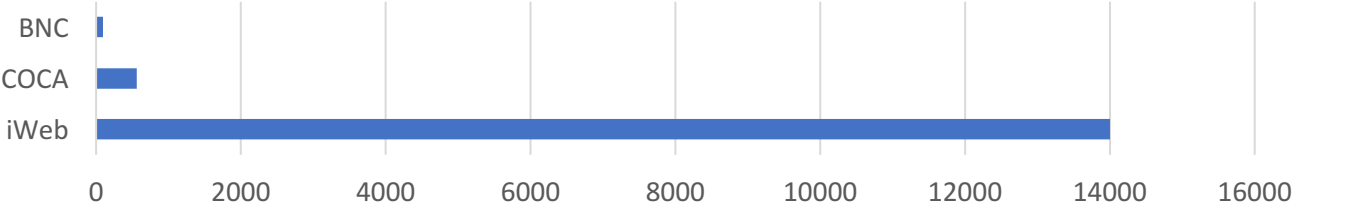


The iWeb corpus (released May 2018)

The BYU corpora (<https://corpus.byu.edu>) are the world’s **most widely-used corpora**. The iWeb corpus (<https://corpus.byu.edu/iweb>) -- released in May 2018 – is perhaps the most useful of the BYU corpora for language learners, teachers, and researchers. There are a number of features that make this such a great corpus.

First, consider the **size** of the corpus. iWeb is about 14 billion words in size, which makes about 25 times as big as COCA (560 million words) and about 140 times as big as the BNC (100 million words).



Second, in just 4-5 seconds you can create a “**Virtual Corpus**” for any topic (e.g. chocolate, basketball, solar power, investments, Buddhism, Harry Potter, or any other topic).

solar energy

[POS]

Find matching strings

Reset

Texts/Virtual

Sort/Limit Op

SEARCH ALL

FIND WEBSITES

MY CORPORA

Create corpus

Edit corpora

(Optional) SAVE AS: solar_energy

OR ADD TO: --SELECT--

SUBMIT

RESET

HELP

HELP	20	TEXT	# WORDS	# HITS	RELEVANCE	PER MILLION WORDS	KEYWORDS
1	<input checked="" type="checkbox"/>	ENERGY.COM	25099	387	15,418.9		solar-powered, blinking, solar, inverter, powered, sunlight, panel, rebate
2	<input checked="" type="checkbox"/>	VIVINTSOLAR.COM	26555	184	6,929.0		solar, financing, residential, lease, finance, quarter, operating, equity
3	<input checked="" type="checkbox"/>	SEIA.ORG	86437	372	4,303.7		photovoltaic, solar, renewable, electricity, utility, energy, installation, capacity
4	<input checked="" type="checkbox"/>	SOLARENERGY.ORG	75142	267	3,553.3		solar, renewable, instructor, energy, training, lab, installation, professional
5	<input checked="" type="checkbox"/>	DASOLAR.COM	24506	85	3,468.5		geothermal, solar, turbine, renewable, electricity, wind, audit, energy
6	<input checked="" type="checkbox"/>	EVENTS.SOLAR	24186	83	3,431.7		exhibitor, solar, attendee, energy, conference, industry, power, attend
7	<input checked="" type="checkbox"/>	CLEANENERGYAUTHORITY.COM	90348	266	2,944.2		photovoltaic, solar, renewable, installer, incentive, homeowner, utility, electricity
8	<input checked="" type="checkbox"/>	SOLARPOWERISTHEFUTURE.COM	28342	71	2,505.1		solar, photovoltaic, sunlight, heating, panel, homeowner, electricity, battery
9	<input checked="" type="checkbox"/>	SOLARRESERVE.COM	10185	25	2,454.6		molten, collapsible, solar, renewable, electricity, energy, storage, emission
10	<input checked="" type="checkbox"/>	ENERGYSAGE.COM	148740	360	2,420.3		solar, installer, electricity, homeowner, renewable, panel, incentive, roof
11	<input type="checkbox"/>	SUNRUN.COM	119391	245	2,052.1		solar, homeowner, electricity, renewable, utility, saving, panel, affordable

Once you’ve created the virtual corpus, you can then search within it (as if it were its own stand-alone corpus), and you can compare results from different virtual corpora that you’ve created. In just 2-3 seconds, you can also find a list of keywords for the virtual corpus, as with the “solar energy” corpus below:

SOLAR_ENERGY1 [539,540 WORDS, 10 WEBSITES] (0.0% OF TOTAL) NOUN VERB ADJ ADV N+N ADJ+N [ALL CORPORA] SAVE LIST

HELP	ENTRY	WORD (CLICK FOR CONTEXT)	FREQ	# WEBSITES	SPECIFIC FREQ 90 5 WEBSITES	ENTIRE CORPUS	EXPECTED
1	↗	MEGAWATT	159	6	922.8	4,471	0.2
2	↗	SOLAR	1367	10	850.0	41,729	1.6
3	↗	KILOWATT	179	6	524.3	8,859	0.3
4	↗	REBATE	377	6	204.6	47,809	1.8
5	↗	HOMEOWNER	478	8	195.3	63,509	2.4
6	↗	INSTALLER	730	7	184.3	102,795	4.0
7	↗	METERING	258	6	180.2	37,155	1.4
8	↗	INVERTER	186	8	105.4	45,801	1.8
9	↗	INCENTIVE	602	6	96.5	161,921	6.2
10	↗	ELECTRICITY	1454	10	89.7	420,622	16.2

1	sela.org	A B C	fossil fuel power plants. # Solar provides benefits to the environment and society. Solar does not produce greenhouse gas emissions, it reduces air pollution
2	vivintsolar.com	A B C	36149115 5149115/ # The Texas market presents a compelling opportunity for Vivint Solar. According to the U.S. Energy Information Administration(EIA), Te
3	solarenergy.org	A B C	be tailored to grade level. Please contact us for details. # SEI's Solar In the Schools (SIS) outreach program starts off its fall season of renewable
4	cleanenergyauthority.com	A B C	and quickly offset the installation cost. Going solar is simple. At Main St. Solar, we handle all of the large and small details for you. Our goal
5	energysage.com	A B C	that these incentive programs will be around. # Every year you don't go solar means thousands of dollars in missed savings. If you wait to go solar,

Third, with iWeb you have access to the **wide range of searches** that you have for all of the other BYU corpora, including: words, phrases, substrings, lemmas, part of speech, synonyms, and customized wordlists. For example, the search **BUY * ADJ @CLOTHES** takes just 1-2 seconds to search through the 14 billion words to find strings like the following (and it doesn't require learning unnecessarily convoluted search syntax).

		CONTEXT	ALL FORMS (SAMPLE): 100 200 500	FREQ	TOTAL 2,079 UNIQUE 1,281 +
1	<input type="checkbox"/>		BUY A NEW BRA	43	<div></div>
2	<input type="checkbox"/>		BUY SOME NEW SHOES	41	<div></div>
3	<input type="checkbox"/>		BUY A NEW HAT	33	<div></div>
4	<input type="checkbox"/>		BUY A NEW BELT	29	<div></div>
5	<input type="checkbox"/>		BOUGHT A NEW BELT	24	<div></div>
6	<input type="checkbox"/>		BUYING A NEW HAT	24	<div></div>
7	<input type="checkbox"/>		BUY NEW RUNNING SHOES	21	<div></div>
8	<input type="checkbox"/>		BUY THE RIGHT SHOES	16	<div></div>

1	theladders.com	A B C	forgot how to tie a tie, so your wife has to buy you clip-on ties for days you are required to wear one. You wear one black sock and
2	naturallivingideas.com	A B C	flowers, and a soft rug or pillow. 2. Buy Some New Shoes # The right pair of shoes can seriously help to reduce the impact that each
3	duluthtrading.com	A B C	going back Saturday to buy 2 more pairs! Will also be buying some cooling underwear! Would recommend to anyone. # I go through khaki pants pretty q
4	duluthtrading.com	A B C	n't need breaking in. Where else in North America can you buy a waterproof shoe with leather uppers for under \$100? Keep making them and I'll keep bu
5	sneakerfreaker.com	A B C	and day despite the fact that they didn't fit. I bought the thinnest socks available (I stopped just short of wearing stockings) stood around a lot,
6	keenfootwear.com	A B C	FINALLY I found the perfect fit. For years, I could buy a Keen shoe and not even have to break them in - they were just the perfect comfortable
7	eastoftheweb.com	A B C	, ' how' bout a cash advance? I got ta buy some new shoes.' It'll be a tough sell, considering he just gave me twenty
8	promocodes.com	A B C	chic or preppy look. Did those summer sandals wear out? Buy them new shoes or sneakers with 20% off at Shoebuy.com. # 2. Cell phones # While
9	gwynniebee.com	A B C	. But I know from personal experience, having to find and buy a special bra for one dress is a pain. Hence I came up with an idea that

Because of iWeb's advanced architecture, even searches for very general searches like **NOUN + NOUN** or **VERB * ADJ NOUN** take just 1-2 seconds to search through the 14 billion word corpus:

		USE LARGE NGRAMS [?]	FREQ	TOTAL 132,044,991 UNIQUE 17,140 +
1	<input type="checkbox"/>	PER CENT	778603	<div></div>
2	<input type="checkbox"/>	HEALTH CARE	549878	<div></div>
3	<input type="checkbox"/>	CUSTOMER SERVICE	526645	<div></div>
4	<input type="checkbox"/>	CREDIT CARD	471662	<div></div>
5	<input type="checkbox"/>	WORLD WAR	331482	<div></div>
6	<input type="checkbox"/>	WEB SITE	322772	<div></div>
7	<input type="checkbox"/>	LAW ENFORCEMENT	275212	<div></div>
8	<input type="checkbox"/>	EMAIL ADDRESS	256952	<div></div>
9	<input type="checkbox"/>	STATE UNIVERSITY	251137	<div></div>
10	<input type="checkbox"/>	SIDE EFFECTS	250283	<div></div>
11	<input type="checkbox"/>	FAMILY MEMBERS	234739	<div></div>
12	<input type="checkbox"/>	CLIMATE CHANGE	229800	<div></div>

1	<input type="checkbox"/>	GO A LONG WAY	47879	<div></div>
2	<input type="checkbox"/>	GOES A LONG WAY	30969	<div></div>
3	<input type="checkbox"/>	TAKE A CLOSER LOOK	30053	<div></div>
4	<input type="checkbox"/>	COME A LONG WAY	28340	<div></div>
5	<input type="checkbox"/>	PLAY AN IMPORTANT ROLE	25440	<div></div>
6	<input type="checkbox"/>	MAKE A BIG DIFFERENCE	23344	<div></div>
7	<input type="checkbox"/>	TAKE A LONG TIME	18846	<div></div>
8	<input type="checkbox"/>	PLAYS AN IMPORTANT ROLE	18244	<div></div>
9	<input type="checkbox"/>	MAKE A HUGE DIFFERENCE	16867	<div></div>
10	<input type="checkbox"/>	PLAY A KEY ROLE	14955	<div></div>
11	<input type="checkbox"/>	TAKE A DEEP BREATH	13970	<div></div>
12	<input type="checkbox"/>	OFFERS A WIDE RANGE	13555	<div></div>

A unique feature of iWeb, which makes it very useful for language learners and teachers, is the ability to **browse** through a list of the **top 60,000 words** (lemmas) in the corpus, and then to see an extremely wide range of information on each of these words.

Word form	<input type="text"/>
Part of speech	<input checked="" type="checkbox"/> NOUN <input checked="" type="checkbox"/> VERB <input checked="" type="checkbox"/> ADJ <input checked="" type="checkbox"/> ADV <input checked="" type="checkbox"/> OTHER <input type="checkbox"/> ALL
Range	<input type="text" value="5145"/> - <input type="text"/>
Pronunciation	Rhymes with <input type="text"/> Type <input type="text" value="EXACT"/>
Syllables / stress	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ×

For example, the following are just a few examples of high frequency words (about word #5000 in the 60,000 word list), medium frequency (~25,000), and low frequency (~45,000) words. For each word in the list, users can hear the word **pronounced**, see **videos** with that word in the text, find related **images** from Google Images, and see a **translation** for their preferred target language.

	RANK	FREQ	Word	PoS	Audio	Video	Image	DE
1	5004	198190	tasty	ADJ				
2	5005	198064	spacious	ADJ				
3	5006	198045	trim	VERB				
4	5007	198027	loving	ADJ				
5	5008	197861	chaos	NOUN				

	RANK	FREQ	Word	PoS	Audio	Video	Image	DE
1	25168	7598	babble	VERB				
2	25169	7598	transform	NOUN				
3	25170	7598	opossum	NOUN				
4	25171	7598	tell-tale	ADJ				
5	25172	7596	pagination	NOUN				

	RANK	FREQ	Word	PoS	Audio	Video	Image	DE
1	45327	1266	kombu	NOUN				
2	45328	1266	immunosuppressant	ADJ				
3	45329	1266	reunify	VERB				
4	45330	1266	five-run	ADJ				
5	45331	1266	schlock	NOUN				

It is even possible to search the 60,000 word list by **pronunciation** (very helpful, because of difficult English spelling). For example, the following is a partial list of two syllable words (accented on the second syllable) that rhyme with *stay*:

Pronunciation	Rhymes with <input type="text" value="stay"/> Type <input type="text" value="EXACT"/>
Syllables / stress	● ○ ○ ○ ○ ○ ○ ○ ○ ○ ×

1	6253	135063	obey	VERB				
2	6717	119213	ballet	NOUN				
3	7088	107874	repay	VERB				
4	7399	99364	buffet	NOUN				
5	7844	89645	decay	NOUN				
6	8755	73424	betray	VERB				
7	8785	72975	replay	NOUN				
8	8977	70291	bouquet	NOUN				

Each of the **top 60,000 words** in the corpus has a “**home page**” such as the following, with links to other pages with more information. Users can save words to a “favorites” list for later review, and go back through a history of all of their “word” pages

trail

(NOUN)

See: [VERB](#)

#1619

[Favorites list](#)

[History](#)

