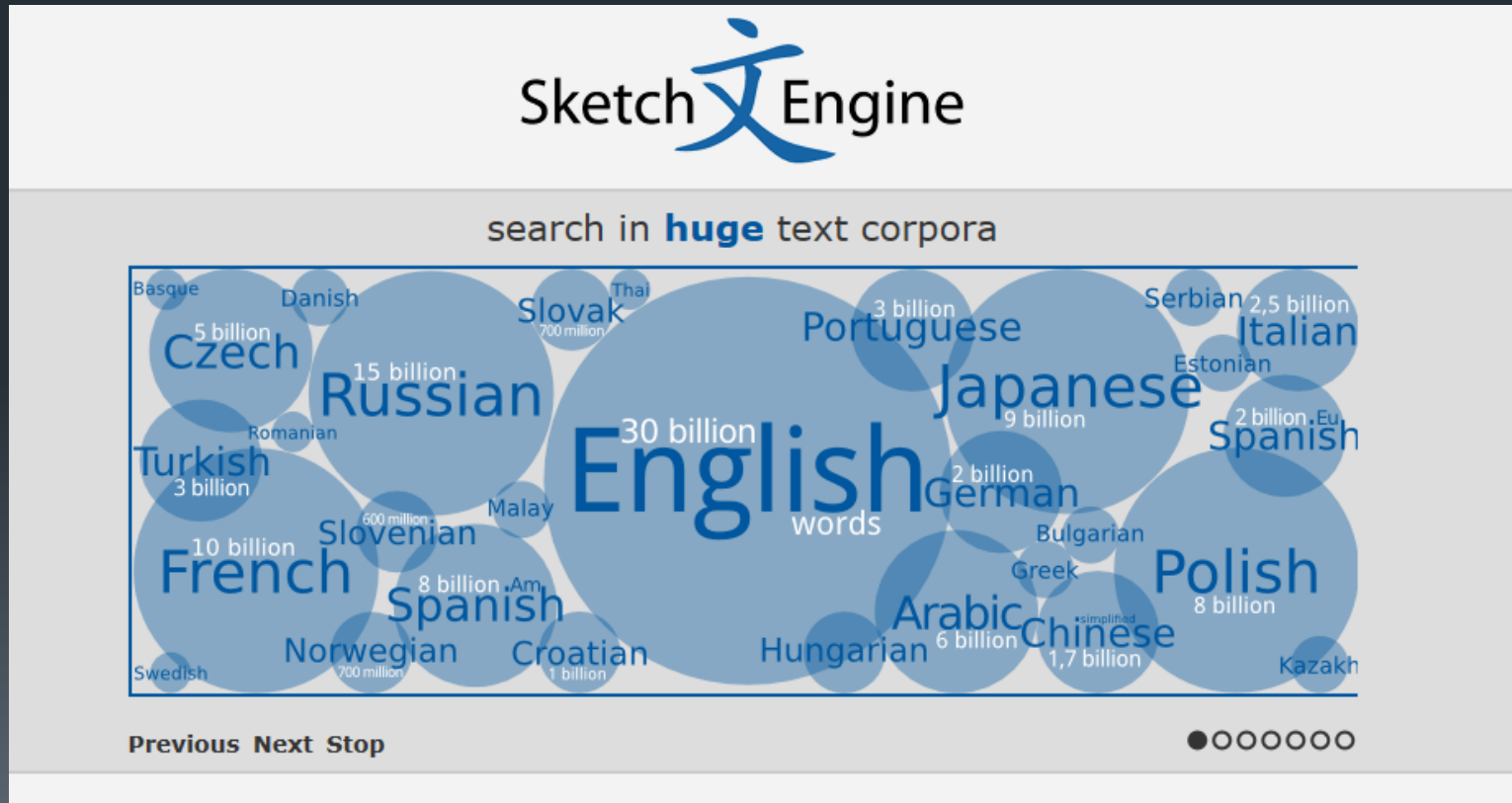# 望月圭子大学院ゼミ発表資料

Sketch Engineの使い方

2014/06/09

川村　駿（M1)

Sketch Engineとは

- 言葉の働きを調査するためのツール
  - Lexical Computing社のAdam Kilgrriff氏によって、2003年に設立
- さまざまなコーパスを一か所に集めた、ハブ的役割を持つサイト

