



1019 U

The information contained herein is for the use of employees of Bell Laboratories and is not for publication. (See GEI 13.9-3)

Title- Computer Detection of Typographical Errors

Date- April 1, 1973

TM- -73-1271-4

Other Keywords- Text
Computing
Mistakes

Author
R. Morris
L. L. Cherry

Location
MH 2C524

Extension
X3878

Charging Case- 39199
Filing Case- 39399-11

ABSTRACT

A computer program written for the UNIX time-sharing system reduces by several orders of magnitude the task of finding words in a document which contain typographical errors. The program is adaptive in the sense that it uses statistics from the document itself for its analysis.

In a first pass through the document, a table of digram and trigram frequencies is prepared. The second pass through the document breaks out individual words and compares the digrams and trigrams in each word with the frequencies from the table. An index is given to each word which reflects the hypothesis that the trigrams in the given word were produced from the same source that produced the trigram table. The words are sorted in decreasing order of their indices and printed. Printing is suppressed for words appearing in a table of 2726 common technical English words. The table is attached as Appendix B.

The author of a 108-page document needed less than ten minutes to scan the output and identify the misspelled words. There were a total of 30 misspelled words among the 386 words output and 23 of those occurred among the first 100 words output.

Pages Text 5	Other 13	Total 18
No. Figures 0	No. Tables 0	No. Refs. 3

April 1, 1973

Introduction

A large proportion of the usage of the UNIX time-sharing system at Murray Hill has been for the preparation of documents. A major part of the task of document production has been made automatic by the existing hardware and software. One of the most costly remaining parts of the overall job is the manual proofreading of the document in draft form.

The method described here substantially reduces the time required to perform part of this proofreading task, namely that of checking the correct spelling of individual words.

Synopsis of the program

A document is read by the program and an attempt is made to recognize the individual words in the text.

Word breakout

- all upper case characters are mapped into lower case,
- apostrophes are deleted from within words
- words broken across lines are reassembled, and
- most ROFF commands are ignored.

Two processes are then performed on the words as they are input. First they are scanned for their digram and trigram statistics and secondly they are sorted, duplicate words are cast out, and one copy of each word is stored in a temporary file.

Typical experiences with large files is that the number of distinct words remaining varies greatly depending on the author, but for a given author, the number of distinct words increases approximately like the square root of the total number of words. Typically a ten thousand word document will contain about 1500 distinct words.

As the words are being broken out, a count is kept of each kind of digram and trigram in the document. For example, if the word 'once' appears in the text, the counts corresponding to each of these five digrams and four trigrams are incremented.

Trigram counting

```
.o  on  nc  ce  e.
.on  onc nce  ce.
```

where the '.' signifies the beginning or ending of a word. These statistics are retained for later processing.

In order that the statistics should give some meaningful results for very short documents, the digram and trigram tables are initialized with some statistics taken from typical technical English text. This has little effect on the results for large documents and can be suppressed by the user if desired.

The table of sorted words is compared with a table of 2726 common technical English words as used at Murray Hill and those words that appear in the table of common words are removed.

Removal of common words

The table of common words contain the results of processing about one million words of technical text appearing in documents produced on the UNIX time-sharing system at Murray Hill. The words selected were those which have the highest probability of appearing in a given document. This is not the same as a list of the most frequently occurring words and, in fact, some very rare words occur in the table. For example, the words 'murray' and 'abstract' are very uncommon words in the documents sampled, yet they both appear once in virtually every document. A copy of this list of common words appears in Appendix B.

The number of words removed by consulting the table varies in practice from about 500 to about 1500 and is dependent more on the author than on the length of the document.

The remaining words from the original document are then read one by one and a number is attached to each word by consulting the table of digrams and trigrams previously produced.

Trigram statistics

The number is an index of peculiarity and reflects the likelihood of the hypothesis that the trigrams in the given word were produced from the same source that produced the trigram table.

Each trigram in the current word is treated separately. An initial and terminal trigram is treated for each word so that the number of trigrams in a word becomes equal to the number of characters in the word. For each trigram $T = (xyz)$ in the current word, the digram and trigram counts from the prepared table are accessed. Each such count is reduced by one to remove the effect of the current word on the statistics. The resulting counts are $n(xy)$, $n(yz)$, and $n(xyz)$. The index $i(T)$ for the trigram T is set equal to

Computing the index

$$i(T) = \frac{1}{2}[\log n(xy) + \log n(yz)] - \log n(xyz)$$

This index is invariant with the size of the document, i.e. it needs no normalization. It measures the evidence against the hypothesis that the trigram was produced by the same source that produced the digram and trigram tables in the sense of References 1 and 2. In the case that one of the digram or trigram counts is zero, the log of that count is taken to be -10 .