Rank #1-60,000

1. a track or mark left by something that has passed 2. a path or track roughly blazed through wild or hilly country 3. evidence pointing to a possible solution

Definition

Pronunciation

[PlayPhrase](#)

[YouGlish](#)

[Yarn](#)

[Links to videos](#)

DE: [Google](#)

[WordRef](#)

[Reverso](#)

[Linguee](#)

[Google images](#)

[Links to translations](#)

[SYNONYMS](#) (more)

path

[road](#)

[track](#)

[path](#)

[trail](#)

[route](#)

[trajectory](#)

[footpath](#)

track

[trail](#), [trace](#)

[TOPICS](#) (more)

Words that co-occur in 22 million web pages

[park](#), [mountain](#), [hike](#), [hike](#), [lake](#), [river](#), [road](#), [forest](#), [bike](#), [trail](#), [creek](#), [route](#), [ride](#), [rock](#), [valley](#), [climb](#), [ride](#), [north](#), [parking](#), [visitor](#)

[COLLOCATES](#) (more)

Words that co-occur nearby

NOUN

[mile](#), [park](#), [mountain](#), [bike](#), [road](#), [lake](#), [creek](#), [river](#)

VERB

[hike](#), [follow](#), [walk](#), [ride](#), [blaze](#), [bike](#), [cross](#), [explore](#)

ADJ

[appalachian](#), [scenic](#), [steep](#), [paved](#), [inca](#), [historic](#), [rocky](#), [indian](#)

ADV

[along](#), [left](#), [uphill](#), [steeply](#), [nearby](#), [downhill](#), [moderately](#), [paralle](#)

NOUN + NOUN (more)

trail NOUN	trail system • trail blazers • trail head • trail mix • trail map • trail junction • trail running • trail bike • trail network • trail ride • trail run • trail runners • trail camera • trail heads • trail guide • trail conditions • trail shoes • trail shoe • trail race • trail users •
NOUN trail	campaign trail • audit trail • creek trail • paper trail • nature trail • bike trail • loop trail • hiking trails • bike trails • nature trails • river trail • rail trail • ridge trail • hiking trail • mountain trail • crest trail • walking trail • freedom trail • heritage trail • lake trail •

CLUSTERS (more) 2, 3, 4 word strings

trail •	trail in • trail to • trails in • trail for • trail on • trail system • trail from • trail with
• trail	hiking trails • campaign trail • appalachian trail • audit trail • creek trail • paper trail • nature trail • walking trails
trail ••	trail of tears • trail as it • trail is not • trail of blood • trail leads to • trail takes you • trail of destruction • trail passes through
•• trail	on the trail • along the trail • on the trails • to the trail • miles of trails • follow the trail • on a trail • off the trail
trail •••	trails of cold steel • trails in the area • trails in the sky • trail to the top • trail to the right • trails in the park • trail on the left • trail to machu picchu
••• trail	on the campaign trail • end of the trail • out on the trail • miles of hiking trails • part of the trail • section of the trail • side of the trail • on the appalachian trail

WEBSITES / VIRTUAL CORPORA (more) Websites that have this as a “keyword”

traillink.com • cnyhiking.com • americantrails.org • redwoodhikes.com • ashevilletrails.com • atlantatrails.com • trailforks.com • ontariotrails.on.ca • brucetrail.org • nynjtc.org •
--

CONCORDANCE LINES (more) The word in context (to see patterns of use)

1	alpineelements.co.uk	awaits visitors , with an extensive lift system , numerous trails	and an array of activities from sedate to gung-ho . Expect a
2	financialforce.com	always be in balance , the application provides ironclad audit trails	and complete drill down to originating transactions .
3	fgl.com	in memory when you buy the dragon , and with some trail	and error an attacker will figure out that changing a certain
4	ovlix.com	Bluff , Devil 's Glen , area ski resorts , Bruce Trail	and more A portion of the land is enrolled in the
5	overlandsummers.com	friends and having fun as you and your group explore on trails	and on the water . # In western Maine , there 's
6	windsorstar.com	council approved a plan Wednesday to build 700 kilometres of trails	and paths over the next 20 years . # Councillors adopted the
7	roadtrailrun.com	shoe for " urban " racing , for varied terrain including trails	and pavement I think many will run and race these in
8	attackpoint.org	the late 1980s , I recall , youth courses went off trail	and requires use of topography to navigate , although there
9	whiteblaze.net	anyone hiked the trail between the Monadnock-Sunapee Greenway Trail	and the New England Trail ? Is it reasonably well blazed ?
10	mtbr.com	Then what ? They 're still in the middle of the trail	and you re traveling out of control if you ca n't stop
11	coloradodirectory.com	. # SPEND A DAY HIKING : The Meridian and Tanglewood Trails	are popular for wildlife and bird watching ! # ROMANCE
12	earthrangers.com	: # This project will support the installation of signs along trails	around Lake Erie , which will raise awareness about the American
13	philadelphiazoo.org	As you explore the 1.5 acres of rocky outcroppings and winding trails	at Carnivore Kindom , keep your eyes peeled for meat-eating
14	whale.to	minds without using implants . # In 1991 , a paper trail	began to appear when the CIA connected U.S. government Global
15	bike-urious.com	enjoy . # Theres something special about micro-displacement trail	bikes So many of us have a sentimental attachment to these
16	ordinarytraveler.com	lots of water and shoes with good traction , as the trail	can be shipped . Stop by the small town of Hawi on
17	olympicnationalparks.com	Full Day : # Nearby are " easy access " hiking trails	designed for people of all ages and abilities . Hiking is ideal
18	swnewsmedia.com	by members of the community . At the end of the trail	enjoy a s'mores cookout over an open fire and archery pumpkin
19	funnewjersey.com	the most challenging trails in the state , but also have trails	for beginners Challenge yourself to hiking their highest
20	outherecolorado.com	trail up Blanca Peak 's Northwest Face route , Follow the trail	for 1.7 miles to 12,730-foot Crater Lake . Continue up the steep

