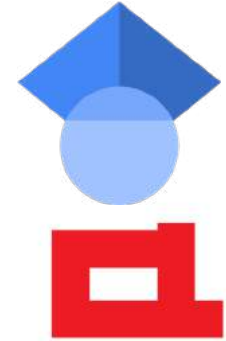


NLP Scholar: An Interactive Visual Explorer for Natural Language Processing Literature



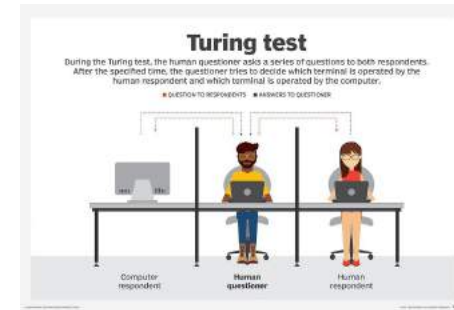
Saif M. Mohammad

Senior Research Scientist, National Research Council Canada

✉ Saif.Mohammad@nrc-cnrc.gc.ca [@SaifMMohammad](https://twitter.com/SaifMMohammad)

Ten-Second History of NLP

- Turing Test proposed
- Formation of the Association for Computational Linguistics (ACL)
 - originally named the Association for Machine Translation and Computational Linguistics (AMTCL)
- First ACL conference
- ACL, ELRA, and AFNLP publish peer-reviewed papers
 - journal articles and conference proceedings (long, short, demo, etc.)
- Broad interdisciplinary field
 - researchers from Computer Science, Linguistics, Information Science, Psychology, Social Sciences, Humanities, and more
 - thousands of papers published every year



1950

1962

1965



Today



Need for a Search Engine Dedicated to NLP Literature

Desirable features:

- easy to use
- visual
- interactive
- captures impact of papers

Metrics of Research Impact (on subsequent scholarly work)



- Often derived from citations
 - number of citations, average citations, h-index, relative citation ratio, and impact factor (Bornmann and Daniel, 2009)
- However, citations do not always reflect quality or importance
Impacted by:
 - systematic biases
 - atypical contributions
 - popularity of area
 - unethical practices (e.g. egregious self citations)

Nonetheless, given the lack of other easily applicable and effective metrics, citation metrics used as an imperfect but useful window into research impact

- often a factor in funding research and hiring scientists

The NLP Scholar Project

Data: extracted and aligned information from

- the ACL Anthology (AA) 
- Google Scholar (GS) 

to create a dataset of tens of thousands of NLP papers and their citations

[NLP Scholar: A Dataset for Examining the State of NLP Research.](#) LREC 2020.

Analyzing NLP Literature:

[Examining Citations of Natural Language Processing Literature.](#) ACL 2020.

[Gender Gap in Natural Language Processing Research: Disparities in Authorship and Citations.](#) ACL 2020.

Practical Tool:

NLP Scholar: An interactive visual explorer for this unified AA-GS dataset

- to find related work

Uses of NLP Scholar

- search for related work in various areas within NLP
- identify the highly cited articles on an interactive timeline
- identify past papers published in a venue of interest (such as ACL or COLING)
 - E.g., papers from ten years back published at ACL that have substantial citations
- examine changes in number of articles and number of citations in an area over time
- identify citation impact of different types of papers
 - e.g., short papers, shared task papers, demo papers, etc.

Other practical applications of the NLP Scholar dataset

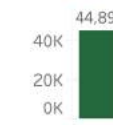
Web browser extensions that allow:

- users of GS to look up the aligned AA information
 - the full ACL BibTeX, poster, slides, access to proceedings from the same venue, etc.
- access from AA to the GS information of the aligned paper
 - number of citations, lists of papers citing the paper, etc.

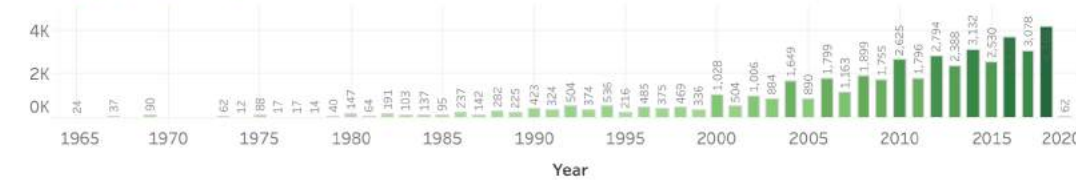
NLP Scholar Dashboard

- A1: #papers
 - 44,895
- A2: #papers by year
 - substantial numbers since the 2000s
 - higher in alternate years (biennial conferences such as LREC, COLING)
- B1: # citations
 - ~1.2 million citations (as of June 2019)
- B2: # citations by year
 - 2000s are the most notable

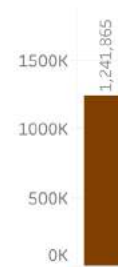
A1. #papers



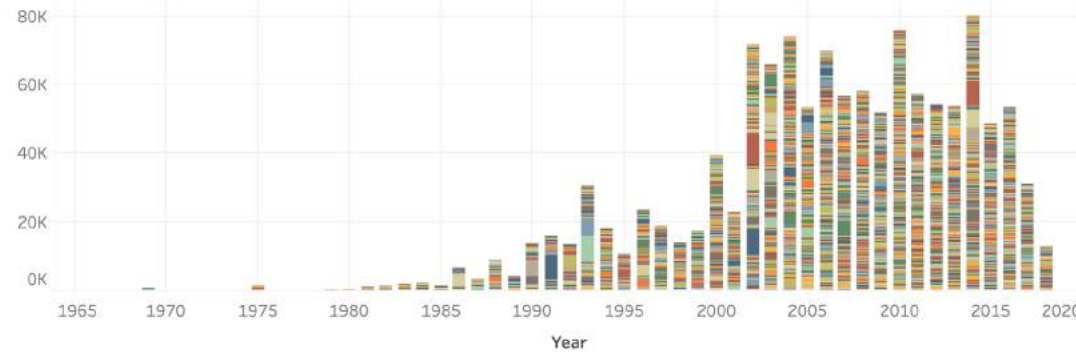
A2. #papers, by year of publication



B1. #citations



B2. #citations, by year of publication



C. Papers

Row	Paper-Id	Paper-Title	Author(s)	Year	Url
1	P02-1040	Bleu: a Method for Automatic Evaluation of Machine Translation	Papineni, Kishore and Roukos, S..	2002	htt.. 9,098
2	W02-1011	Thumbs up? Sentiment Classification using Machine Learning Techniques	Pang, Bo and Lee, Lillian and Val..	2002	htt.. 8,187
3	D14-1162	Glove: Global Vectors for Word Representation	Pennington, Jeffrey and Socher..	2014	htt.. 7,965
4	J93-2004	Building a Large Annotated Corpus of English: The Penn Treebank	Marcus, Mitch and Santorini, Be..	1993	htt.. 7,527
5	J91-4003	The Generative Lexicon	Pustejovsky, James	1991	htt.. 6,593
6	P02-1053	Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsuper..	Turney, Peter	2002	htt.. 5,642
7	D14-1179	Learning Phrase Representations using RNN Encoder-Decoder for Stati..	Cho, Kyunghyun and van Merrie..	2014	htt.. 5,344
8	J93-2003	The Mathematics of Statistical Machine Translation: Parameter Estima..	Brown, Peter F. and Della Pietra..	1993	htt.. 5,047
9	J90-1003	Word Association Norms, Mutual Information, and Lexicography	Church, Kenneth and Hanks, Pat..	1990	htt.. 4,845
10	P07-2045	Moses: Open Source Toolkit for Statistical Machine Translation	Koehn, Philipp and Hoang, Hieu ..	2007	htt.. 4,581

OK SK 10K
#citations

D. Authors

Row	Author-name	#citations
1	Manning, Christoph..	54,587
2	Koehn, Philipp	19,412
3	Och, Franz Josef	18,620
4	Socher, Richard	17,506
5	Lee, Lillian	17,458
6	Jurafsky, Dan	16,405
7	Hovy, Eduard	16,292
8	Klein, Dan	15,881
9	Ney, Hermann	15,097
10	Dyer, Chris	14,745

