SD202 - Lab 2 - Spark
Student: José Lucas Barretto

Question 1) To compute the number of occurrences for each word, the only necessary change to the previous code was to map each word to the value 1 (in the previous code, each word was mapped to a value of 2). This way, when we sum the values for each (word, 1) pair, we will get the number of occurrences for that word. I obtained the following results:

Steven : 1	Jobs : 23	(/dʒɒbz/; : 1	was : 33	an : 10	American : 2	business : 2
inventor, : 1	chief : 1	executive : 1	officer : 1	co-founder : 2	of : 41	Apple : 11
Inc.; : 1	CEO : 3	shareholder : 1	The : 6	Walt : 1	board : 1	following : 1
are : 2	widely : 1	as : 13	revolution : 1	: 10	in : 41	put : 2
adoption : 4	at : 6	he : 19	raised : 2	Bay : 2	during : 2	1960s.[4] : 1
College : 1	before : 1	dropping : 1	India : 1	enlightenment : 1	studying : 1	Jobs's : 8
FBI : 1	stated : 4	used : 1	marijuana : 1	LSD : 2	college.[7] : 1	once : 1
told : 3	"one : 1	two : 3	three : 1	things" : 1	his : 15	life.[8] : 1
co-founded : 1	Wozniak's : 1	computer. : 1	visionaries : 1	wealth : 1	year : 1	II, : 2
mass-produced : 2	after : 4	tour : 1	commercial : 1	potential : 1	Xerox : 1	Alto, : 1
mouse-driven : 1	graphical : 1	led : 1	development : 3	1983, : 1	breakthrough : 1	GUI, : 1
introduced : 1	desktop : 1	feature : 1	vector : 1	long : 1	power : 1	struggle, : 1
forced : 1	out : 3	1985.[9] : 1	Apple, : 1	took : 2	members : 1	NeXT, : 1
platform : 1	specialized : 1	computers : 1	higher-education : 1	visual : 1	when : 4	funded : 1
Lucas's : 1	Lucasfilm : 1	1986.[10] : 1	new : 2	company, : 1	would : 6	eventually : 3
produce : 1	computer-animated : 1	Story—an : 1	event : 1	possible : 1	financial : 1	Within : 1
months : 2	revived : 1	verge : 1	"Think : 1	different" : 1	campaign, : 1	worked : 1
closely : 1	designer : 1	Jonathan : 1	Ive : 1	line : 1	have : 3	larger : 1
iMac, : 1	iTunes : 2	iPhone, : 1	App : 1	iPad. : 1	2001, : 1	OS : 3

replaced : 1	based : 1	platform, : 1	modern : 1	Unix-based : 1	diagnosed : 1	tumor : 1
2003 : 1	respiratory : 1	related : 1	Family : 1	Jandali : 8	adoptive : 3	Clara : 8
Hagopian. : 1	"John" : 1	الفتاح : 1	March : 1	1931), : 1	Homs, : 2	into : 1
Arab : 1	household.[11] : 1	is : 1	son : 2	millionaire : 1	go : 1	college : 2
traditional : 1	housewife.[11] : 1	undergraduate : 1	University : 2	Lebanon, : 1	student : 1	jail : 1
political : 2	law, : 1	decided : 3	economics : 1	latter : 1	subject : 1	where : 1
Schieble : 8	1, : 1	1932), : 1	Catholic : 1	Swiss : 1	German : 1	Wisconsin.[12][11][13] : 1
doctoral : 1	candidate, : 1	teaching : 1	assistant : 1	taking, : 1	although : 1	both : 2
age.[14] : 1	Mona : 1	full : 1	sister, : 1	notes : 1	her : 5	daughter : 2
Jandali: : 1	wasn't : 1	Muslim. : 1	But : 1	there : 2	Arabs : 1	unusual."[14] : 1
Walter : 1	Isaacson, : 1	official : 2	biographer, : 1	father : 3	cut : 1	completely" : 1
continued : 1	Reinhold : 1	(1922–1993), : 1	Calvinist : 1	sometimes : 1	abusive" : 1	father.[12] : 1
family : 4	lived : 1	Germantown, : 1	Wisconsin.[12][15] : 1	ostensible : 1	Dean; : 1	dropped : 1
high : 2	school, : 2	around : 1	several : 1	years : 2	looking : 1	work.[12][15] : 1
joined : 1	United : 2	States : 2	Coast : 2	machinist.[15] : 1	World : 1	leave : 1
docked : 1	They : 1	engaged : 1	ten : 1	days : 1	married : 2	1946.[12] : 1
Armenian : 1	before, : 1	but : 3	husband : 1	killed : 1	war. : 1	series : 1
moves, : 1	settled : 1	1952.[12] : 1	hobby, : 1	rebuilt : 1	career : 1	"aggressive, : 1
personality."[15] : 1	start : 1	halted : 1	pregnancy, : 1	leading : 1	them : 3	consider : 1
1955.[12] : 1	1954 : 1	summer : 1	very : 1	love : 2	... : 3	forbade : 1
And : 2	give : 1	adoption."[17] : 1	aggravate : 1	felt : 3	young : 1	against : 1
child : 2	wedlock : 1	single : 1	illegal : 1	only : 2	women : 1	According : 1
Jandali, : 1	deliberately : 1	process: : 1	"without : 1	upped : 1	move : 1	anyone : 1
bring : 1	thought : 1	this : 1	best : 1	"doctor : 1	sheltered : 1	quietly : 1