Sketch Engineとは

- 世界中の言語のコーパスを収録している

Afrikaans/ Alabian/ Arabic/ Azerbaijani/ Bengali/ Bosnian/ Bulgarian/ Catalan/ Croatian/ Czech/ Danish/ Dutch/ English/ Estonian/ Filipino/ Finnish/ French/ German/ Greek/ Gujarati/ Hebrew/ Hindi/ Hungarian/ Chinese-simplified/ Chinese-traditional/ Igbo/ Indonesian/ Irish/Italian/ Japanese/ Kazakh/ Korean/ Kyrgyz/ Latin/ Latvian/ Lithuanain/ Macedonian/ Malay/ Malayalam/ Maltense/ Maori/ Nerwegian/ Persian/ Polish/ Portuguese/ Romanian/ Russian/ Serbian/ Setswana/ Slovak/ Slovenian/ Spanish/ Swahili/ Swedish/ Tamil/ Telugu/ Thai/ Turkish/ Turkmen/ Ukrainian/ Uzbek/ Vietnamese/ Welsh

- 全63言語

http://www.sketchengine.co.uk/

Sketch Engineでできること

- Concordance
  - 指定した語を中心に、どのような文脈で使用されているのかを調べる

Sketch Engineでできること

- Concordance
  - 「knowledge」で検索した場合

Sketch Engineでできること

- 共起される語や品詞の指定
- テキストの種類の指定

Sketch Engineでできること

- Word List
  - コーパス内に使用されている単語のリストを作ることができる

Sketch Engineでできること

- Word List
  - British National Corpus (BNC)に多く使われている単語

Corpus: British National Corpus			

Page [1] [Go] Next >

word	Freq
the	5,415,473
of	3,027,293
to	2,566,874
and	2,510,049
a	2,040,346
in	1,788,884
that	1,062,370
is	972,923
was	877,918
I	862,214
for	831,741
it	821,805
on	694,931
be	648,529
with	640,381
The	619,060
as	603,494
you	574,714
at	487,203
by	486,695
he	462,137
are	455,051
have	453,293
not	432,451
had	418,575
from	410,171
his	381,674
which	361,184
or	358,453
this	343,534

they	336,525
but	321,463
an	319,130
n't	316,344
were	312,027
her	288,929
we	265,155
one	261,418
been	259,938
has	254,355
she	251,063
all	249,118
their	245,769
will	245,301
do	242,909
would	240,940
there	234,784
It	232,004
can	224,462
up	204,844
more	201,697
said	194,767
out	194,625
who	190,875
so	189,339
about	188,454
if	181,213
He	178,015
what	174,851
when	170,095
them	167,096
no	159,245
into	157,157

could	156,542
its	153,970
In	153,620
him	153,168
time	149,602
some	147,341
other	146,413
than	144,563
like	140,611
two	140,178
only	139,953
then	131,682
me	129,348
A	128,904
my	127,621
over	126,478
did	126,320
your	125,156
But	124,887
also	118,800
know	118,237
any	116,158
very	115,142
just	114,599
people	113,684
now	113,342
may	111,429
first	110,884
This	110,039