OK 20K 40K 60K 80K
#citations

E. Search by year of publication, title term (unigram, bigram), or author name

Year of publication

1965 2019

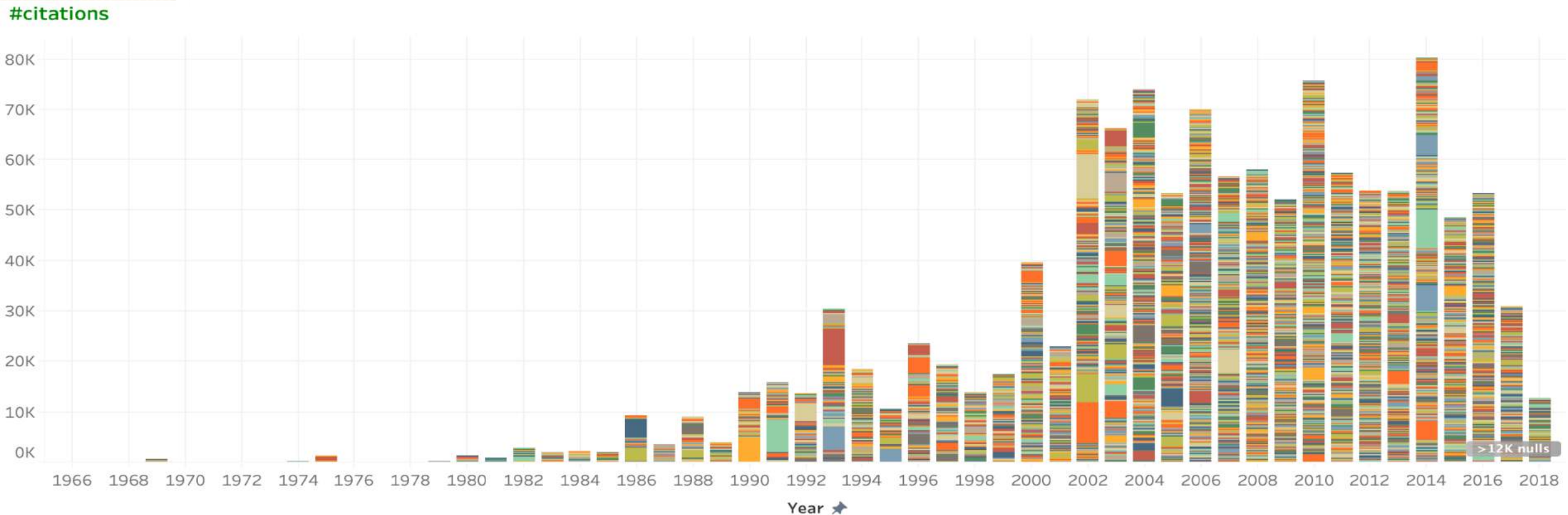
Unigram

Bigram

Author Name

B2: # citations by year

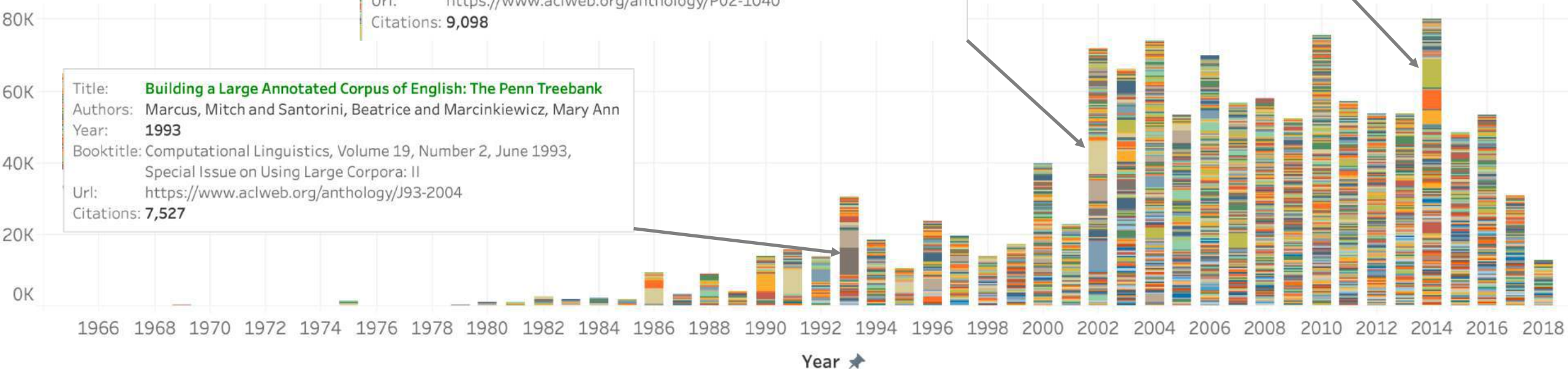
- ~1.2 million citations (as of June 2019)
- colored segments correspond to each of the papers
- the height of a segment is proportional to the number of citations the paper has received



B2: # citations by year

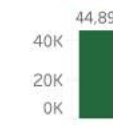
- ~1.2 million citations (as of June 2019)
- colored segments correspond to each of the papers
- the height of a segment is proportional to the number of citations the paper has received

#citations, by year of publication

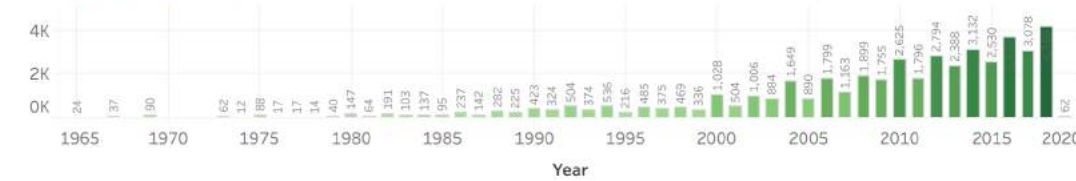


- A1: #papers
 - 44,895
- A2: #papers by year
 - substantial numbers since the 2000s
 - higher in alternate years (biennial conferences such as LREC, COLING)
- B1: # citations
 - ~1.2 million citations (as of June 2019)
- B2: # citations by year
 - 2000s are the most notable
- C: most cited papers
 - *machine translation, sentiment analysis, word embeddings, syntax, semantics*
- D: most cited authors
- E: search options

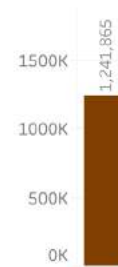
A1. #papers



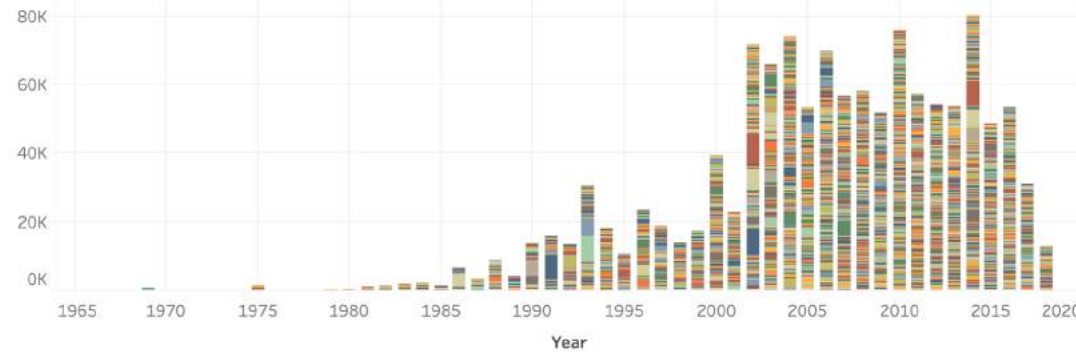
A2. #papers, by year of publication



B1. #citations



B2. #citations, by year of publication



C. Papers

Row	Paper-Id	Paper-Title	Author(s)	Year	Url	#citations
1	P02-1040	Bleu: a Method for Automatic Evaluation of Machine Translation	Papineni, Kishore and Roukos, S..	2002	htt..	9,098
2	W02-1011	Thumbs up? Sentiment Classification using Machine Learning Techniques	Pang, Bo and Lee, Lillian and Val..	2002	htt..	8,187
3	D14-1162	Glove: Global Vectors for Word Representation	Pennington, Jeffrey and Socher..	2014	htt..	7,965
4	J93-2004	Building a Large Annotated Corpus of English: The Penn Treebank	Marcus, Mitch and Santorini, Be..	1993	htt..	7,527
5	J91-4003	The Generative Lexicon	Pustejovsky, James	1991	htt..	6,593
6	P02-1053	Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsuper..	Turney, Peter	2002	htt..	5,642
7	D14-1179	Learning Phrase Representations using RNN Encoder-Decoder for Stati..	Cho, Kyunghyun and van Merrie..	2014	htt..	5,344
8	J93-2003	The Mathematics of Statistical Machine Translation: Parameter Estima..	Brown, Peter F. and Della Pietra..	1993	htt..	5,047
9	J90-1003	Word Association Norms, Mutual Information, and Lexicography	Church, Kenneth and Hanks, Pat..	1990	htt..	4,845
10	P07-2045	Moses: Open Source Toolkit for Statistical Machine Translation	Koehn, Philipp and Hoang, Hieu ..	2007	htt..	4,581

