive, uneasy, temperament, wary, sadness
20: sensibly, stand out, watched, talkative, highly, distinguished, vocative, diligently, tested, arranged, fate, looked, pleased, said, mentioned, stated, exceptionally, properly, clearly, optimism, trance, beholder, admirable, foremost, insight, vision, genuinely, amaze, sisal, clever, really, files, perceived, beautifully, anonymous, delightful, heard, listener, visual, exceptional, faithfully, incognito, intelligent, seen, genius, nicely, listened, appeared
21: skirts, lacy, tiles, spindle, bracelet, horses, jewel, jumpers, woven, bikes, diamond, silver paper, boots, pants, rucksack, mending, shirts, hoist, togs, holes, handles, pelmet, knickers, tractor, tapestry, railings, pavement, covering, floors, manhole, weave, torn, silk, curtain, floor, jacket, balls, tacks, floats, hooves, buttons, cloth, lining, string, tracks, drawer, vest, pegs, braid, scissors
22: priestess, parson, singing, virginals, shrine, poems, grandeur, aries, saga, souvenirs, beethoven, planets, rites, jupiter, musician, glamour, viola, splendour, opera, earl, prophet, leo, brahms, classical, gazer, serenade, priest, mythology, ballet, mozart, poet, pagan, venus, ludwig, angels, sculpture, gazers, bishop, rhymes, stars, cathedral, songs, concerto, mars, symphony, bach, trio, statesman, mythical, recital
23: restrained, unrest, heaved, handicap, grabbed, clamped, eradicate, broken, harmed, headed, gripped, destroyed, released, tightly, hit, disturbed, immobile, destroy, fight, mania, repress, blunder, avoidance, agitate, hindered, tried, annoy, feinted, accidentally, retaliate, poked, casualty, killed, annihilate, annoying, disturbance, menace, malfunction, mismanaged, tackled, attack, disrupt, offend, convicted, disturb, fixed, prevent, abducted, tense, hinder
24: shake, cushion, snakes, totter, shiver, caress, repellent, snake, pains, flea, belting, pincer, gasp, fasten, fester, repulsive, withered, flimsy, tooth, itch, squirm, throbbing, pinching, bruise, trotted, clinging, mender, lounging, jerk, bouts, uncomfy, wriggling, flinch, immunity, bitten, smack, stiff, pant, nausea, convulse, crawling, stifle, injection, germs, hay fever, lithe, rasp, sore, painful
25: idiots, fairies, elf, legend, pansies, mac, goblin, shrunken, grotesque, fairy tale, choirs, dozy, dwarf, toadstool, fantasies, pure, portrait, dwarfs, causeway, wee, smallest, ladder, conjurors, livery, shorter, fairy, mime, tales, incredible, idiot, bad luck, fables, bewitched, mirrors, tiny, fable, wand, enchantment, weird, myth, fantasy, mythology

TABLE 55
Edinburgh Associative Thesaurus: Response Stimulus. Correspondence between the clusters (described by stimuli) and a random subset of their instances (responses).