A variety of methods were tried to combine the trigram indices to obtain an index for the whole word including:

- the average of the trigram indices,
- the square root of the average of the squares of the trigram indices, and
- the largest of the trigram indices.

All were moderately effective; the first tended to swamp out the effect of a really strange trigram in a long word and the latter was insensitive to a sequence of strange trigrams whose cumulative evidence should have made a word suspicious. The second appeared on trial to be a satisfactory compromise and was adopted.

The words with their attached indices are sorted on the index in decreasing order and printed in a pleasing three-column format. The indices have been normalized in such a way that a word with an index greater than 10 contains trigrams that are not representative of the remainder of the document. Such a word almost certainly appears just once in the document.

Printing the results

The results for a 108-page document typed into the UNIX system and described below are representative of the results for large documents. The total number of words in this document was 19917. The number of distinct words was 1207 and of these, 821 were removed by reference to the list of common words, leaving the 386 words in Appendix A to be output. The total running time of the program for this document was about 3 minutes on a PDP 11-45 minicomputer.

An example

The sample document uses a smaller than average vocabulary for technical writing because it was a very detailed and unusually repetitive description of a limited subject area. On the other hand, it is typical of documents of its size and larger that are typed into the UNIX system.

An author of the document took less than 10 minutes to scan the output and identify the misspelled words. A total of 30 misspelled words were found. The value of the trigram analysis is indicated by the fact that 23 of the misspelled words occurred among the first 100 words output and only 3 occurred among the last 100 words output.

Proofreading the original document and finding this many typographical errors would be a nearly impossible task. Experiments with other documents have shown that an author will find barely half of the misspelled words in his own document if the document is more than about 20 pages long, but that he will find virtually every misspelling by looking once through the output of this program.

Two keys to the very high rate of error locating are that the output from the program is exceedingly short and every word can be looked at rather carefully in only a few minutes and secondly that the percentage of errors is sufficiently high, especially at the beginning of the output, that the effort is psychologically rewarding. The author, seeing some errors among the first few words, is induced to continue scanning the list with considerable care.

The misspelled words located by the author of the sample document have been marked in Appendix A.

Considerable reduction in the volume of output could be achieved by discarding words that occurred in the original document more than once and retaining for further processing only those which occurred just once (hapax legomena) in the document. This extension met with considerable customer resistance since it concealed consistent misspellings. On the other hand, no attention is now paid to the potential value of treating specially words which occur just once. They evidently would repay some special attention.

Some possible extensions

The program does not by any means do the whole job of proofreading, and the draft document still has to be read, but with considerably less care. The program does not find missing or extra words or semantic nonsense.

*Drawbacks of the
program*

The principal annoyance in using the program is a consequence of the initial word breakout which maps all letters into lower case and removes apostrophes and the like. The result of this processing and the (necessary) lack of any information about where the offending word occurred in the original makes it rather difficult to find on occasion, even with the use of a context editor.

Trigram Analysis

Early attempts to separate misspelled words from correct English words by the use of preset trigram tables failed completely. It appeared that the trigram frequencies of a document, no matter how long, did not approach the trigram frequencies of a document by some other author or on some other subject.

A table of trigram frequencies was prepared using the 2006 most frequent words in the English language according to Ref. 3 and this table was compared against actual samples of technical and non-technical English text. The only authors whose writings showed a useful correlation were Edmund Burke and Thomas Jefferson. In every other case (including, for example, Mark Twain) the prepared table was absolutely useless.

On the other hand, in any particular document, the trigram frequencies obtained from one part of the document were highly correlated with frequencies from the remaining parts of the document, particularly in technical writing. For this reason, the trigram frequencies were gathered from the document being analysed even though this nearly doubled the running time of the program.

[1] Good, I. J. *Probability and the Weighing of Evidence*. Charles Griffin & Co., London, 1950.

[2] Kullback, S. and Leibler, R. A. On Information and Sufficiency. *Annals of Math. Stat.*, 22 (1951), pp. 79-86.

[3] Kučera, H. and Francis, W. *Computational Analysis of Present-Day American English*. Brown Univ. Press, Providence, 1967.