OK SK 10K
#citations

D. Authors

Row	Author-name	#citations
1	Manning, Christoph..	54,587
2	Koehn, Philipp	19,412
3	Och, Franz Josef	18,620
4	Socher, Richard	17,506
5	Lee, Lillian	17,458
6	Jurafsky, Dan	16,405
7	Hovy, Eduard	16,292
8	Klein, Dan	15,881
9	Ney, Hermann	15,097
10	Dyer, Chris	14,745

E. Search by year of publication, title term (unigram, bigram), or author name

Year of publication

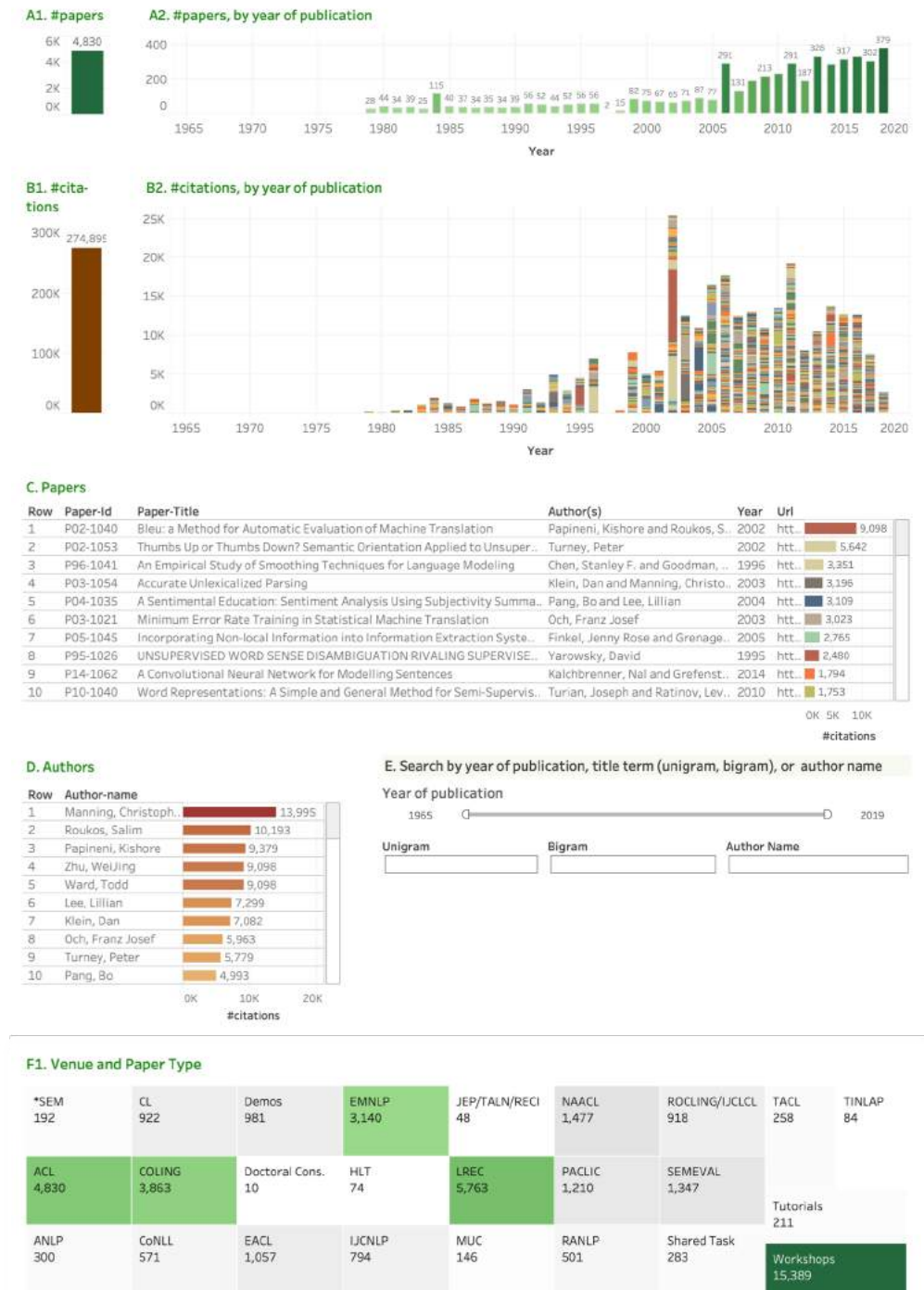
1965 2019

Unigram Bigram Author Name

Four Other Dashboards

Main dashboard (A to E) + sixth element (F) for focused search

- F1: venues and paper types
- F2: title unigrams
- F3: title bigrams
- F4: language mentions in the title



Treemap of Venues and Paper Types

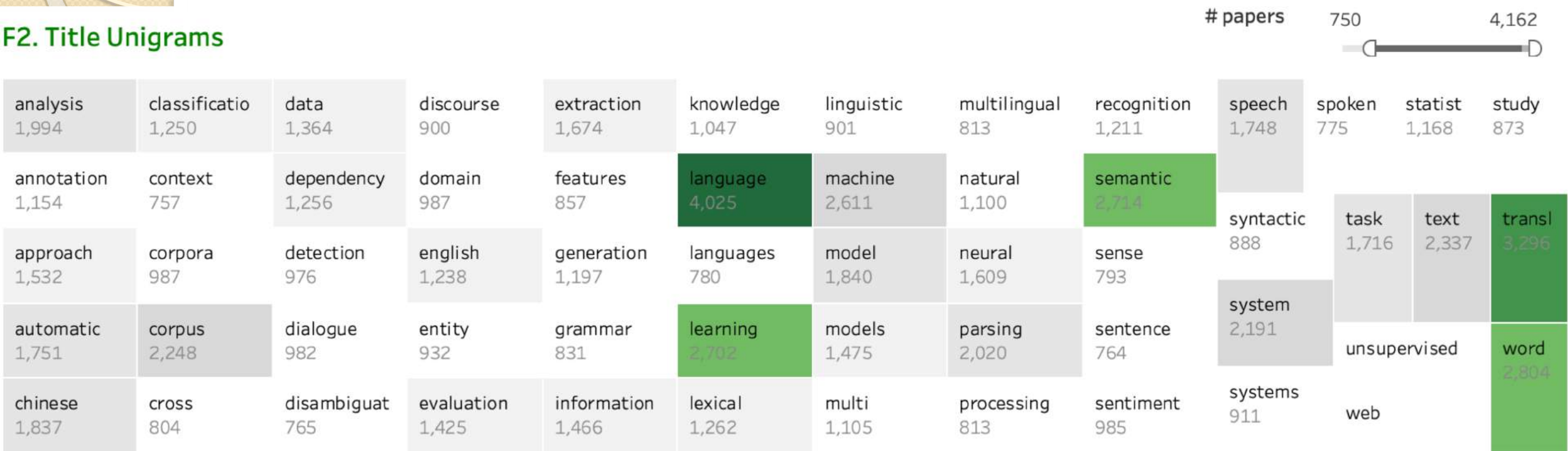
F1. Venue and Paper Type

*SEM 192	CL 922	Demos 981	EMNLP 3,140	JEP/TALN/RECI 48	NAACL 1,477	ROCLING/IJCLCL 918	TACL 258	TINLAP 84
ACL 4,830	COLING 3,863	Doctoral Cons. 10	HLT 74	LREC 5,763	PACLIC 1,210	SEMEVAL 1,347	Tutorials 211	
ANLP 300	CoNLL 571	EACL 1,057	IJCNLP 794	MUC 146	RANLP 501	Shared Task 283	Workshops 15,389	

- Items are in alphabetic order
- Below each item is the number of papers
- Darker shades of green indicate higher numbers

