
Comparison of Sketch Engine and Colterm for Scientific Texts

Conference on Language Technologies &
Digital Humanities Ljubljana, 2018

Klara Eva Kukoičič

—



**Monolingual
terminology base**



30%

**Automatic
extraction tools**



~~Terms written down manually~~

—

PURPOSE

- The KAS Corpus of Academic Slovene
 - Terminology and automatic term extraction
 - Comparison of Sketch Engine and CollTerm
-

CORPUS OF ACADEMIC SLOVENE

- Action Plan for Language Infrastructure & Action Plan for Language Education
 - 50.000 scientific texts
 - 1 billion words
-

ANALYSED CANDIDATES

- Multi-word extraction
- Bi-, tri-, four-grams
- 100 candidates from each list
- 300 candidates - CollTerm
- 300 candidates - Sketch Engine

600 candidates

CATEGORIES

1. Term
 2. Term from other field
 3. Scientific writing
 4. Irrelevant
-

ANALYSIS



CANDIDATES
SORTED INTO
CATEGORIES

RANKING OF
THE SAME
TERMS

PERCENTAGE
OF POSITIVE
CANDIDATES
AMONG FIRST
N-CANDIDATES

CANDIDATES IN CATEGORIES

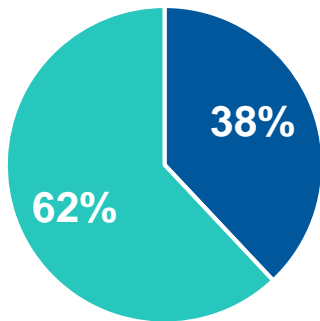
	Category	CollTerm	Sketch Engine
1	Term	68	54
2	Other field	13	3
3	Scientific writing	106	130
4	Irrelevant	113	113

PERCENTAGE OF POSITIVE CANDIDATES

		CollTerm	Sketch Engine
1	Term	81 (27 %)	57 (19 %)
2	Not term	219 (73 %)	243 (81 %)