MH-1271-rhm-pdp

ROBERT MORRIS
L. L. CHERRY

Attached

Appendix A
Appendix B

17 nd	6 cetera	4 miscellaneous
17 heretofore	5 zeros	4 messages
17 erroronously	5 virtually	4 legs
16 suer	5 ultimately	4 ired
16 seized	5 truncate	4 interconnections
16 poiter	5 therewith	4 integrity
16 lengthy	5 thereafter	4 intact
16 inaccessible	5 spectre	4 incorporating
16 disagreement	5 rewritten	4 illustratively
16 bwirte	5 raises	4 illustrates
15 violating	5 prefix	4 guarded
15 unaffected	5 pesudonym	4 exhausted
15 tape	5 neames	4 eventually
15 swapped	5 namees	4 erroneously
15 shortly	5 multiplied	4 erroneous
15 mutilated	5 interrelationship	4 denied
15 multiprogramming	5 inefficient	4 dedicated
15 likewise	5 icalc	4 decremented
15 datum	5 handler	4 corrected
15 dapt	5 flag	4 complexities
15 cumulatively	5 exercised	4 circular
15 consulted	5 erroneous	4 choosing
15 consolidation	5 dumped	4 candidate
15 checking	5 dump	4 bulk
15 accordinng	5 deficiency	4 borrowing
14 typical	5 controller	4 borrow
14 tabular	5 contiguous	4 arbitrarily
14 supplying	5 changing	4 ancw
14 subtle	5 bottoms	3 wishes
14 shortcoming	5 bitis	3 verifying
14 pivotal	5 ascertain	3 verify
14 invalid	5 accomodate	3 verifies
14 infrequently	4 unnecessarily	3 unused
14 flexible	4 traversing	3 undertaken
14 flags	4 tracing	3 unallocated
14 conceptually	4 totally	3 treating
14 bwaite	4 tops	3 treat
14 broadly	4 thirteen	3 traverses
14 amy	4 tallyed	3 traversed
14 adds	4 summarized	3 traces
14 accompanying	4 strictly	3 trace
13 overwritten	4 simultaneous	3 therein
13 occupying	4 retrieval	3 termed
13 lookup	4 quotient	3 tally
13 flagged	4 proceeding	3 systematically
9 iin	4 proceeded	3 suffer
8 subrouutine	4 preceeding	3 successfully
8 adjunct	4 popularity	3 subroute
7 drawbacks	4 periods	3 subdirectory
6 thee	4 owners	3 subdirectories
6 odification	4 owner	3 sub
6 od	4 occupies	3 structurally
6 indicator	4 nonzero	3 simplify
6 imminent	4 nonowners	3 simplified
6 formats	4 nonexistence	3 sequential

3 separately	3 implementing	3 behalf
3 separated	3 iget	3 assures
3 searching	3 identifying	3 assure
3 searches	3 identifies	3 arrows
3 searched	3 hierarchy	3 arising
3 scheme	3.hierarchal	3 amodification
3 saved	3 grown	3 alloc
3 salvage	3 functionally	3 administrated
3 retrieving	3 fixptr	3 administrate
3 retrieved	3 findb	3 accomplishes
3 retrieve	3 extension	3 accommodate
3 retained	3 extending	3 accesses
3 respects	3 exits	2 writes
3 resides	3 exceptions	2 volumes
3 reside	3 entering	2 updated
3 requests	3 ending	2 unlinking
3 requesting	3 ended	2 unlink
3 renaming	3 encounter	2 transfers
3 rename	3 elimination	2 transferred
3 relieve	3 discrepancies	2 symbolic
3 referring	3 diagrammatically	2 succeed
3 referenced	3 diagrammatical	2 subroutine
3 recording	3 determination	2 storing
3 recorded	3 designating	2 specially
3 recalled	3 deletion	2 simulate
3 pseudonym	3 deleting	2 sharing
3 provisions	3 deletes	2 shared
3 priorly	3 deleted	2 returning
3 presumed	3 default	2 restore
3 presently	3 dealt	2 removing
3 preparation	3 dealing	2 reads
3 precaution	3 dcheck	2 proceeds
3 preassigned	3 corrects	2 proceed
3 permitting	3 correctness	2 portions
3 permission	3 copy	2 pointing
3 permissible	3 copies	2 physically
3 outer	3 copied	2 outset
3 opens	3 consistency	2 ot
3 occurrence	3 consist	2 offset
3 necessity	3 consequence	2 obtains
3 manipulating	3 conjunction	2 obtaining
3 manipulate	3 computers	2 naming
3 itrunc	3 complements	2 namei
3 interprets	3 closing	2 mdir
3 interpret	3 closes	2 marking
3 interference	3 clears	2 mared
3 interfere	3 cleared	2 maintaining
3 interact	3 claimed	2 locates
3 intends	3 checks	2 locate
3 initiation	3 charts	2 linking
3 indexing	3 cautiously	2 leg
3 indexes	3 bypassed	2 iwrite
3 independently	3 buffering	2 inventive
3 indentified	3 bread	2 inner
3 inconsistent	3 bilateral	2 indistinction