Each of the top 60,000 words also has more detailed pages, including a “dictionary” page, related topics, collocates, clusters, websites, and concordance lines. Samples of each of these pages are given below.

DICTIONARY page

Includes a **definition**, links to Google **Images**, **pronunciation**, **videos**, and **translations** (to desired language) at (up to) four different sites. Also includes **synonyms** and words with more specific meanings (**hyponyms**) and more general meanings (**hypernyms**) (both from WordNet). It also includes **frequency information** (including rank order, number of tokens, and two “range” measures of how well the word is spread throughout the 95,000 websites. Finally, it also includes the frequency of the different **forms** of the lemma (e.g. verb forms), and **related words**.

1. a raised horizontal surface 2. a document stating the aims and principles of a political party 3. a woman's shoe with a very high thick sole 4. any military structure or vehicle bearing weapons 5. the combination of a particular computer and a particular operating system

[PlayPhrase](#) [YouGlish](#) [Yarn](#)

[ZH-CN: Google](#) [WordRef](#) [Reverso](#) [Linguee](#)

SYNONYMS (more)

[policy](#) [policy](#), [proposal](#), [platform](#), [manifesto](#) [stage](#) [stand](#), [platform](#), [podium](#), [dais](#)

MORE SPECIFIC MEANING (click on blue word)

footplate	(British) the platform in the cab of a locomotive on which the engineer stands to operate the controls
turntable	a circular horizontal platform that rotates a phonograph record while it is being played
stage	a large platform on which people can stand and can be seen by an audience
runway	a narrow platform extending from the stage into the audience in a theater or nightclub etc.
footboard	a narrow platform on which to stand or brace the feet
dock	a platform built out from the shore into the water and supported by piles
block	a platform from which an auctioneer sells
scaffold	a platform from which criminals are executed (hanged or beheaded)

FREQUENCY INFORMATION ?

977 Freq: 1,686,440 Range: 0.53 Range10: 0.29

WORD FORMS

[platform](#) (1,165,033), [platforms](#) (426,932)

RELATED WORDS

[platform](#) (n) , [cross-platform](#) (j) , [multi-platform](#) (j) , [platform-specific](#) (j) , [platform-independent](#) (j)

MORE GENERAL MEANING (click on blue word)

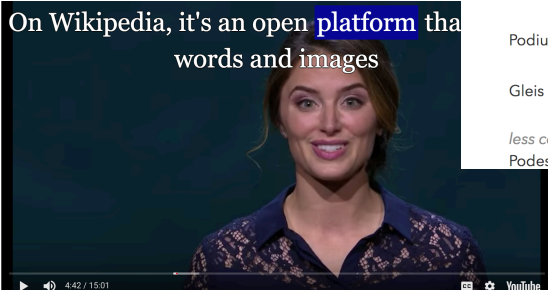
level	a flat surface at right angles to a plumb line
structure	a thing constructed
document	writing that provides information (especially information of an official nature)
construction	a thing constructed
shoe	footwear shaped to fit the foot (below the ankle) with a flexible upper of leather or plastic and a sole and heel of heavier material

“One click” link to Google Images (alternative for China)



“One-click” links to translations (4 different sites)

“One click” links to videos from PlayPhrase, YouGlish, and Yarn



English ↔ German

[platform](#)

Dictionary English-German

[platform](#) *noun*

[Plattform](#) *f*

Social networks can serve as a platform for business.

The platform allows the sharing of files.

Bahnsteig *m*

My train will be departing from the platform soon.

Mind the gap between the platform and the train.

Podium *nt*

The speaker stood on a platform.