gave : 1	well-educated, : 1	wealthy."[18] : 1	changed : 1	however, : 1	adopt : 1	boy : 1
placed : 2	Jobs, : 1	neither : 1	refused : 1	sign : 1	papers.[12] : 1	She : 1
matter : 1	attempt : 1	different : 1	promised : 1	attend : 1	college.[12] : 1	When : 2
admitted : 1	Chrisann : 2	Brennan, : 1	frightened : 1	scared : 1	take : 2	away : 1
we : 2	case, : 1	mistake. : 1	him."[18] : 1	mother's : 1	already : 1	say : 1
loved : 1	Clara.[19] : 1	Laurene : 1	"he : 1	really : 2	blessed : 1	parents."[19] : 1
become : 1	upset : 1	parents" : 1	"were : 1	regard : 1	"my : 1	sperm : 2
bank. : 1	That's : 1	harsh, : 1	just : 1	way : 1	bank : 1	more."[12] : 1
"I : 1	am : 1	dad. : 1	Mr. : 1	are, : 1	place."[17] : 1	Paul : 11
February : 2	24, : 2	1955 : 1	– : 1	October : 2	5, : 2	2011) : 1
entrepreneur, : 1	magnate, : 1	and : 53	industrial : 1	designer. : 1	He : 5	the : 66
chairman, : 2	(CEO), : 1	majority : 1	Pixar;[3] : 1	a : 45	member : 1	Disney : 1
Company's : 1	directors : 1	its : 2	acquisition : 1	Pixar; : 1	founder, : 1	NeXT. : 2
Steve : 6	Wozniak : 2	recognized : 1	pioneers : 1	microcomputer : 1	1970s : 1	1980s. : 1
born : 2	San : 8	Francisco : 6	to : 42	parents : 4	who : 5	had : 11
him : 6	up : 6	for : 12	birth; : 1	Area : 2	then : 3	attended : 1
Reed : 1	1972 : 1	out,[5] : 1	traveled : 2	through : 1	1974 : 1	seeking : 1
Zen : 1	Buddhism.[6] : 1	declassified : 1	report : 1	that : 22	acquaintance : 1	knew : 1
while : 1	reporter : 1	taking : 1	or : 1	most : 1	important : 1	did : 5
1976 : 1	sell : 1	I : 6	personal : 2	gained : 1	fame : 1	later : 3
one : 1	first : 6	highly : 1	successful : 1	computers. : 1	In : 6	1979, : 1
PARC, : 1	saw : 1	which : 2	user : 1	interface : 1	(GUI). : 1	This : 1
unsuccessful : 1	Lisa : 1	followed : 1	by : 4	Macintosh : 2	1984. : 1	addition : 2
being : 1	computer : 3	with : 15	sudden : 1	rise : 1	publishing : 1	industry : 2