2 indirect	1 buffer
2 indices	
2 indication	
2 indexed	
2 increments	
2 incrementing	
2 incremented	
2 handled	
2 freed	
2 fget	
2 duplication	
2 duplicated	
2 duplicate	
2 divides	
2 distinctive	
2 distinctions	
2 disk	
2 directories	
2 creation	
2 creating	
2 creates	
2 corrections	
2 content	
2 complicate	
2 completes	
2 compares	
2 commonly	
2 commences	
2 chdir	
2 calculations	
2 calculates	
2 calculated	
2 calculate	
2 bwrite	
2 bitset	
2 backup	
2 allocation	
2 allocating	
2 allocates	
2 allocated	
2 allocate	
2 allocatable	
2 accounting	
2 accounted	
2 accessing	
1 pointers	
1 newly	
1 mget	
1 map	
1 locating	
1 fix	
1 entry	
1 enters	
1 directory	
1 buffers	

Common Technical English Words

ability	affairs	annual	aside	because
able	affect	anode	ask	become
about	affected	anodes	asked	becomes
above	affects	another	asking	becoming
absence	aforementioned	answer	aspect	been
abstract	after	answered	aspects	before
academic	afternoon	answers	assign	began
accept	again	any	assigned	begin
acceptable	against	anyone	assignment	beginning
accepted	age	anything	assistance	begins
accepts	agencies	apart	associated	behavior
access	agency	apparatus	association	behind
accessed	ago	apparent	assume	being
accessible	agreed	apparently	assumed	belief
accomplished	agreement	appeal	assumes	believe
accordance	ahead	appear	assuming	believed
according	aid	appearance	assumption	bell
accordingly	air	appeared	asymmetric	below
account	algebraic	appearing	at	beneath
accuracy	algot	appears	atmosphere	benefit
accurately	algorithm	application	atom	bent
achieve	algorithms	applications	attack	besides
achieved	all	applied	attempt	best
achievement	allow	applies	attempts	better
across	allowed	apply	attention	between
act	allows	applying	attitude	beyond
acting	almost	approach	attorney	bill
action	alone	appropriate	attractive	binary
actions	along	appropriately	audio	biometrika
activated	already	approximate	august	bit
activates	also	approximated	author	bits
activation	alter	approximately	authority	black
active	altered	approximation	automatic	block
activities	alternate	april	available	blocks
activity	alternative	arbitrary	average	blue
actual	alternatively	are	avoid	board
actually	alternatives	area	aware	body
actuated	although	areas	away	book
adapted	always	argued	axes	books
add	america	argument	axis	both
added	american	arguments	back	bottom
adding	among	arise	background	bought
addition	amount	arises	bad	bound
additional	amplifier	arithmetic	balance	boundaries
additions	amplitude	around	ball	boundary
additive	an	arrangement	bandwidth	bounds
address	analog	arrangements	bank	box
addressed	analogous	array	bar	branch
adequate	analyses	arrays	base	branches
adjacent	analysis	arrive	based	break
administration	analyze	arriving	basic	breakdown
adopted	and	art	basis	brief
advance	angle	article	be	bring
advantage	animal	arts	bear	broad
advantageously	announced	as	became	broken

Common Technical English Words

brought	central	clustering	components	construct
brown	century	clusterings	composite	constructed
bubble	certain	clusters	composition	constructing
budget	certainly	cm	comprise	construction
build	chain	code	comprises	contact
building	chairman	coding	comprising	contacts
buildings	chance	coefficient	computation	contain
built	chang	coefficients	computational	contained
burning	change	coffee	computations	containing
bus	changed	cold	compute	contains
buses	changes	collection	computed	contents
business	channel	college	computer	continue
busy	channels	color	computing	continued
but	chapter	column	conceived	continues
buy	character	columns	concentrated	continuing
by	characteristic	combinations	concentration	continuity
cable	characteristics	combined	concept	continuous
calculation	characters	combining	concepts	contract
california	charge	come	concern	contrast
call	charged	comes	concerned	control
called	charging	coming	concerning	controllable
calling	chart	command	conclude	controlled
calls	check	comment	conclusion	controlling
came	checked	comments	conclusions	controls
can	chemical	commerce	condition	convenient
cannot	chief	commercial	conditional	convention
capabilities	children	committee	conditions	conventional
capability	choice	common	conductivity	conventions
capable	choices	communicates	conductor	conversely
capacity	choose	communication	conductors	conversion
capital	chosen	communications	conference	convert
card	circle	communism	confidence	converted
care	circuit	community	configuration	converter
career	circuitry	compact	configurations	converting
careful	circuits	companies	congruent	cooperation
carefully	circumstances	company	connected	coordinate
carried	citizens	comparable	connection	coordinates
carries	city	compared	connections	core
carroll	civil	comparison	connects	corporation
carry	claim	compensate	consequently	correct
carrying	claims	compensation	consider	correctly
case	class	competent	considerable	correlation
cases	classes	competition	consideration	correspond
cathode	classified	compilers	considerations	corresponding
cathodes	clean	compiling	considered	corresponds
catholic	clear	complement	considering	cosines
caught	clearly	complete	consistent	cost
cause	clock	completed	consisting	costs
caused	close	completely	consists	could
causes	closed	completion	constant	council
causing	closely	complex	constants	count
cell	closer	complexity	constrained	counted
cells	closest	complicated	constraint	counter
center	closure	component	constraints	counting
centered	cluster			countries