FIRST 20 CANDIDATES

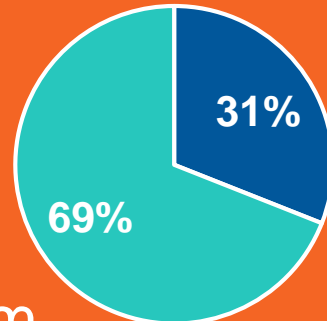
■ TERM ■ NOT TERM



CollTerm

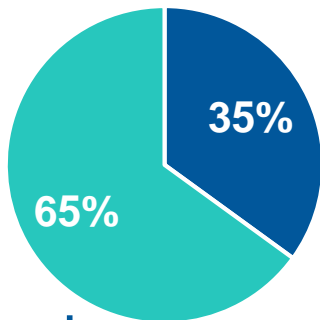
FIRST 50 CANDIDATES

■ TERM ■ NOT TERM



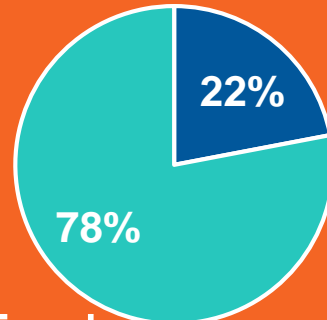
CollTerm

■ TERM ■ NOT TERM



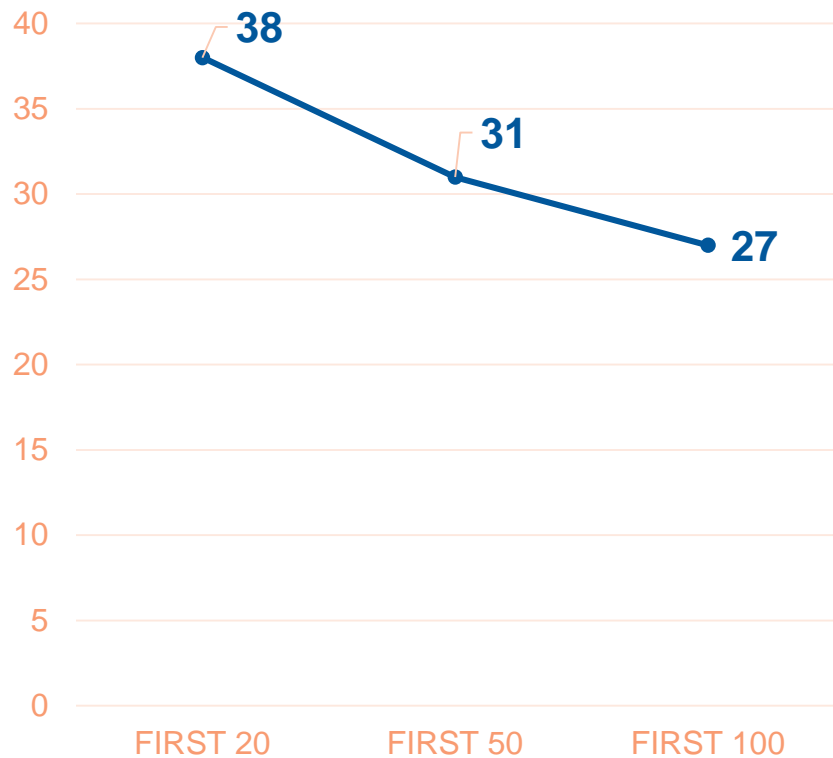
Sketch Engine

■ TERM ■ NOT TERM

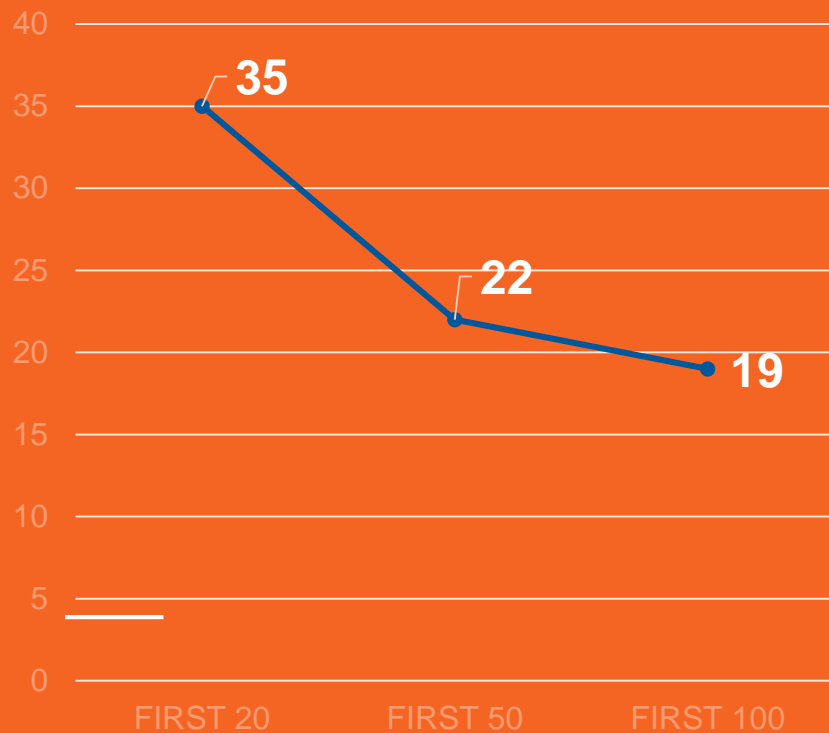


Sketch Engine

CollTerm



Sketch Engine



Ranking no. in ColIT	Ranking no. inv Sk E
2	30
5	21
7	9
13	205
16	67
17	0
18	135
19	10
21	4541
22	0
24	54
88	0