1985 : 1	LaserWriter, : 1	laser : 1	printer : 1	graphics. : 1	Following : 1	After : 3
leaving : 1	few : 2	found : 1	company : 1	state-of-the-art : 1	markets. : 1	addition, : 2
helped : 1	initiate : 1	effects : 1	spinout : 1	graphics : 1	division : 1	George : 1
Pixar, : 1	fully : 1	film, : 1	Toy : 1	made : 3	part : 1	because : 1
support. : 1	1997, : 1	merged : 1	merger, : 1	became : 2	former : 1	company; : 1
bankruptcy. : 1	Beginning : 1	1997 : 1	advertising : 1	develop : 1	products : 1	cultural : 1
ramifications: : 1	Store, : 3	iPod, : 1	original : 1	Mac : 2	completely : 1	X, : 1
on : 6	NeXT's : 1	NeXTSTEP : 1	giving : 1	foundation : 1	time. : 1	pancreatic : 1
neuroendocrine : 1	died : 1	2011, : 1	arrest : 1	tumor. : 1	Background : 1	biological : 5
were : 10	Abdulfattah : 2	Joanne : 5	Schieble. : 1	His : 2	father, : 2	(Arabic: : 1
عبد : 1	الجندي) : 1	(b. : 2	15, : 1	grew : 4	Syria : 1	Muslim : 1
self-made : 1	not : 8	mother : 3	While : 1	Beirut, : 1	activist : 1	spent : 2
time : 2	activities.[11] : 1	Although : 1	initially : 1	wanted : 3	study : 2	science.[11] : 1
pursued : 1	PhD : 1	Wisconsin, : 1	met : 1	Carole : 1	August : 1	descent, : 1
farm : 2	As : 2	course : 1	same : 1	Simpson, : 1	maternal : 1	grandparents : 1
happy : 1	their : 5	dating : 1	"it : 1	Middle-Eastern : 1	so : 3	much : 2
lot : 1	Michigan : 1	Wisconsin. : 1	So : 1	it's : 2	additionally : 1	states : 1
Schieble's : 2	"threatened : 1	off : 1	if : 1	she : 6	relationship.[12] : 1	household,[15] : 1
"alcoholic : 1	bore : 1	resemblance : 1	James : 1	tattoos, : 1	Midwest : 1	1930s : 1
Guard : 2	engine-room : 1	War : 1	ship : 1	Francisco.[15] : 1	bet : 1	find : 1
wife : 2	promptly : 1	went : 1	blind : 1	date : 1	Hagopian : 1	(1924–1986). : 1
Clara, : 1	immigrants, : 1	been : 3	Francisco's : 1	Sunset : 1	District : 1	cars, : 1
"repo : 1	man", : 1	suited : 1	tough : 1	Meanwhile, : 1	attempts : 1	ectopic : 1
pregnant : 1	Syria. : 2	has : 2	"was : 2	sadly, : 1	tyrant, : 1	marry : 1

me, : 2	from : 2	me : 2	baby : 5	biographer : 1	dying : 1	time, : 1
want : 3	him, : 1	23 : 1	they : 5	too : 2	marry.[12] : 1	strong : 1
stigma : 1	bearing : 1	raising : 1	it : 2	mother, : 1	abortions : 1	dangerous, : 1
option : 1	1954.[15] : 1	involve : 1	telling : 1	left : 1	without : 1	knowing, : 1
including : 1	shame : 1	onto : 1	everyone."[17] : 1	herself : 1	care : 1	unwed : 1
mothers, : 1	delivered : 1	babies, : 1	arranged : 1	closed : 1	adoptions."[12] : 1	birth : 1
1955, : 1	chose : 1	couple : 3	"Catholic, : 1	mind, : 1	girl : 1	instead.[18] : 1
blue : 1	collar : 1	whom : 1	education, : 1	court : 1	family[18] : 1	consented : 1
releasing : 1	girlfriend, : 1	17-year-old : 1	[Steve] : 1	six : 1	life : 1	going : 1
me. : 1	Even : 1	won : 1	difficult : 1	return : 1	shared : 1	comment : 1
Steve, : 1	aware : 1	that[18] : 1	deeply : 1	indulged : 1	Many : 1	later, : 1
also : 2	noted : 1	having : 1	referred : 2	"adoptive : 1	my : 1	1,000%."[12] : 1
With : 1	parents, : 1	egg : 1	was, : 1	thing, : 1	nothing : 1	Mrs. : 1

Question 2) Results for top 5 words in number of occurrences.

```
the : 66
and : 53
a : 45
to : 42
of : 41
```

Question 3) Results for top 5 words with largest number of occurrences among the words containing at least 5 characters.

```
Apple : 11
Jobs's : 8
Jandali : 8
Clara : 8
Schieble : 8
```

Question 4) Results for top 10 pages in Wikipedia with largest in-degree.

```
United States : 8145
France : 7799
```

```
Communes of France : 5740
Departments of France : 5299
Regions of France : 4064
City : 3832
Romania : 3527
Category:Rivers in Romania : 2978
Tributary : 2799
England : 2277
```