Common Technical English Words

country	deep	development	distinguish	effectively
counts	defect	deviation	distortions	effects
couple	defects	deviations	distribution	efficiency
coupled	define	device	divided	efficient
coupling	defined	devices	division	efficiently
course	defines	diagonal	do	effort
courses	defining	diagram	document	efforts
court	definite	diagrams	documented	eight
cover	definition	dial	does	either
covered	definitions	diameter	doing	electric
create	degree	dictates	dollars	electrical
created	delay	dictionary	domain	electrode
credit	delayed	did	dominant	electrodes
crisis	delays	die	done	electronic
criterion	delete	dielectric	doped	electrons
critical	demand	difference	dotted	element
crop	density	differences	double	elements
cross	department	different	doubt	eliminate
crossovers	departures	differential	down	eliminated
cubic	depend	difficult	dr	else
culture	dependent	difficulties	draw	embedded
current	depending	difficulty	drawing	embodiment
currently	depends	diffusion	drawn	embodiments
curve	depicted	digit	drew	emotional
curves	depth	digital	drink	emphasis
cut	derived	digits	drive	employed
cutoff	deriving	dimension	driving	employees
cutting	descendant	dimensions	drop	empty
cycle	descendants	direct	dropped	enable
cycles	describe	directed	due	enabled
daily	described	direction	duration	enclosed
dallas	describing	directions	during	encoding
damage	description	directly	dust	encountered
danger	design	director	duty	encourage
dark	designated	discharge	dynamic	end
data	designed	discharges	each	ends
date	designs	disclosed	earlier	energy
day	desirable	discloses	early	engineering
days	desire	disclosure	earth	english
dead	desired	disconnected	ease	enough
deal	despite	discovered	easier	enter
debugging	destination	discrepancy	easily	entered
december	detail	discrete	east	entire
decide	detailed	discuss	easy	entirely
decided	details	discussed	economic	entities
decision	detected	discussion	economically	entitled
decisions	detection	display	economy	entries
declared	detects	displays	edge	envelope
decoded	determine	disposed	edges	environment
decoder	determined	dissimilarities	editing	epitaxial
decoding	determines	dissimilarity	editor	equal
decomposition	determining	distance	education	equally
decrease	develop	distances	educational	equals
decreases	developed	distinct	effect	equation
decreasing	developing	distinction	effective	equations

Common Technical English Words

equilibrium	expects	features	for	gating
equipment	expense	february	force	gave
equipped	expensive	feed	forced	general
equivalent	experience	feedback	forces	generality
equivalently	experiment	feel	foregoing	generalization
error	experimental	feeling	foreign	generalized
errors	experiments	feet	forest	generally
escape	experts	felt	form	generate
especially	explain	few	format	generated
essential	explained	fewer	formed	generates
essentially	explanation	field	former	generating
establish	explicit	fields	forming	generation
established	explicitly	fifteen	forms	generator
establishing	exponential	fifth	formula	geometries
estimated	exposed	fig	formulation	geometry
et	express	figs	forth	german
etc	expressed	figure	fortran	germany
evaluated	expression	figures	forward	get
even	extend	file	found	gets
evening	extended	filed	four	getting
event	extensive	files	fourth	give
events	extent	filled	fraction	given
ever	external	filling	frame	gives
every	extinguished	film	frames	giving
everybody	extra	filter	free	glass
everyone	extreme	filtering	freedom	global
everything	extremely	final	french	glow
evidence	fabricated	finally	frequencies	go
evidenced	face	financial	frequency	goal
evident	faces	find	frequently	goes
evil	facilities	finding	fresh	going
exact	fact	finds	friend	gone
exactly	factor	fine	friendly	good
examine	factors	finished	from	goods
examined	facts	finite	front	government
example	faculty	fire	full	governments
examples	failed	firm	fully	grant
exceeds	failure	first	function	granted
excellent	fair	fiscal	functional	graph
except	fairly	fit	functions	graphic
exception	faith	fitting	fund	great
exchange	fall	five	fundamental	greater
exclusive	falls	fixed	funds	greatest
executed	false	flexibility	furnish	greatly
executing	familiar	flip	furnishes	green
execution	families	floor	further	gross
exercise	family	flop	furthermore	ground
exhibit	far	flops	future	grounded
exist	fashion	flow	gain	grounds
existence	fast	follow	gap	group
existing	faster	followed	gas	grouped
exists	favor	following	gaseous	grouping
exit	fear	follows	gate	groups
expect	feasible	food	gates	grow
expected	feature	foot		growing