Gleis *nt*

The train leaves from platform seven.

less common:

Podest *nt* · Bühne *f* · Rampe *f* · Standfläche *f* · Pritsche *f*

Soziale Netzwerke können als Plattform für Geschäfte dienen.

Die Plattform ermöglicht das Teilen von Dateien.

Mein Zug wird in Kürze vom Bahnsteig abfahren.

Beachten Sie die Lücke zwischen dem Bahnsteig und dem Zug.


Der Redner stand auf einem Podium.

Der Zug fährt von Gleis sieben ab.

Shows words that tend to co-occur in the 22 million webpages in the corpus. In many cases these provide better insight into meaning and usage than collocates (the standard tool for finding textually related words), and yet we’re not aware of any other corpus that has these. Users can click on any of these words to follow a “chain” of related words.

TOPICS	HORMONE	NOUN		TOPICS	COSMOLOGICAL	ADJ		TOPICS	VIBRATE	VERB	
	Same text	Word	PoS		Same text	Word	PoS		Same text	Word	PoS
1	10151	symptom	n	1	851	universe	n	1	2373	vibration	n
2	9606	blood	n	2	493	theory	n	2	1390	sound	n
3	8527	diet	n	3	417	cosmology	n	3	1300	frequency	n
4	7866	stress	n	4	414	physics	n	4	853	wave	n
5	7471	gland	n	5	373	existence	n	5	640	energy	n
6	7304	muscle	n	6	367	galaxy	n	6	629	noise	n
7	6974	fat	n	7	363	bang	n	7	620	string	n
8	6824	testosterone	n	8	360	cosmic	j	8	528	tone	n
9	6801	body	n	9	303	gravity	n	9	525	instrument	n
10	6783	estrogen	n	10	296	quantum	n	10	476	phone	n
11	6582	pregnancy	n	11	282	particle	n	11	464	particle	n
12	6442	protein	n	12	278	gravitational	j	12	453	resonance	n
13	6426	cell	n	13	277	relativity	n	13	447	ear	n
14	6414	supplement	n	14	268	observation	n	14	420	battery	n
15	6399	disease	n	15	266	infinite	j	15	413	loud	j
16	6290	tissue	n	16	265	argument	n	16	407	motion	n
17	6014	treatment	n	17	265	physicist	n	17	400	pitch	n
18	5983	doctor	n	18	262	expansion	n	18	394	molecule	n
19	5973	vitamin	n	19	256	philosopher	n	19	390	signal	n
20	5840	acid	n	20	250	philosophical	i	20	367	air	n

Words that occur nearby. Click on any word to see the collocates for the new word, or click on the “text” icon to see the word in context. The gray boxes show the most normal placement of the node word (e.g. *bread and butter*, but *slice of bread*). You can also sort by Mutual Information score and set frequency thresholds (Advanced Options).

COLLOCATES			BREAD	NOUN	See also as: VERB	Advanced options				Collocates	Clusters	Topics	Dictionary	Websites	KWIC				
+ NOUN		NEW WORD		?	+ ADJ		NEW WORD		?	+ VERB		NEW WORD		?	+ ADV		NEW WORD		?
18101	6.82	butter			13416	4.01	white			14936	4.18	eat			2509	6.35	freshly		
16525	9.53	loaf			9486	4.44	fresh			14031	6.63	bake			740	3.62	lightly		
14090	7.59	slice			8116	3.29	whole			3300	7.79	toast			437	3.25	evenly		
11801	7.37	banana			5963	11.23	unleavened			2308	3.10	spread			137	4.00	thinly		
10912	4.62	recipe			5906	7.42	baked			1854	4.90	dip			101	3.85	deliciously		
10595	9.33	crumb			5848	6.39	homemade			1846	5.35	slice			74	4.87	thickly		
10267	5.75	cheese			4879	8.38	sliced			1803	2.65	cook			68	4.38	lengthwise		
8447	7.02	wheat			4751	4.28	french			1604	3.65	taste			42	4.01	ever-more		
8324	3.15	piece			4750	9.79	crusty			1399	4.19	soak			37	2.60	diagonally		
8310	6.47	flour			3899	3.19	daily			1390	3.55	top			37	2.65	liberally		
7944	4.54	wine			3793	4.36	delicious			1245	7.23	knead			30	3.04	nutritionally		
7421	6.95	pasta			3690	3.59	sweet			1028	10.26	sourdough			28	3.13	therewith		
6171	5.39	grain			3401	4.40	brown			944	4.01	stuff			18	4.13	industrially		
5834	4.88	cake			3243	8.41	toasted			905	5.18	fry			16	3.23	wherewith		
5695	6.15	garlic			3147	2.59	quick			888	9.01	butter			16	4.32	make-ahead		
5650	6.15	dough			2676	3.72	flat			839	4.20	sprinkle			15	5.05	crackly		
5450	7.89	pudding			2523	2.90	warm			625	3.08	toss			13	2.74	coarsely		
5440	6.21	sandwich			2313	7.58	stale			558	2.64	freeze			9	3.41	best-ever		
5297	5.31	rice			2296	2.73	soft			527	6.23	yeast			9	4.02	mystically		
4554	4.31	meat			2129	7.01	gluten-free			514	3.43	coat			8	2.86	blooming		

CLUSTERS page

The most common 2, 3, and 4 word strings. You can decide how “wide” or “tight” you want the clusters to be (e.g. whether they include highly frequent words like articles and prepositions).