Sketch Engineでできること

- Word List
  - リスト化する単語の種類を選ぶこともできる
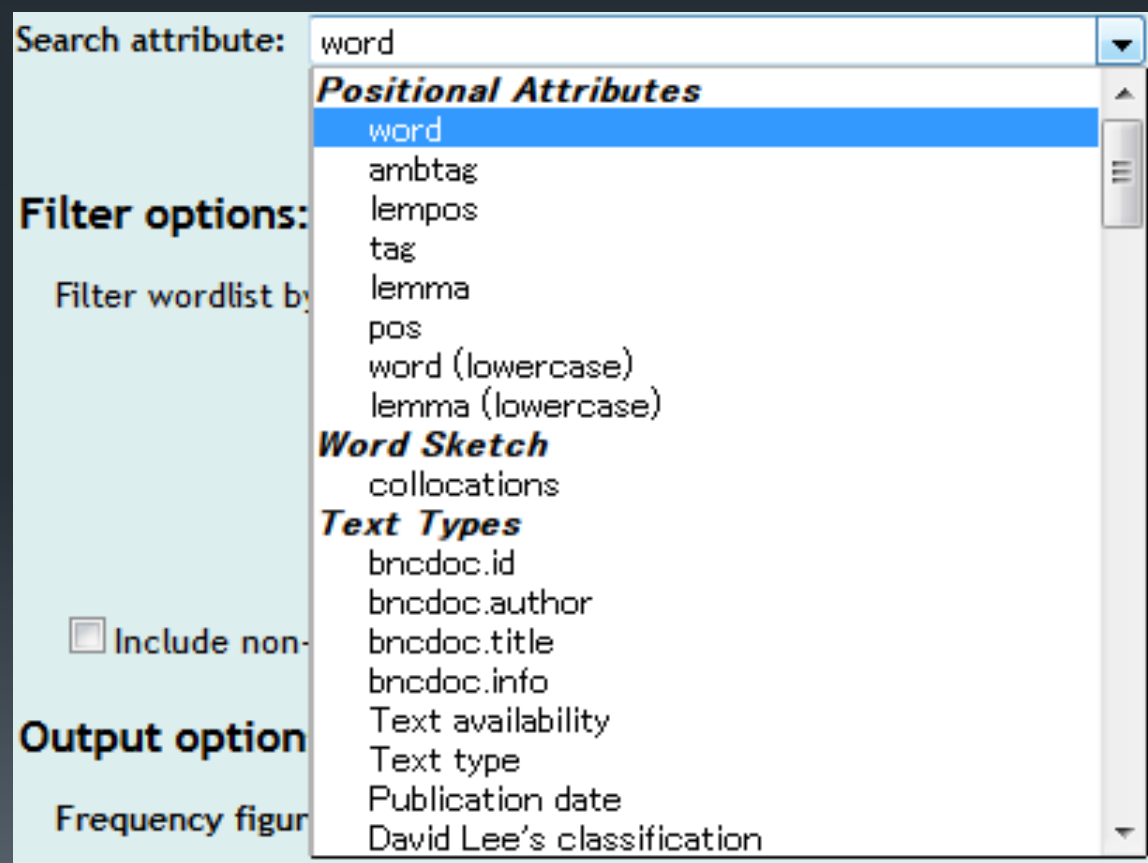
Sketch Engineでできること

- Word Sketch
  - ある語がコーパス内で、どのような語と共起しているかを検索できる

基本的に上部のみで検索可

さらに詳細な設定も可能

Sketch Engineでできること
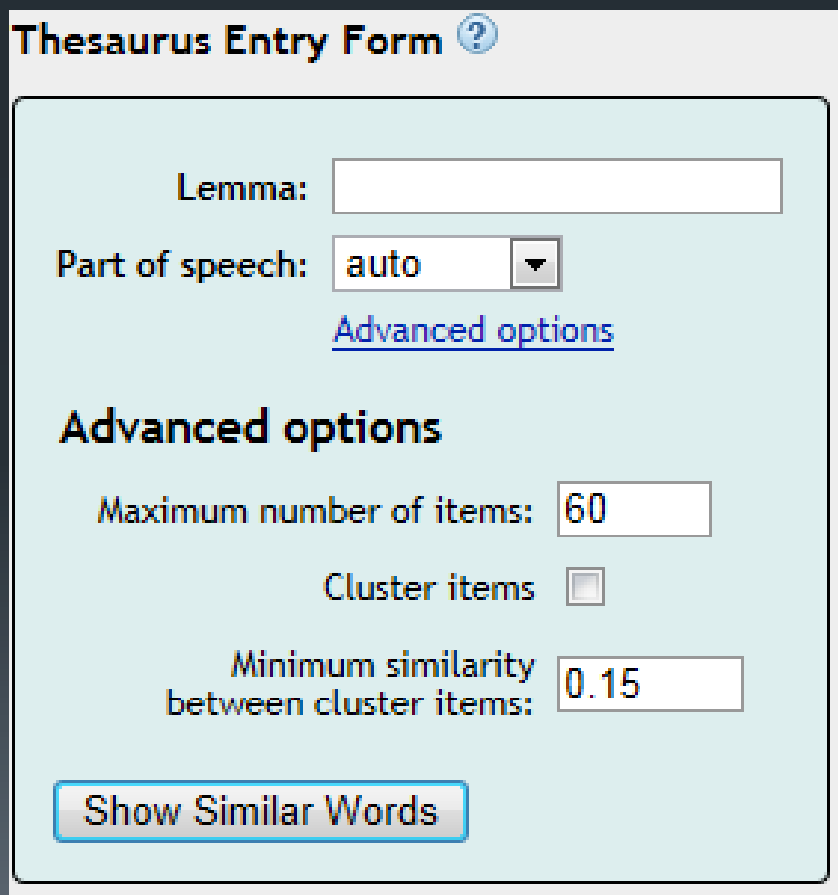
- Word Sketch
  - 「knowledge」と共起される語を調べる

knowledge *(noun)*
British National Corpus freq = 14381 (128.2 per million)

usually in singular (99.8 %, percentile 17.6)