Treemap of Common Title Unigrams

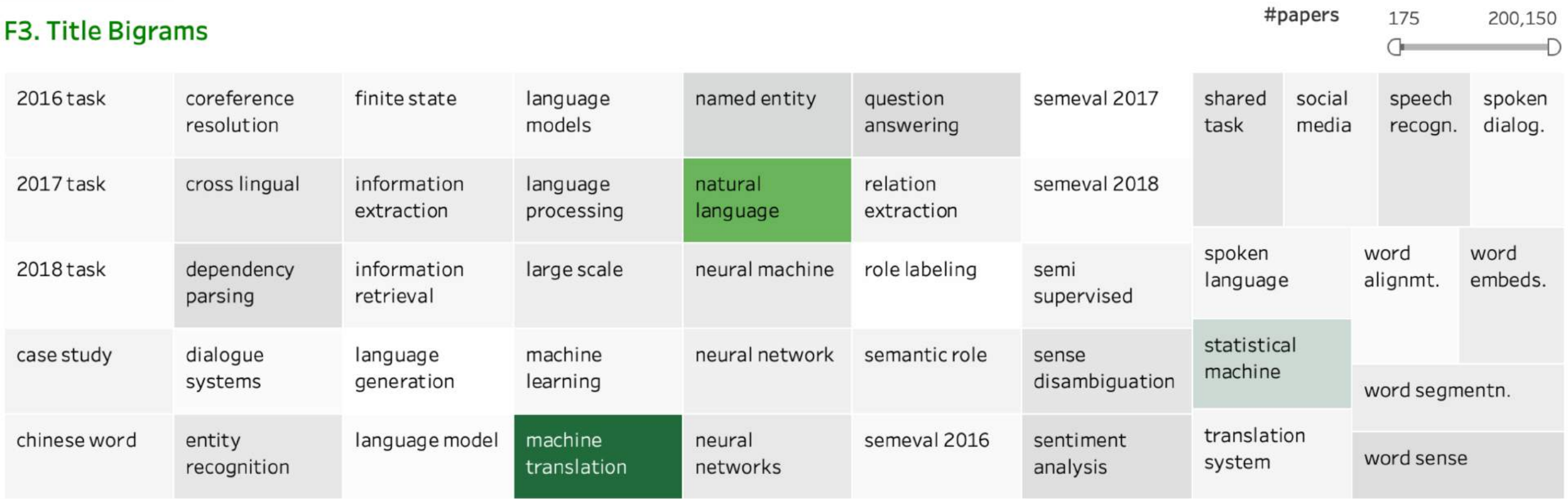
F2. Title Unigrams



- Items are in alphabetic order
- Below each item is the number of papers
- Darker shades of green indicate higher numbers

Treemap of Common Title Bigrams

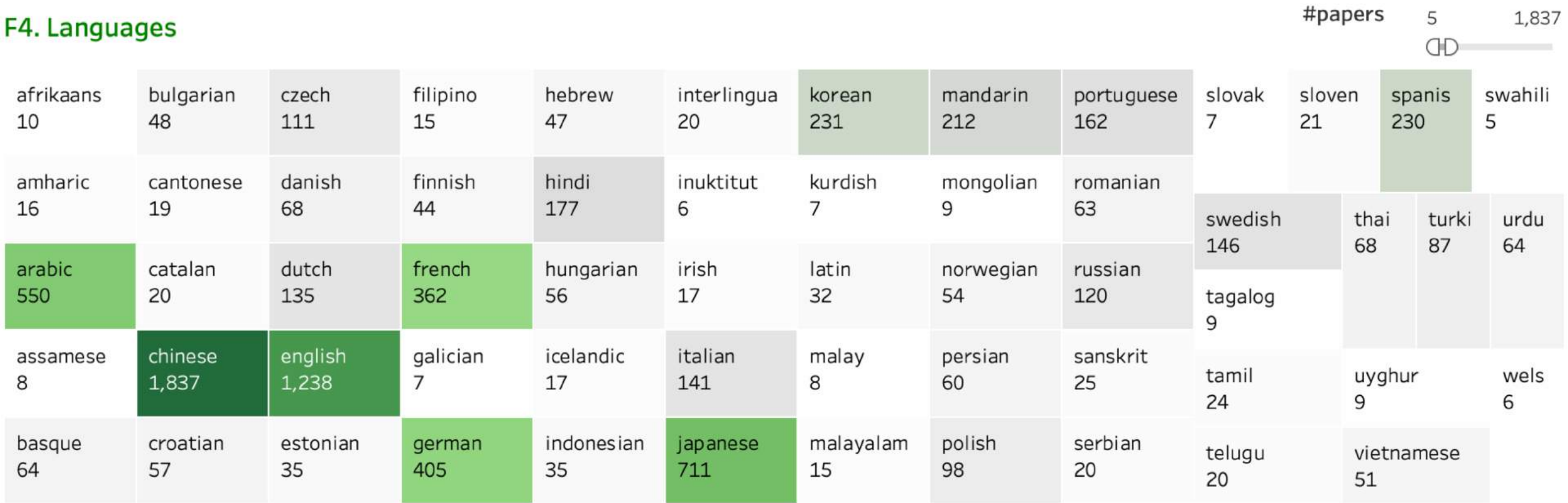
F3. Title Bigrams



- Items are in alphabetic order
- Below each item is the number of papers
- Darker shades of green indicate higher numbers

Treemap of Common Languages

F4. Languages

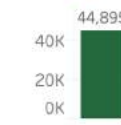


- Items are in alphabetic order
- Below each item is the number of papers
- Darker shades of green indicate higher numbers

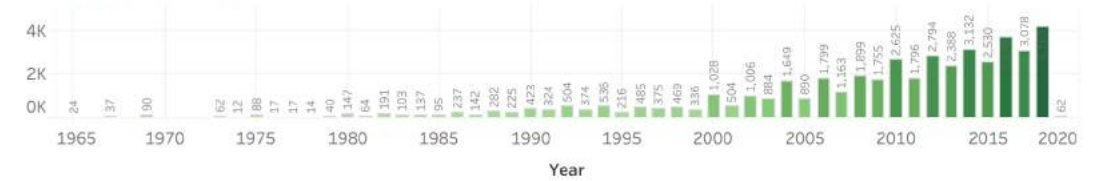
NLP Scholar Dashboard

Main View

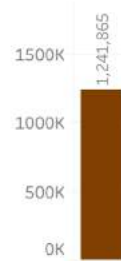
A1. #papers



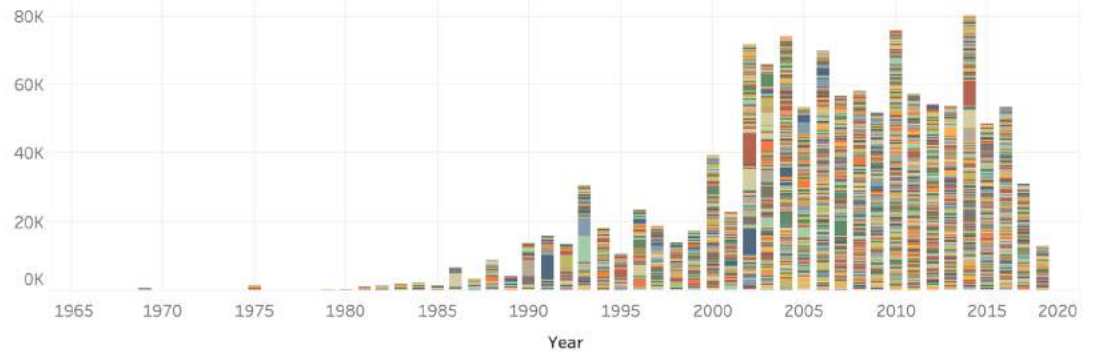
A2. #papers, by year of publication



B1. #citations



B2. #citations, by year of publication



C. Papers

Row	Paper-Id	Paper-Title	Author(s)	Year	Url
1	P02-1040	Bleu: a Method for Automatic Evaluation of Machine Translation	Papineni, Kishore and Roukos, S..	2002	htt.. 9,098
2	W02-1011	Thumbs up? Sentiment Classification using Machine Learning Techniques	Pang, Bo and Lee, Lillian and Vai..	2002	htt.. 8,187
3	D14-1162	Glove: Global Vectors for Word Representation	Pennington, Jeffrey and Socher..	2014	htt.. 7,965
4	J93-2004	Building a Large Annotated Corpus of English: The Penn Treebank	Marcus, Mitch and Santorini, Be..	1993	htt.. 7,527
5	J91-4003	The Generative Lexicon	Pustejovsky, James	1991	htt.. 6,593
6	P02-1053	Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsuper..	Turney, Peter	2002	htt.. 5,642
7	D14-1179	Learning Phrase Representations using RNN Encoder-Decoder for Stati..	Cho, Kyunghyun and van Merrie..	2014	htt.. 5,344
8	J93-2003	The Mathematics of Statistical Machine Translation: Parameter Estima..	Brown, Peter F. and Della Pietra..	1993	htt.. 5,047
9	J90-1003	Word Association Norms, Mutual Information, and Lexicography	Church, Kenneth and Hanks, Pat..	1990	htt.. 4,845
10	P07-2045	Moses: Open Source Toolkit for Statistical Machine Translation	Koehn, Philipp and Hoang, Hieu ..	2007	htt.. 4,581

OK 5K 10K
#citations

D. Authors

Row	Author-name	#citations
1	Manning, Christoph..	54,587
2	Koehn, Philipp	19,412
3	Och, Franz Josef	18,620
4	Socher, Richard	17,506
5	Lee, Lillian	17,458
6	Jurafsky, Dan	16,405
7	Hovy, Eduard	16,292
8	Klein, Dan	15,881
9	Ney, Hermann	15,097
10	Dyer, Chris	14,745

OK 20K 40K 60K 80K
#citations

E. Search by year of publication, title term (unigram, bigram), or author name

Year of publication

1965 2019

Unigram

Bigram

Author Name

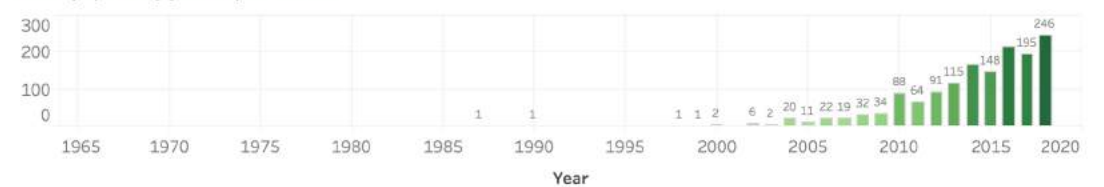
NLP Scholar Dashboard

After entering terms associated with sentiment analysis in the Search Box.