CLUSTERS **BREAD** **NOUN** See also: VERB LIMIT: Loose Medium **Tight** N+N  Collocates **Clusters** Topics Dictionary Websites KWIC  

8798	bread crumbs	9582	white bread	11973	bread and butter	7443	loaf of bread	407	bread on the table	2078	thing since sliced bread
4410	bread recipe	9565	banana bread	5076	bread and wine	3849	piece of bread	406	bread and butter pudding	1565	feast of unleavened bread
4279	bread pudding	5546	unleavened bread	3213	bread of life	3578	slice of bread	301	bread and butter pickles	477	days of unleavened bread
3434	bread flour	4508	wheat bread	1330	bread and water	2903	slices of bread	285	bread in the oven	473	day our daily bread
2998	bread dough	4083	sliced bread	1224	bread and cheese	2725	loaves of bread	284	bread and the wine	397	breaking of the bread
2850	bread machine	4065	garlic bread	764	bread and pasta	2510	whole wheat bread	274	bread on the side	367	day of unleavened bread
1749	bread recipes	3627	fresh bread	746	bread and milk	1539	whole grain bread	254	bread in five minutes	306	top of the bread
1744	bread knife	3620	french bread	675	bread and other	1370	freshly baked bread	172	bread to the hungry	262	substance of the bread
1589	bread made	3425	daily bread	580	breads and cereals	1320	pieces of bread	169	bread for the world	250	foods such as bread
1561	bread making	3347	crusty bread	575	bread from heaven	891	gluten free bread	166	bread to soak up	239	how to make bread
1480	bread baking	3102	rye bread	558	bread is made	843	irish soda bread	160	bread with olive oil	221	loaf of french bread
1374	bread slices	2922	sourdough bread	530	bread and circuses	741	type of bread	157	bread of the presence	217	slices of white bread
1343	bread maker	2918	pita bread	521	breads and pastries	672	whole grain breads	143	bread which came down	216	used to make bread
1125	bread rolls	2881	homemade bread	440	bread to make	660	kind of bread	134	bread that came down	212	side of the bread
1087	bread basket	2867	baked bread	396	bread and meat	619	fresh baked bread	124	bread and olive oil	202	serve with crusty bread
1031	bread cubes	2581	make bread	389	bread and pastries	605	breaking of bread	124	bread in the morning	194	different types of bread
939	bread loaf	2377	panera bread	379	bread at home	594	eat the bread	124	bread is golden brown	193	sides of the bread
931	bread pan	2230	corn bread	343	bread to eat	580	types of bread	123	bread from the oven	189	slice of white bread
910	bread crumb	2139	soda bread	332	bread and then	564	make the bread	120	breads and baked goods	162	chocolate chip banana bread
864	bread alone	2029	baking bread	329	bread is so	557	cheese and bread	120	bread and the cup	154	make your own bread
811	bread bowl	2000	grain bread	321	bread and drink	511	milk and bread	115	bread to mop up	152	make my own bread

KWIC / concordance page

See 100 – 1000 random lines, with the surrounding words highlighted for part of speech. You can re-sort the words by words to the left or right, all with the goal of seeing the patterns in which the word occurs.

KWIC **CRACK** **VERB** See also as: NOUN ADJ # lines: 100 200 500 1000  Collocates Clusters Topics Dictionary Websites **KWIC**  