Bilingual word sketch available for: French (French web corpus).

modifier	5,129	1.1	pp_of-p	3,376	3.9	and/or	2,981	1.4	object_of	2,829	1.7	pp_obj_of-p	2,594	3.0
scientific	161	8.78	language	97	6.38	skill	475	9.12	acquire	161	8.55	sociology	115	9.62
prior	76	8.6	greek	11	6.08	expertise	70	8.16	gain	105	7.61	lack	137	7.95
intimate	62	8.33	workings	9	5.93	understanding	131	7.96	possess	45	7.41	advancement	17	7.3
detailed	99	8.11	God	43	5.92	experience	242	7.53	impart	15	7.19	corpus	15	6.9
technical	84	7.74	grammar	13	5.87	wisdom	18	6.72	deny	62	7.11	acquisition	29	6.83
thorough	39	7.67	english	25	5.81	competence	15	6.35	lack	32	6.86	pursuit	19	6.81
personal	146	7.65	anatomy	7	5.76	belief	40	6.28	glean	11	6.78	breadth	12	6.76
linguistic	42	7.42	bsl	6	5.73	consent	21	6.27	accumulate	13	6.56	depth	21	6.57
specialist	70	7.41	subject	61	5.71	appreciation	11	6.05	pool	9	6.5	best	59	6.27
schematic	27	7.37	chemistry	10	5.61	recklessness	6	5.94	update	14	6.48	fount	6	6.19
shared	33	7.33	topic	14	5.47	enthusiasm	13	5.69	disseminate	9	6.46	diffusion	7	6.18
background	74	7.24	nature	42	5.35	faith	16	5.51	test	32	6.4	frontier	10	6.06
sound	30	7.23	existence	17	5.29	ability	33	5.48	share	49	6.4	basis	53	5.84
common	103	7.06	biology	6	5.27	enjoyment	6	5.37	detail	11	6.38	body	105	5.79
previous	68	7.04	past	15	5.19	attitude	28	5.29	require	145	6.37	archaeology	6	5.74
inside	23	7.03	literature	13	5.11	intelligence	9	5.1	embody	12	6.28	kind	66	5.61
specialised	24	7.01	circumstance	20	5.1	suspicion	6	4.93	transmit	12	6.22	transmission	8	5.6
working	46	6.97	history	38	5.04	technique	22	4.93	advance	16	6.18	source	43	5.45
first-hand	20	6.94	probability	6	5.03	awareness	9	4.88	extend	40	6.11	conception	9	5.37
superior	27	6.94	cause	18	4.89	love	19	4.83	broaden	8	6.03	theory	41	5.36
practical	50	6.92	world	75	4.85	information	54	4.73	presuppose	7	6.01	validity	6	5.34
human	84	6.78	truth	11	4.83	qualification	7	4.6	utilise	8	5.97	mode	12	5.28
specialist	19	6.75	german	7	4.75	perspective	7	4.59	demonstrate	23	5.79	absence	14	5.25
in-depth	18	6.74	Christ	7	4.7	concept	13	4.54	attain	8	5.77	application	36	5.2
existing	51	6.74	technique	19	4.69	confidence	10	4.52	derive	21	5.76	store	14	5.11

Sketch Engineでできること

- Thesaurus
  - コーパス内でシソーラス（類語辞典）を作成することができる

Sketch Engineでできること

- Thesaurus
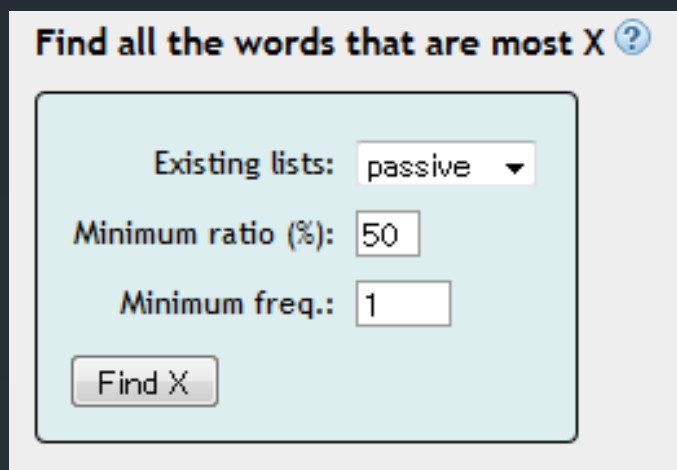 - 検索結果がリストだけでなく、視覚的にもわかりやすく表示される