1: 40, 22, psalm, 18, route, 30, unready, life-span, million, century, teen,. . .	20, feeder, chests, youth, weeps, jack bruce, aggregate, basie, proverbs
2: david, alan, auntie, martin, hardy, ed, elegy, ronnie, godard, goodbye,. . .	annual, ph.d., d.c., cher, kith, tiddlywinks, matthew, karl, ding-dong
3: humph, persists, choosers, constance, la, prolongs, hillock, izzy, lope,. . .	stirs, an, last one, chinese, bricky, staggering, scoring, funfair, oracle
4: degrade, letting, debts, deficit, troubles, bills, redress, emitted,. . .	allot, pretending, creditors, characteristic, istanbul, separated, married
5: proof, teaching, solving, review, effects, unequal, specialize, qualify,. . .	schedule, adjective, legacy, coalition, broadcast, same as, q.c., fend
6: classes, catalogue, commerce, posts, meetings, sports, manager, goods,. . .	reproduction, trams, travel, afar, housing, business, bob, receptacle
7: tower, pier, county, marchioness, busby, jutland, mountains, midlands,. . .	steamboat, and clyde, londoner, by way of, bognor, stalactites, schooner
8: contents, diffraction, element, magnetism, submarine, mended, scientist,. . .	volt, apparatus, micro, coma, carved, urinate, taxonomy, h-bomb, rapier
9: charlady, cider, corn, apple-tree, hen, prickly, pans, reddish, butter,. . .	of time, wafer, supping, scotch, yardley, stic, scrub, amateurs, fowls
10: glycerine, potash, sugars, absorb, salts, acids, gland, choler, spuds,. . .	porton, veal, squeezer, tumbler, steak, alcohol, granulated, toothbrush
11: bird, eery, doves, rake, jokes, witches, bitch, ghosts, smart, mighty, dad,. . .	sheep, wearer, weird, apfel, invincible, carnation, brolly, sewing-machine
12: tribe, terrorist, russia, slavery, immigration, overlord, c. of e., waster,. . .	white collar, amazons, of ireland, claude, capitalists, wars, town council
13: pi, variables, lb, 3/4, per, frequency, rounded, quantities, rectangle,. . .	4th, pound, coordinates, cape kennedy, zebedee, lateral, sum, milestone
14: realised, freely, persuaded, thoughtlessly, hopeful, apt, accepted, forgave,. . .	obliged, wanting, thinking, revealed, tangible, rewarded, accuse
15: thugs, gangsters, hideaway, basher, thug, knifing, foundered, swords, hides,. . .	assassination, trooper, demonstrator, cluster, portrait, sabre, pillage
16: speaker, infamous, liar, humanism, manners, respect, bigotry, atheist,. . .	radical, tolerant, town crier, advisor, infidelity, christianity, boyfriend
17: attach, flagpole, pow, orders, about turn, draft, flags, prestige, cadet,. . .	call up, services, norm, vietnam, girl guides, inward, charger, send away
18: thereafter, therein, besides, forthwith, whatsoever, otherwise, penultimate,. . .	all under, precisely, dedication, brethren, is, therefore, morrow, instance
19: beguilingly, aghast, distraught, quietly, coaxing, relaxing, slowness,. . .	raving, terrifying, thin air, amazement, idiocy, withdrawal, weakness
20: sensibly, stand out, watched, talkative, highly, distinguished, vocative,. . .	careful, hypnotise, feelingly, wellknown, very good, knighted, doom
21: skirts, lacy, tiles, spindle, bracelet, horses, jewel, jumpers, woven,. . .	of cancer, tin, spokes, rockets, tweed, socks, rods, armpit, wooden, gauze
22: priestess, parson, singing, virginals, shrine, poems, grandeur, aries, saga,. . .	angels, pictures, crosses, churchy, victorious, lord, soc, miracles
23: restrained, unrest, heaved, handicap, grabbed, clamped, eradicate, broken,. . .	ative, quicken, destructive, corsets, whirlpool, kept back, stopped
24: shake, cushion, snakes, totter, shiver, caress, repellent, snake, pains,. . .	sickly, trenchant, tug, and judy, kip, writhe, censored, breath, shrug
25: idiots, fairies, elf, legend, pansies, mac, goblin, shrunken, grotesque,. . .	cinderella, magnitude, homosexual, oscillator, fraser, gnomes, unsightly

TABLE 56
Edinburgh Associative Thesaurus: Stimulus Response. Descriptive features (responses) selected for each cluster (of stimuli) sorted by coefficient value.

1: stupid, bad, love, happy, child, laugh, sex, funny, wrong, fool, mother, me, friend, woman, false, joy, nice, good, male, man, hate, daft, fun, person, boy, kind, girl, mad, women, friendly, smile, happiness, fear, youth, loving, no, people, pleasure, yes, sad
2: church, god, religion, priest
3: tree, flower, flowers, green, trees, plant, yellow, garden, grass, leaves, red, rose, roses, grow, hedge, bush, leaf, purple, plants, blossom, woods, summer, fruit
4: ship, army, war, soldier, police, america, law, ships, boat, officer, judge, prisoner, violence, criminal, prison, navy, uniform, general, castle, king, guns, illegal, court, lawyer, captain, chief, vietnam
5: correct, vote, look, blank, degree, difficult, word, people, make, talk, number, committee, questions, years, idea, twenty, see, first, clever, subject, new, theory, find, rubbish, psychology, maths, politics, computer, things, order, of, facts, time, study, exam, thought, answer, useless, part, work, one, test, life, class, words, observe, think, question
6: paper, school, book, picture, king, magazine, t.v., news, story, letter, teacher, work, art, board, maths, write, charles, numbers, science, newspaper, writing, drawing, read, words, ink, pen, pictures, learn, office, desk, rubbish, lesson, books, notes, computer, lines, page, letters, reading, journal, learning, study
7: now, not, yes, to, all, never, them, ever, you, why, finish, gone, must, again, think, here, then, no, that, near, more, often, other, nothing, go, so, soon, else, it, do, time, i, when, is, where, something, anything, there, yet, will, if, but, this, what, for
8: food, drink, eggs, bread, orange, cheese, soup, fat, eat, beer, cake, meat, tea, sugar, sweet, pudding, cooking, breakfast, chocolate, butter, jam, beef, stew, sweets, steak, fruit, cream, pie, sour, apple, milk, wine, dinner, cakes, sauce, meal, lemon, juice
9: car, train, fire, cars, stop, bus, bang, lorry, slow, road, fast, run, wheel, gun, speed, traffic, crash, motor, engine, wheels, travel, transport, vehicle
10: money, rich, buy, steal, thief, give, price, pounds, gain, bank, wealth, pound, goods, loot, shop, dollar, broke, miser, debt, poverty, theft, coins, pay, savings, spend, pennies, cost, cash
11: bird, dog, horse, animal, sheep, lion, cat, fly, animals, mice, cow, fox, cows, dogs, tail, tiger, farm, horses, hound, cattle
12: pain, death, blood, hurt, sad, doctor, hospital, anger, fight, fear, medicine, ill, sick, disease, dead, fright, cry, stomach, illness, knife, wound, scream, sore, hate
13: door, town, house, round, straight, carpet, houses, up, london, mountain, stone, wall, lift, wide, top, building, wood, bridge, circle, football, north, iron, climb, ball, out, ladder, high, roof, hill, stairs, game, flat, tall, floor, earth, down, brick, road, square, height, flats, hole, concrete, fall
14: water, sea, fish, river, bath, wet, beach, swim, wind, flow, deep, salt, rain, sand, stream, seas, shore, surf, sink, boat
15: light, cold, snow, smoke, sun, stars, glass, bright, heat, cancer, ice, windows, star, dirt, gold, mist, air, chemistry, shine, sky, soap, hot, clear, rain, dark, white, moon, winter, weather, glow, bulb, water
16: hair, shirt, foot, arm, trousers, legs, sharp, sew, smooth, dress, cloth, socks, clothes, face, nose, knees, head, shoes, brush, leg, holes, toe, skirt, coat, nail, cut, cotton, silk, wool, needle, shoe, hand, scissors, skin, clean, scarf, thread, stocking, lace, vest, sock, material
17: music, noise, dance, band, song, play, piano, record, ear, loud, beethoven, sound, violin, guitar, theatre, radio, voice, hear, player
18: sleep, bed, tired, lazy, chair, awake, quiet, rest, baby, night, comfort, easy, ease, comfortable, lie, soft, peace, sofa, relax, asleep, settee, sit, armchair, sleepy, pillow