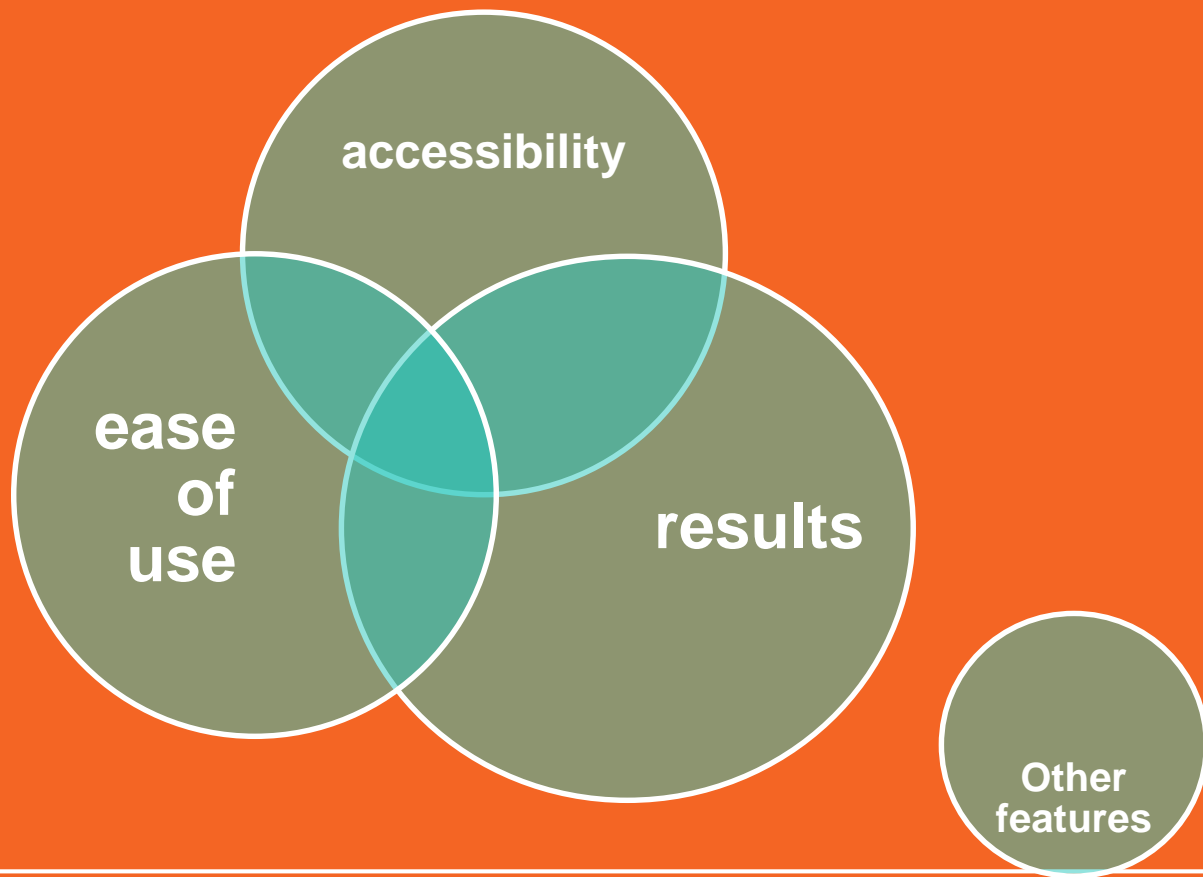
Ranking no. in ColIT	Ranking no. in Sk E
32	546
36	4727
39	40
40	4440
44	4422
50	4681
59	7
63	201
70	4669
71	454
72	4106
83	4594

Ranking no. in ColIT	Ranking no. in Sk E
2	30
5	21
7	9
13	205
16	67
17	0
18	135
19	10
21	4541
22	0
24	54
88	0

Ranking no. in ColIT	Ranking no. in Sk E
32	546
36	4727
39	40
40	4440
44	4422
50	4681
59	7
63	201
70	4669
71	454
72	4106
83	4594

Ranking no. in Sk E	Ranking no. ColIT
2	185
7	59
9	7
10	19
11	192
20	195
21	5
29	194
33	2
34	0
40	39
54	24

Ranking no. In Sk E	Ranking no. in ColIT
66	0
67	16
70	219
96	215





Assessing the recall

Comparison of Sketch Engine and Collterm for Scientific Texts

THANK YOU FOR
YOUR ATTENTION