Common Technical English Words

growth	hospital	increasing	interests	kennedy
guaranteed	hot	increment	interface	kept
guarantees	hour	indeed	interior	key
guard	hours	independence	intermediate	kill
guess	house	independent	internal	kind
guests	housing	index	international	kinds
had	how	india	interpreted	knew
hair	however	indicate	interstage	know
half	human	indicated	interval	knowledge
hall	hundred	indicates	intervals	known
hand	idea	indicating	into	knows
handle	ideal	indicative	introduced	labeled
handling	ideas	individual	introduction	labelled
hands	identical	individuals	invention	labels
happen	identification	industrial	inverse	laboratories
happened	identified	industry	inverted	laboratory
hard	identify	inequality	inverter	lack
hardly	identity	influence	investigation	lag
hardware	idle	information	invites	laid
has	if	informed	involve	land
have	illustrate	initial	involved	language
having	illustrated	initialize	involves	languages
he	illustrating	initially	involving	large
head	illustration	initiate	ion	larger
health	illustrative	initiated	ions	largest
heard	image	input	is	last
heat	immediate	inputs	issue	late
heavily	immediately	inserted	issued	later
heavy	impact	inside	issues	latter
held	implement	instability	it	law
help	implementation	install	items	laws
helped	implemented	instance	its	lay
hence	implicitly	instances	itself	layer
here	implied	instantaneously	james	layers
herein	implies	instead	january	layout
hereinafter	imply	instruction	jersey	lead
hidden	importance	instructions	job	leader
hierarchical	important	insurance	jobs	leaders
high	impossible	integer	john	leadership
higher	improve	integers	joined	leading
highest	improved	integral	jointly	leads
highly	improvement	integrated	jr	learn
hill	impurities	intellectual	judge	learned
him	impurity	intended	judged	learning
himself	in	intensity	judgment	least
his	inches	intent	judicious	leave
historical	include	inter	july	leaving
history	included	interaction	jump	led
hold	includes	interactions	june	left
holding	including	interchange	just	leftmost
hole	incoming	interconnected	justice	legal
home	incorporated	interconnection	justification	length
hook	increase	interest	justify	less
hope	increased	interested	keep	let
horizontal	increases	interesting	keeping	letter

Common Technical English Words

letters	lot	meaningful	modifications	needs
level	low	means	modified	negative
levels	lower	meant	modular	neighboring
liability	lowest	measure	moment	neither
liable	machine	measured	money	net
liberal	machinery	measurement	monotone	network
library	machines	measurements	monotonic	neutral
lie	made	measures	monotonically	never
lies	magnetic	measuring	month	nevertheless
life	magnitude	mechanical	months	new
light	main	mechanism	moon	news
like	maintain	medical	moral	next
likely	maintained	meet	more	nice
limit	maintenance	meeting	moreover	night
limitations	major	meetings	morgan	nine
limited	majority	member	morning	no
limiting	make	members	most	nobody
limits	makes	membership	motor	node
line	making	memory	mounted	nodes
linear	man	men	move	noise
linearly	management	mentioned	moved	non
lines	manager	merely	moves	none
link	manner	merged	moving	nonlinear
links	manual	merging	much	nor
list	manufacturer	message	multi	normal
listing	manufacturers	met	multidimensional	normalized
lists	manufacturing	metal	multiple	normally
literature	many	metallization	multiplication	north
little	mapping	method	multiplier	not
live	march	methods	multipliers	notation
lived	margins	metric	multistage	note
load	mark	middle	multivariate	noted
loaded	marked	might	murder	notes
loading	market	miles	murray	nothing
local	mask	military	must	notice
locally	masking	million	mutually	noting
located	mass	mind	my	novel
location	master	minds	name	november
locations	match	minimal	named	now
log	matched	minimize	names	nuclear
logic	matches	minimizes	narrow	number
logical	matching	minimizing	nation	numbered
long	material	minimum	national	numbers
longer	materials	minister	nations	numeric
longest	mathematical	minor	natural	numerical
look	mathematically	minutes	naturally	object
looked	matrices	missing	nature	objective
looking	matrix	mistakes	near	objects
looks	matter	mixture	nearer	observation
loop	matters	mode	nearest	observe
loops	maximum	model	nearly	observed
lose	may	models	necessarily	obtain
loss	me	modern	necessary	obtained
losses	mean	modes	need	obvious
lost	meaning	modification	needed	obviously