TABLE 57
Edinburgh Associative Thesaurus: Stimulus Response. Correspondence between the clusters (described by responses) and a random subset of their instances (stimuli).

1: stupid, bad, love, happy, child, laugh, sex, funny, wrong, fool, mother, me,. . .	amuse, obstinate, corrupt, chuckle, humour, thankful, kindliness, honest
2: church, god, religion, priest	catholic, prayer, lord, satan, fervour, hall, ceremony, praise, sins, abbey
3: tree, flower, flowers, green, trees, plant, yellow, garden, grass, leaves,. . .	bile, hideout, flower-pot, lemon, danger, enlarge, sap, plant, birds
4: ship, army, war, soldier, police, america, law, ships, boat, officer, judge,. . .	barrister, hero, elizabeth, leader, cox, decker, rifles, powerful, bayonet
5: correct, vote, look, blank, degree, difficult, word, people, make, talk,. . .	addition, therein, distribution, convention, organise, quaint, observe
6: paper, school, book, picture, king, magazine, t.v., news, story, letter,. . .	bench, feature, digits, rhombus, tales, reagents, graph, copy, seller
7: now, not, yes, to, all, never, them, ever, you, why, finish, gone, must,. . .	after, presences, compulsion, belonging, knowing, they, nearer, is
8: food, drink, eggs, bread, orange, cheese, soup, fat, eat, beer, cake, meat,. . .	bars, swede, desert, cooked, peeler, rice pudding, oxo, empty, cellar
9: car, train, fire, cars, stop, bus, bang, lorry, slow, road, fast, run,. . .	bang, fume, drivers, tongs, stuck, hire, conductor, triumph, dangerous
10: money, rich, buy, steal, thief, give, price, pounds, gain, bank, wealth,. . .	beneficent, inherited, executive, yen, unemployed, mean, conned, decrease
11: bird, dog, horse, animal, sheep, lion, cat, fly, animals, mice, cow, fox,. . .	biting, foul, egg, shepherds, tame, scraggy, hack, conjurors, soar
12: pain, death, blood, hurt, sad, doctor, hospital, anger, fight, fear,. . .	annoyed, war, fate, despair, prescription, skull, pincer, frightened
13: door, town, house, round, straight, carpet, houses, up, london, mountain,. . .	base, ton, elevate, county, plan, semi-detached, path, floor, chisel
14: water, sea, fish, river, bath, wet, beach, swim, wind, flow, deep, salt,. . .	boil, ice, fire, wind, mersey, depth, douse, seascape, resort, borne, slosh
15: light, cold, snow, smoke, sun, stars, glass, bright, heat, cancer, ice,. . .	breaking, haze, fan, weight, vim, lenses, cough, detergent, shadows, polish
16: hair, shirt, foot, arm, trousers, legs, sharp, sew, smooth, dress, cloth,. . .	blanco, yards, finger, dermis, scuff, snob, samson, grips, coarse, shirts
17: music, noise, dance, band, song, play, piano, record, ear, loud, beethoven,. . .	brass, quietly, mike, song, grand, hearing, bugle, tuba, puppet
18: sleep, bed, tired, lazy, chair, awake, quiet, rest, baby, night, comfort,. . .	breathless, reverie, mistress, slothful, fatigued, homely, camp