A1. #papers



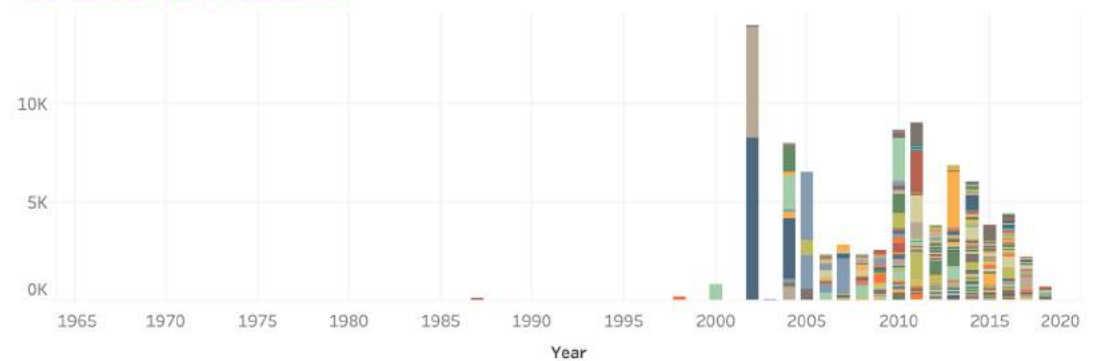
A2. #papers, by year of publication



B1. #citations



B2. #citations, by year of publication



C. Papers

Row	Paper-Id	Paper-Title	Author(s)	Year	Url	#citations
1	W02-1011	Thumbs up? Sentiment Classification using Machine Learning Techniques	Pang, Bo and Lee, Lillian and Vai..	2002	htt..	8,187
2	P02-1053	Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsuper..	Turney, Peter	2002	htt..	5,642
3	H05-1044	Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis	Wilson, Theresa and Wiebe, Jan..	2005	htt..	3,487
4	P04-1035	A Sentimental Education: Sentiment Analysis Using Subjectivity Summa..	Pang, Bo and Lee, Lillian	2004	htt..	3,109
5	D13-1170	Recursive Deep Models for Semantic Compositionality Over a Sentimen..	Socher, Richard and Perelygin, ..	2013	htt..	2,798
6	L10-1-531	SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysi..	Baccianella, Stefano and Esuli, ..	2010	htt..	2,263
7	J11-2001	Lexicon-Based Methods for Sentiment Analysis	Taboada, Maite and Brooke, Jul..	2011	htt..	1,982
8	P05-1015	Seeing Stars: Exploiting Class Relationships for Sentiment Categorizati..	Pang, Bo and Lee, Lillian	2005	htt..	1,743
9	P07-1056	Biographies, Bollywood, Boom-boxes and Blenders: Domain Adaptation ..	Blitzer, John and Dredze, Mark ..	2007	htt..	1,735
10	C04-1200	Determining the Sentiment of Opinions	Kim, SooMin and Hovy, Eduard	2004	htt..	1,723

D. Authors

Row	Author-name	#citations
1	Pang, Bo	13,039
2	Lee, Lillian	13,039
3	Vaithyanathan, Shi..	8,187
4	Turney, Peter	6,148
5	Ng, Andrew Y.	5,158
6	Manning, Christoph..	4,518
7	Wilson, Theresa	4,502
8	Wiebe, Janyce	4,409
9	Potts, Christopher	4,257
10	Socher, Richard	3,823

E. Search by year of publication, title term (unigram, bigram), or author name

Year of publication
1965 to 2019

Unigram
#emotional
affect
emotion
emotional
emotions
orientation
sentiment
stance
valence

Bigram

Author Name

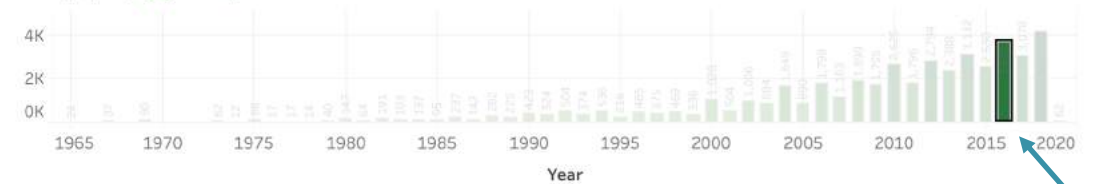
NLP Scholar Dashboard

Clicking on the 2016 bar
in the #papers by year viz (A2)

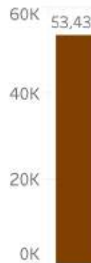
A1. #papers



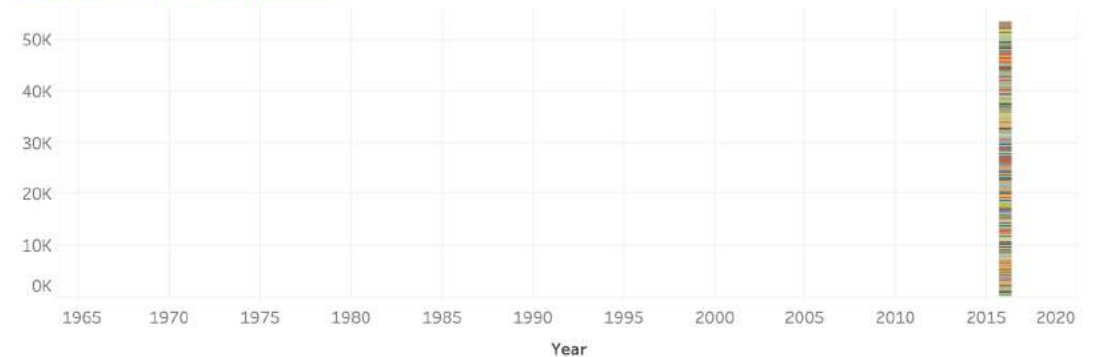
A2. #papers, by year of publication



B1. #citations



B2. #citations, by year of publication



C. Papers

Row	Paper-Id	Paper-Title	Author(s)	Year	Url	#citations
1	N16-3020	"Why Should I Trust You?": Explaining the Predictions of Any Classifier	Ribeiro, Marco Tulio and Singh, ..	2016	htt..	1,387
2	P16-1162	Neural Machine Translation of Rare Words with Subword Units	Sennrich, Rico and Haddow, Bar..	2016	htt..	1,028
3	N16-1030	Neural Architectures for Named Entity Recognition	Lample, Guillaume and Balleste..	2016	htt..	957
4	N16-1174	Hierarchical Attention Networks for Document Classification	Yang, Zichao and Yang, Diyi and..	2016	htt..	952
5	D16-1264	SQuAD: 100,000+ Questions for Machine Comprehension of Text	Rajpurkar, Pranav and Zhang, Ji..	2016	htt..	748
6	P16-1101	End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF	Ma, Xuezhe and Hovy, Eduard	2016	htt..	611
7	S16-1001	SemEval-2016 Task 4: Sentiment Analysis in Twitter	Nakov, Preslav and Ritter, Alan..	2016	htt..	567
8	K16-1002	Generating Sentences from a Continuous Space	Bowman, Samuel and Vilnis, Lu..	2016	htt..	561
9	S16-1002	SemEval-2016 Task 5: Aspect Based Sentiment Analysis	Pontiki, Maria and Galanis, Dim..	2016	htt..	549
10	D16-1044	Multimodal Compact Bilinear Pooling for Visual Question Answering an..	Fukui, Akira and Park, Dong Huk..	2016	htt..	430