Common Technical English Words

occasion	organized	perfect	polarity	private
occupied	orientation	perform	police	probability
occupy	oriented	performance	policies	probably
occur	origin	performed	policy	problem
occurred	original	performing	political	problems
occurrence	originally	performs	polynomial	proc
occurrences	orthogonal	perhaps	pool	procedure
occurring	other	period	poor	procedures
occurs	others	periodically	popular	process
october	otherwise	peripheral	population	processed
odd	ought	permanent	portion	processes
of	our	permit	position	processing
off	ourselves	permits	positive	processor
offer	out	permitted	possibility	produce
offered	outline	person	possible	produced
office	outlined	personal	possibly	produces
officer	output	personnel	post	producing
officers	outputs	persons	potential	product
official	outside	pertinent	power	production
officials	over	phase	powerful	products
often	overall	phases	practical	professional
oil	overlap	phenomena	practice	professor
old	own	philosophy	preceding	program
omitted	page	physical	precise	programmer
on	pain	pick	precision	programmers
once	pair	picked	predetermined	programming
one	pairs	pickup	predicted	programs
ones	paper	picture	prepared	progress
only	paragraph	piece	presence	project
onto	parallel	pieces	present	projections
open	parameter	place	presented	projects
opened	parameters	placed	presents	proof
opening	part	places	press	propagate
operate	partial	placing	pressure	propagated
operated	particular	plan	presumably	propagating
operates	particularly	planar	pretty	propagation
operating	parties	plane	prevent	proper
operation	partitioning	planned	prevents	properly
operations	parts	planning	previous	properties
opinion	party	plans	previously	property
opportunities	pass	plant	price	proportional
opportunity	passed	plants	prices	proposal
opposite	passes	plausible	primarily	proposals
optimal	passing	play	primary	proposed
optimization	past	played	prime	protect
optimum	patent	playing	priming	protection
options	path	plays	principal	proved
or	paths	please	principle	provide
order	patient	plotted	principles	provided
ordered	pattern	plurality	print	provides
ordering	patterns	plus	printed	providing
orders	pay	point	prints	provision
ordinary	peak	pointed	prior	public
organization	people	pointer	priority	publication
organizations	per	points	privacy	published

Common Technical English Words

pulse	receiving	remote	review	sciences
pulses	recent	remove	reviewed	scientific
pure	recently	removed	rich	scope
purely	recognize	removes	right	sea
purpose	recognized	repeat	rights	search
purposes	record	repeated	rise	second
put	records	repeatedly	risk	secondary
puts	recover	replaced	roads	seconds
quality	rectangular	replaces	role	secret
quantization	recursive	report	roles	section
question	recursively	reported	room	sections
questions	red	reports	root	security
quick	reduce	represent	roots	see
quickly	reduced	representation	rose	seeing
quiet	reduces	representations	rotation	seek
quite	reduction	representative	roughly	seeking
radio	refer	represented	round	seem
raised	reference	representing	rounded	seemed
random	references	represents	route	seems
randomly	referred	request	routed	seen
range	refers	require	routine	segment
rank	reflect	required	routines	segments
rapid	refused	requirement	routing	seldom
rapidly	regard	requirements	routings	select
rare	regarded	requires	row	selected
rate	regarding	requiring	rows	selection
rated	regardless	requisite	rule	selectively
rates	region	research	rules	self
rather	regions	reserved	run	semiconductor
ratio	register	reset	running	send
rational	registers	resetting	runs	sense
reach	regression	resistance	safe	sensitive
reached	regular	resistivity	said	sensitivity
reaction	rejected	resistor	sales	sent
read	relate	resistors	same	separate
readable	related	resolution	sample	separation
reader	relates	resources	samples	september
readers	relating	respect	sampling	sequence
readily	relation	respective	san	sequences
reading	relations	respectively	satisfied	sequentially
ready	relationship	response	satisfies	serial
real	relative	responsibility	satisfy	series
reality	relatively	responsible	save	serious
realize	relay	responsive	say	serve
realized	release	rest	saying	served
realizing	released	restored	says	serves
really	relevant	restrict	scalar	service
rear	reliability	restricted	scale	serviced
reason	relief	result	scaled	services
reasonable	remain	resulting	scaling	serving
reasonably	remainder	results	scan	set
reasons	remained	return	schematically	sets
receive	remaining	returned	school	setting
received	remains	returns	schools	seven
receives	remembered	reverse	science	several