Sketch Engineでできること

- Thesaurus
  - 類義語をクリックするとSketch-Diff（後述）の画面にリンクする

knowledge/understanding British National Corpus freqs = 14381 | 7615

| knowledge | 6.0 | 4.0 | 2.0 | 0 | -2.0 | -4.0 | -6.0 | understanding |

and/or	2,981	1,332	1.4	1.1
sympathy	0	33	--	7.7
patience	0	16	--	7.4
empathy	0	7	--	7.0
tolerance	0	10	--	6.9
compassion	0	7	--	6.7
acceptance	0	17	--	6.5
insight	0	14	--	6.4
kindness	0	6	--	6.3
respect	0	18	--	6.3
sensitivity	0	9	--	6.0
knowledge	29	131	5.0	7.3
awareness	9	29	4.9	6.8
enjoyment	6	8	5.4	6.4
intelligence	9	13	5.1	5.9
appreciation	11	13	6.0	6.8
belief	40	9	6.3	4.3
experience	242	30	7.5	4.6
skill	475	36	9.1	5.5
enthusiasm	13	0	5.7	--
recklessness	6	0	5.9	--
consent	21	0	6.3	--
competence	15	0	6.4	--
wisdom	18	0	6.7	--
understanding	131	0	8.0	--
expertise	70	0	8.2	--

subject_of	686	238	0.8	0.5
dawn	0	9	--	8.9
reach	0	6	--	2.6
go	9	0	0.6	--
make	18	0	0.7	--
help	13	7	3.1	2.3
give	13	0	1.0	--
tell	6	0	1.0	--
lead	15	6	3.2	1.9
enable	17	6	4.8	3.3
come	24	7	2.5	0.8
require	6	0	1.8	--
bring	8	0	1.9	--
play	6	0	1.9	--
seem	10	0	2.2	--
develop	6	0	2.4	--
become	15	0	2.4	--
involve	8	0	2.7	--
increase	9	0	3.2	--
affect	10	0	3.7	--
grow	10	0	3.9	--
gain	8	0	4.1	--
influence	8	0	4.5	--
inform	6	0	4.5	--

adj_subject_of	363	62	1.9	0.6
important	9	0	2.9	--
possible	11	0	3.4	--
concerned	7	0	3.7	--
useful	6	0	4.2	--
necessary	16	0	4.6	--
essential	8	0	4.7	--
impossible	9	0	5.0	--
available	35	0	5.0	--
relevant	10	0	5.1	--

Sketch Engineでできること

- Find X
  - コーパス内に特定の語形がどれほど使われているかを検索する

- 検索できる語形
  - Passive
  - Singular
  - -ing
  - spoken

Sketch Engineでできること

- Find X
  - 「-ing」で検索
  - 「Let」を表示

Sketch Engineでできること

- Sketch-Diff(erence)
  - コーパス内の２語、あるいは２つのサブコーパスにおける
    ある語の使用方法の違い（主にコロケーションの違い）を検索できる

Word Sketch Differences Entry Form ②

Lemma:	knowledge
Part of speech:	noun ▼
Sketch diff by:	⦿ lemma
	Second lemma: understanding
	○ subcorpus
	First subcorpus: None (whole corpus) ▼ *info*
	Second subcorpus: None (whole corpus) ▼ *info*
	○ word form
	First word form:
	Second word form:
	Advanced options
Show Diff	

Sketch Engineでできること

- Sketch-Diff
  - 「intelligent」と「clever」のコロケーションの違いを検索

赤い表示
＝「clever」とより共起される

緑の表示
＝「intelligent」と
　　　より共起される

個人使用時の料金

2014 prices

Individual academic licence via PayPal (inclusive of VAT):

monthly	GBP 5.20
3-monthly	GBP 14.60
yearly	GBP 52.00

All other prices in GBP, per year, exclusive of UK VAT.

1GBP ≒172JPY
（2014/06/08現在）

VAT（付加価値税）
＝20%

Academic and commercial licences

	Single user	5 users	10 users	20 users	50 users	100 users	200 users
Academic	52	208	364	624	1,040	1,404	1,664
Freelance translator	78						
Commercial – low intensity	130	520	910	1,560	2,600	3,562	4,160
Commercial – mid intensity	312	1,248	2,184	3,744	6,240	8,736	9,984
Commercial – high intensity/lexicography	832	3,328	5,824	9,984	16,640	22,464	26,624

high-intencity：年100回以上の使用　　mid-intensity：年30回以上の使用
Academicは辞書編纂目的以外であれば、教育や調査のために使ってもよいというアカウント