D. Authors

Row	Author-name	#citations
1	Dyer, Chris	2,886
2	Hovy, Eduard	2,000
3	Haddow, Barry	1,919
4	Sennrich, Rico	1,730
5	Birch, Alexandra	1,583
6	Manning, Christoph..	1,455
7	Singh, Sameer	1,414
8	Ribeiro, Marco Tulio	1,387
9	Guestrin, Carlos	1,387
10	Goldberg, Yoav	1,326

E. Search by year of publication, title term (unigram, bigram), or author name

Year of publication
1965 2019

Unigram Bigram Author Name

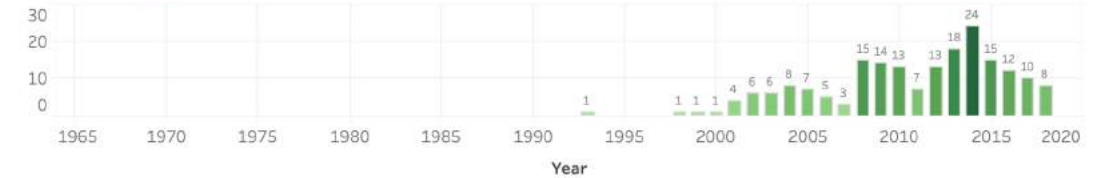
NLP Scholar Dashboard

Clicking on 'Manning, Christopher' in the Authors List (D)

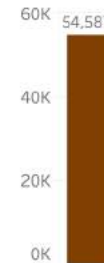
A1. #papers



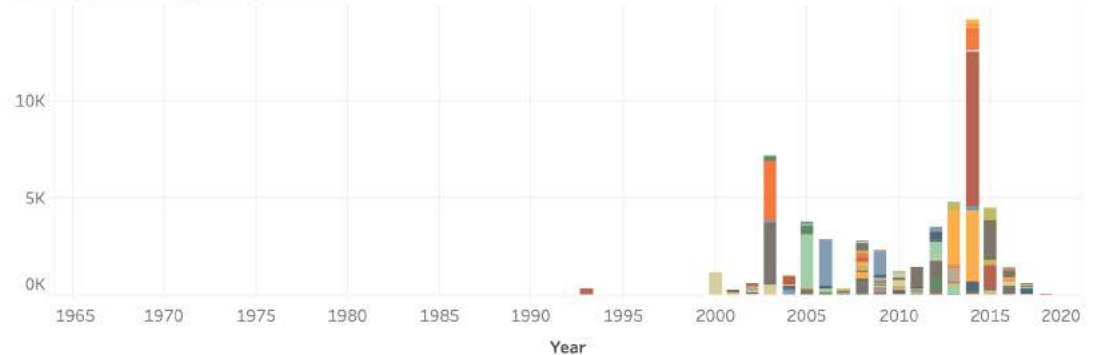
A2. #papers, by year of publication



B1. #citations



B2. #citations, by year of publication



C. Papers

Row	Paper-Id	Paper-Title	Author(s)	Year	Url	#citations
1	D14-1162	Glove: Global Vectors for Word Representation	Pennington, Jeffrey and Socher..	2014	htt..	7,965
2	P14-5010	The Stanford CoreNLP Natural Language Processing Toolkit	Manning, Christopher D. and Su..	2014	htt..	3,543
3	P03-1054	Accurate Unlexicalized Parsing	Klein, Dan and Manning, Christo..	2003	htt..	3,196
4	N03-1033	Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network	Toutanova, Kristina and Klein, D..	2003	htt..	3,083
5	D13-1170	Recursive Deep Models for Semantic Compositionality Over a Sentimen..	Socher, Richard and Perelygin, ..	2013	htt..	2,798
6	P05-1045	Incorporating Non-local Information into Information Extraction Syste..	Finkel, Jenny Rose and Grenage..	2005	htt..	2,765
7	L06-1-260	Generating Typed Dependency Parses from Phrase Structure Parses	de Marneffe, MarieCatherine a..	2006	htt..	2,414
8	D15-1166	Effective Approaches to Attention-based Neural Machine Translation	Luong, Minh-Thang and Pham, ..	2015	htt..	1,961
9	D09-1026	Labeled LDA: A supervised topic model for credit attribution in multi-lab..	Ramage, Daniel and Hall, David ..	2009	htt..	1,168
10	W00-1308	Enriching the Knowledge Sources Used in a Maximum Entropy Part-of-S..	Toutanova, Kristina and Mannin..	2000	htt..	1,164

0K 5K 10K

#citations

D. Authors

Row	Author-name	#citations
1	Manning, Christoph..	54,587
2	Koehn, Philipp	19,412
3	Och, Franz Josef	18,520
4	Socher, Richard	17,505
5	Lee, Lillian	17,458
6	Jurafsky, Dan	16,405
7	Hovy, Eduard	16,292
8	Klein, Dan	15,881
9	Ney, Hermann	15,097
10	Dyer, Chris	14,745

0K 20K 40K 60K 80K

#citations

E. Search by year of publication, title term (unigram, bigram), or author name

Year of publication

1965 2019

Unigram

Bigram

Author Name

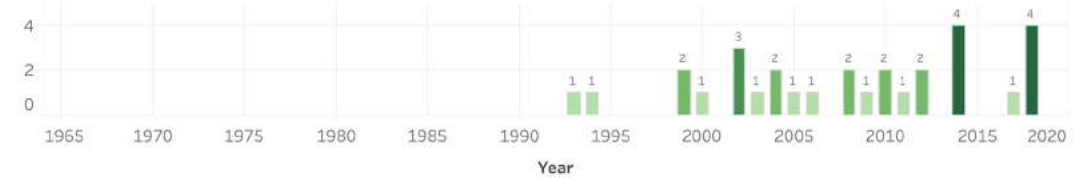
NLP Scholar Dashboard

Clicking on 'Lee, Lillian' in the Authors List (D)

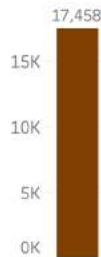
A1. #papers



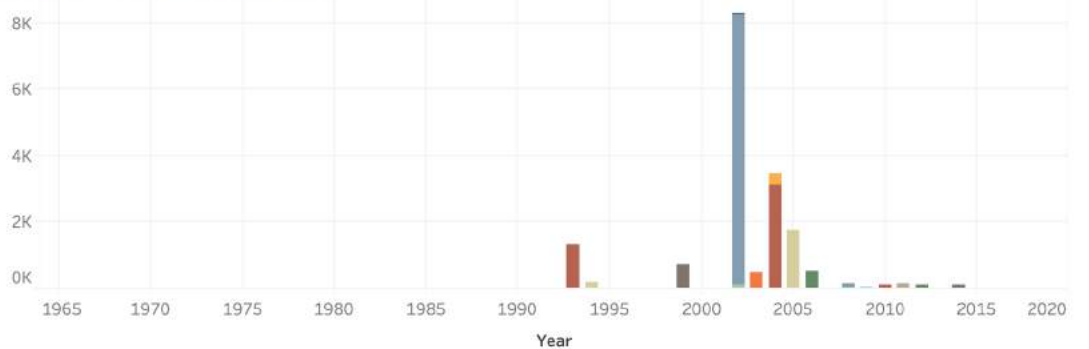
A2. #papers, by year of publication



B1. #citations



B2. #citations, by year of publication



C. Papers

Row	Paper-Id	Paper-Title	Author(s)	Year	Url	#citations
1	W02-1011	Thumbs up? Sentiment Classification using Machine Learning Techniques	Pang, Bo and Lee, Lillian and Vail...	2002	htt..	8,187
2	P04-1035	A Sentimental Education: Sentiment Analysis Using Subjectivity Summa..	Pang, Bo and Lee, Lillian	2004	htt..	3,109
3	P05-1015	Seeing Stars: Exploiting Class Relationships for Sentiment Categorizati..	Pang, Bo and Lee, Lillian	2005	htt..	1,743
4	P93-1024	DISTRIBUTIONAL CLUSTERING OF ENGLISH WORDS	Pereira, Fernando and Tishby, N..	1993	htt..	1,322
5	P99-1004	Measures of Distributional Similarity	Lee, Lillian	1999	htt..	697
6	W06-1639	Get out the vote: Determining support or opposition from Congressiona..	Thomas, Matt and Pang, Bo and..	2006	htt..	536
7	N03-1003	Learning to Paraphrase: An Unsupervised Approach Using Multiple-Seq..	Barzilay, Regina and Lee, Lillian	2003	htt..	504
8	N04-1015	Catching the Drift: Probabilistic Content Models, with Applications to G..	Barzilay, Regina and Lee, Lillian	2004	htt..	341
9	P94-1038	Similarity-Based Estimation of Word Cooccurrence Probabilities	Dagan, Ido and Pereira, Fernan..	1994	htt..	183
10	W11-0609	Chameleons in Imagined Conversations: A New Approach to Understand..	DanescuNiculescuMizil, Cristia..	2011	htt..	147