Common Technical English Words

severe	social	stands	subsequently	talk
shall	society	start	subset	talked
shape	soft	started	subsets	talking
share	software	starting	substantial	tapes
sharp	solid	state	substantially	target
sheet	solution	stated	substituting	task
shift	solutions	statement	substitution	tax
shifting	solved	statements	substrate	teacher
shifts	some	states	succeeding	teachers
ship	someone	static	success	teaching
shop	something	station	successful	teachings
short	sometimes	stations	successive	team
shot	somewhat	statistical	successively	technical
should	somewhere	statistically	such	technique
show	son	status	suddenly	techniques
showed	soon	stay	sufficient	technology
showing	sophisticated	steady	sufficiently	teeth
shown	sort	step	suggest	telephone
shows	sought	steps	suggested	temp
side	sound	still	suggests	temperature
sides	source	stimuli	suitable	temporary
sign	sources	stimulus	suitably	ten
signal	space	stop	sum	tend
signals	spaced	stopped	summarize	tends
significance	spaces	storage	summary	term
significant	spacing	store	summing	terminal
significantly	spatial	stored	supervision	terminals
similar	speaking	stores	supervisory	terminated
similarly	special	straight	supplied	terminating
simple	specific	straightforward	supplies	termination
simpler	specifically	strange	supply	terms
simplest	specifications	strategy	support	test
simplicity	specified	stream	supporting	tested
simply	specify	street	suppose	testing
simultaneously	specifying	streets	supposed	tests
since	speech	strength	sure	texas
single	speed	stress	surface	text
sit	spent	strict	surprised	than
site	spirit	strong	surrounds	that
situation	spite	strongly	survey	the
situations	splitting	structure	switch	their
six	spoke	structures	switched	them
size	spot	struggle	switches	themselves
sizes	spread	student	switching	then
skilled	spring	students	symmetric	theorem
slightly	square	studied	symmetry	theoretical
slot	squares	studies	system	theory
slots	stable	study	systematic	there
slow	staff	style	systems	thereby
slowly	stage	subject	table	therefore
small	stages	subjected	tables	thereof
smaller	stand	subjects	take	these
smallest	standard	submitted	taken	they
snow	standards	subroutines	takes	thick
so	standing	subsequent	taking	thickness

Common Technical English Words

thin	traverse	unusual	voltages	word
thing	treated	up	volume	words
things	treatment	update	wait	work
think	tree	upon	waiting	worked
thinking	trees	upper	wall	workers
third	trial	us	walls	working
thirty	triangle	usage	want	works
this	tried	use	wanted	world
those	triggered	used	wants	worst
though	trip	useful	war	worth
thought	trivial	useless	was	would
three	trouble	user	washington	write
threshold	truck	users	water	writer
through	true	uses	wavelength	writers
throughout	truly	using	way	writing
thus	try	usual	ways	written
time	trying	usually	we	wrong
times	tube	utility	week	wrote
timing	turn	utilization	weeks	year
tip	turned	utilized	weight	years
tips	turning	utilizing	weighting	yes
title	turns	valid	weights	yet
to	twenty	validity	well	yield
today	twice	value	went	yields
together	two	values	were	york
toggled	type	van	western	you
told	types	variable	what	young
tolerances	typical	variables	whatever	your
tone	typically	variance	when	zero
too	ultimate	variation	whenever	zone
took	unaltered	varied	where	zones
top	unchanged	variety	whereby	
total	undefined	various	wherein	
touch	under	vary	wherever	
toward	underlying	vast	whether	
towards	understand	vector	which	
track	understanding	vectors	while	
trade	understood	verification	white	
traditional	unequal	verified	who	
traffic	unfortunately	version	whole	
train	unidirectionality	versions	whom	
training	unidirectionally	vertical	whose	
transaction	uniform	very	why	
transfer	union	via	wide	
transformation	unique	view	width	
transformations	unit	viewed	will	
transistor	united	vis	william	
transistors	units	visible	willing	
transition	unity	vision	wind	
translated	universe	visit	wise	
translation	university	vital	wish	
transmission	unless	vocational	with	
transmit	unrelated	voice	within	
transmitted	unstable	voids	without	
travel	until	voltage	women	