WEBSITE		SORT	SORT	SORT
151	imsdb.com	the hell it 's made of , but I ca n't	crack	it saw I DONALD Good . Clear it off the table so
152	bleubirdblog.com	and cheese . Pick out your favorite bottle of wine and	crack	it saw I Light a nice- candle- to set the mood .
153	iwantoneofthose.com	they deserve every last drop if they do finally manage to	crack	it I if they do n't , aside from you crowing ,
154	hpc-uk.org	was said and got frustrated and kicked the car window ,	cracking	it I So again Mother A felt the situation was made worse
155	eoshd.com	do ! With all of Adobes resources you'd think they'd have	cracked	it I I # Apples decision to- start afresh with FCPX on the
156	www.afl.com.au	people in that position probably would have had the right to	crack	it I I # " I thought , you know what ? There
157	futureofpersonalhealth.com	for good must be our next accomplishment . If we can	crack	its molecular code , we will have extraordinarily scalable
158	sciencemuseum.org.uk	of hydrogen and oxygen , marking the first step towards	cracking	its molecular structure . Each molecule consists of one oxygen
159	venturebeat.com	to be somewhat recognizably human . Characters can swear and	crack	jokes and possess active sex drives . It 's not like ,
160	techcrunch.com	around the world . Have them play DJ while you can	crack	jokes between songs . # 's most popular station Radio Mozart
161	oscarfishlover.com	like he is protecting something other than territory . I always	crack	jokes that he is " playing with himself " on the rocks
162	miaminewtimes.com	seamlessly to working as a writer . He hustled . He	cracked	jokes I He loved to talk . And talk . And talk
163	pcgscoinfoacts.com	# Die State III : As above , but with die	crack	just beginning at star 12 ; hairline crack extends to border ;
164	rouxbe.com	made these a few times now and the tops are always	cracked	kind of look like an oatmeal raisin cookie . What should I
165	nathanbransford.com	up . The Coldwater River . Confirming her fears , ice	cracked	loud as a pistol shot . Carter undid his seatbelt . Tori
166	grc.com	on this exponential rise , 2021 was where the cost to	crack	made it begin to be feasible for nongovernmental agencies .
167	puzzlepirates.com	is fully intact (those who lobby for overhauled graphics also	crack	me up btw , as they completely miss the point that the
168	mixtapedtorrent.com	neva trapped a day in his life !! I Yaw	crack	me up cause yaw 2 Lame !! Blow pops . / .
169	jenwoodhouse.com	is so gorgeous , Jen ! (And your commentary was	cracking	me up I I love that youve built and created so
170	jensfavoritecookies.com	thoughts on " Orange Funfetti Cupcakes " # God , you	crack	me up I Jen . I have NO doubt that you could
171	theuglyvolvo.com	and held and still just fussed and fussed . # You	crack	me up I Ugly Volvo !! too , once had a
172	stage32.com	" Fargo " and " No Country for Bad Men "	crack	me up I I usually fast forward through the violence in the

Websites page

Create Virtual Corpus

Freq	Per 1000	% Pages	Website (click for topics)	Words
469	36.6	89%	compartes.com	truffle, chocolate, gourmet, luxury, signature, delicious, gift, dark, flavor, box, fresh, best, most, new, take,
932	33.1	97%	amanochocolate.com	cocoa, chocolate, bean, butter, sugar, flavor, temperature, cause, grow, eat, fine, amount, too, food, quality, process, often, small, large, question,
1372	30.8	96%	chocoley.com	dipping, dipping, chocolate, truffle, melted, melt, dip, harden, candy, gourmet, glossy, refrigerator, coating, fridge, excess, milk, cream, bowl, sheet, tap,
395	28.1	74%	fanniemay.com	confection, chocolate, delicious, gift, rich, best, day, may, more,
302	26.7	95%	mrchocolate.com	chocolate, milk, signature, dark, gift, box,
1031	25.6	94%	ecolechocolat.com	chocolatier, ganache, chocolate, cocoa, saucepan, welcome, stir, melt, mixture, bean, flavor, pan, butter, medium, pour, milk, cream, bowl, sugar, recipe,
604	24.0	85%	ghirardelli.com	bittersweet, cocoa, chocolate, melted, baking, preheat, stir, vanilla, prepared, melt, mixer, butter, refrigerator, mixture, bake, spoon, pan, bowl, medium, flour,

Move between page showing the websites in which the word is a keyword (see above), and the page showing all keywords from a specific website (right).

In other large web-based corpora, the web pages are just the result of “blindly” following links. In iWeb, we know a lot about each of the 95,000 websites, which makes possible a wide range of searches that cannot be done anywhere else.

WORDS FROM: [chocoley.com](#)

	Frequency	% texts		Word	PoS
1	115	0.49		dipping	n
2	41	0.26		dipping	j
3	1372	0.96		chocolate	n
4	48	0.25		truffle	n
5	51	0.38		melted	j
6	214	0.68		melt	v
7	118	0.50		dip	v
8	45	0.32		harden	v
9	148	0.38		candy	n
10	41	0.27		gourmet	n
11	33	0.29		glossy	j

Summary

The iWeb corpus has a number of features that sets it apart from any other corpus. These include its **size** (14 billion words, 22 million web pages, 95,000 websites), its **searches** (range of query types, and the ease and speed of its searches), and the ease and speed with which you can create “**Virtual Corpora**” for any topic of interest.

In addition, it is different from any of the other BYU corpora in the attention that it gives to the **top 60,000 words** in the corpus, and the wide range of information for each word, including frequency information, definitions, synonyms, WordNet entries, related topics, concordances (new display in iWeb), clusters, websites that have the word as a “keyword”, and KWIC/concordance lines.

All of these features make iWeb the ideal corpus for researchers, teachers, and learners.