0K 5K 10K

#citations

D. Authors

Row	Author-name	#citations
1	Manning, Christoph..	54,587
2	Koehn, Philipp	19,412
3	Och, Franz Josef	18,520
4	Socher, Richard	17,506
5	Lee, Lillian	17,458
6	Jurafsky, Dan	16,405
7	Hovy, Eduard	16,292
8	Klein, Dan	15,881
9	Ney, Hermann	15,097
10	Dyer, Chris	14,745

E. Search by year of publication, title term (unigram, bigram), or author name

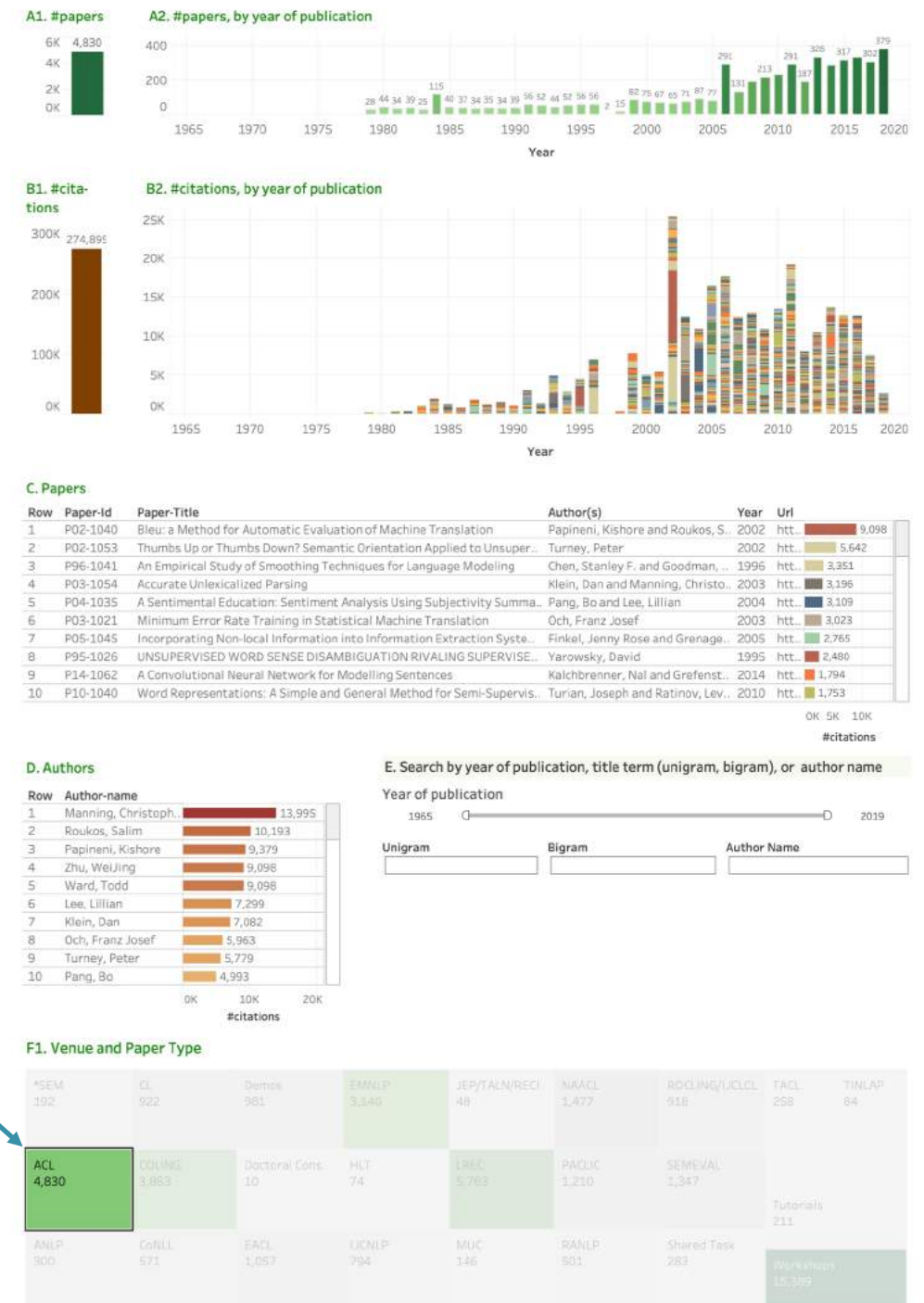
Year of publication

1965 2019

Unigram Bigram Author Name

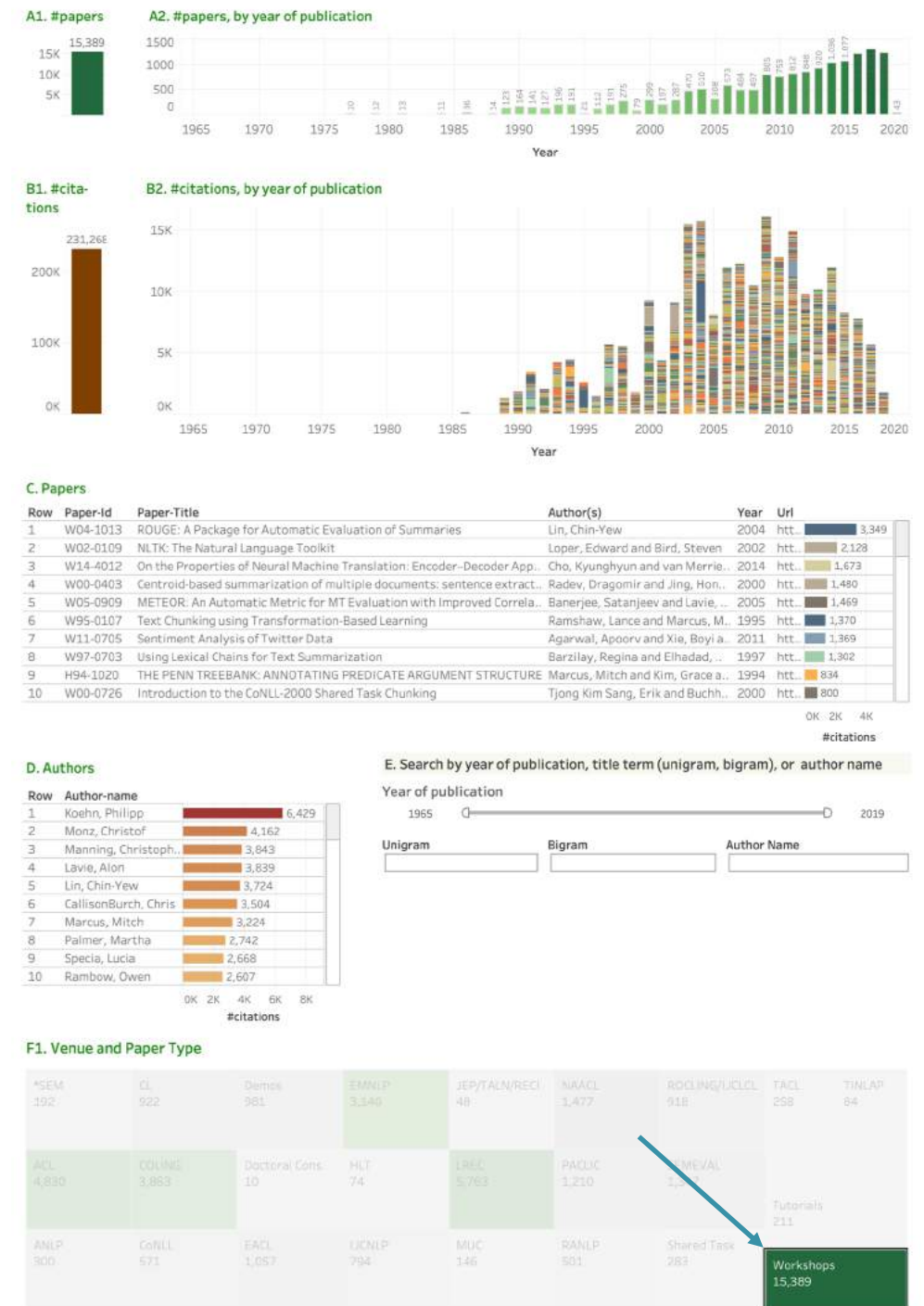
NLP Scholar Dashboard

Clicking on 'ACL' in the
Venue and Paper Type treemap (F1)



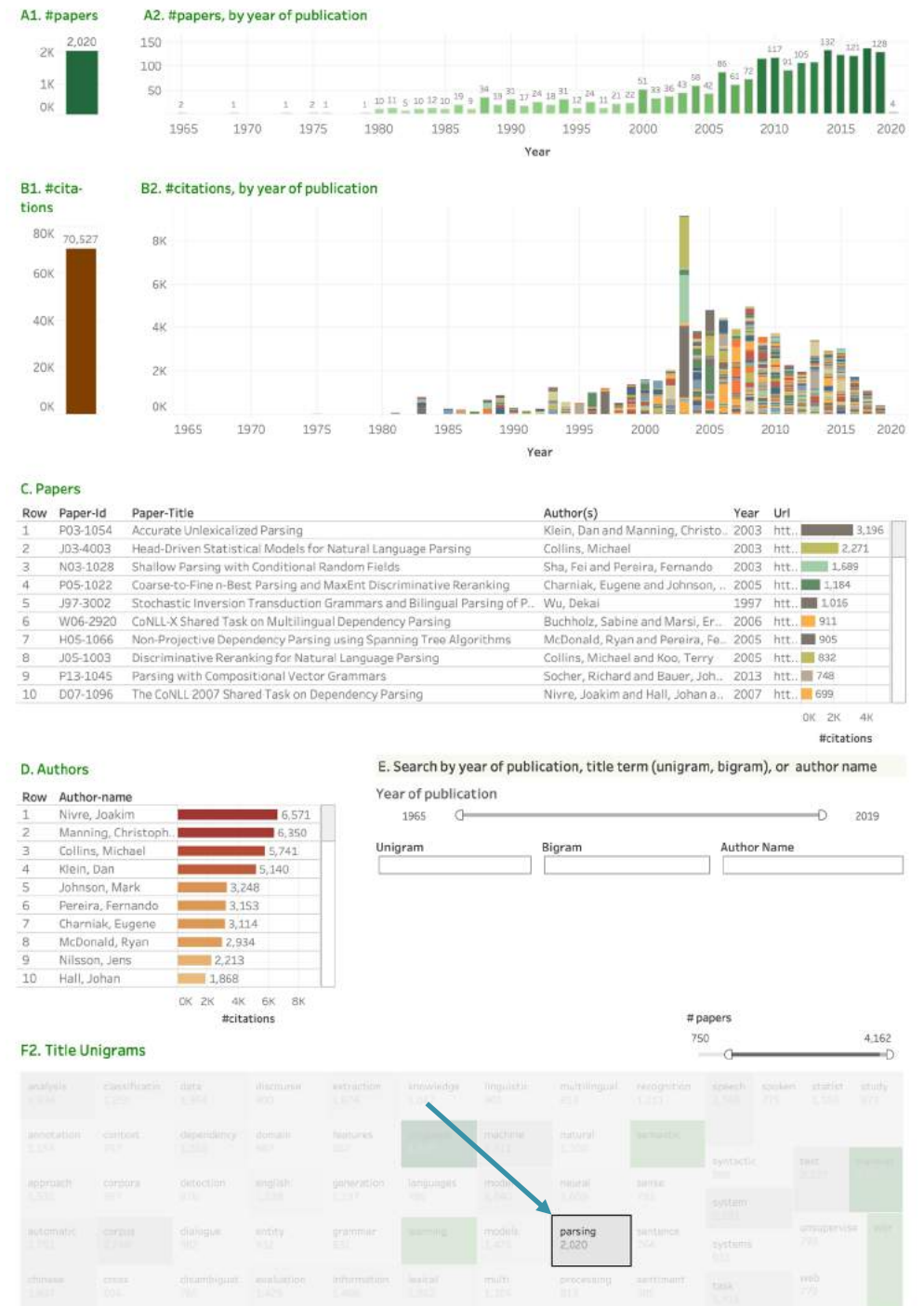
NLP Scholar Dashboard

Clicking on 'Workshops' in the
Venue and Paper Type treemap (F1)



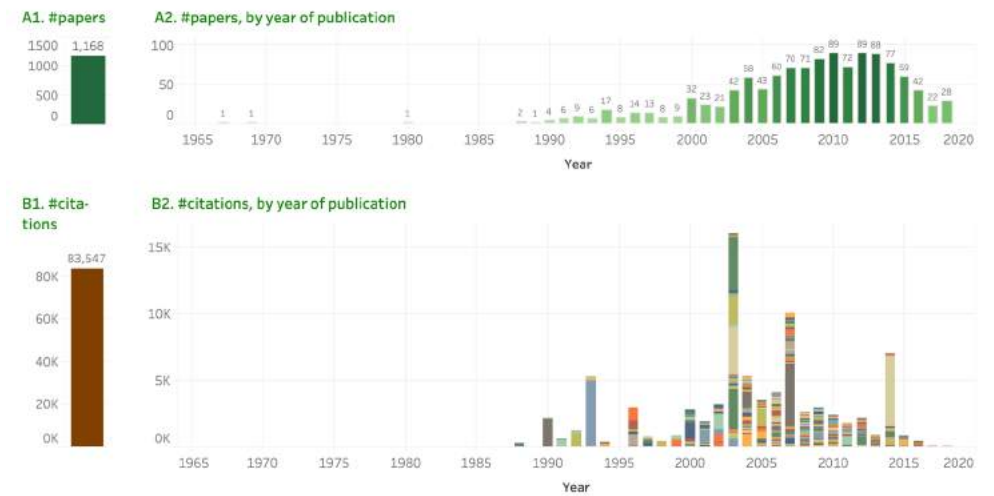
NLP Scholar Dashboard

Clicking on 'parsing' in the
Title Unigrams treemap (F2)



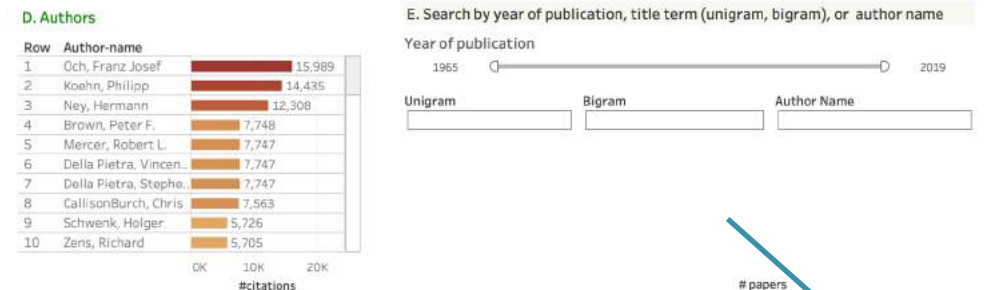
NLP Scholar Dashboard

Clicking on 'statistical' in the
Title Unigrams treemap (F2)



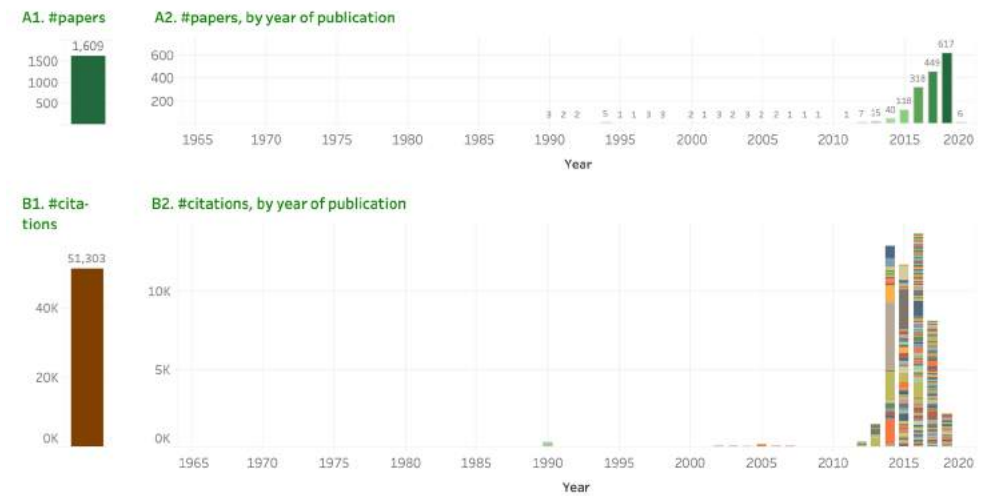
C. Papers

Row	Paper-id	Paper-Title	Author(s)	Year	Url	#citations
1	D14-1179	Learning Phrase Representations using RNN Encoder-Decoder for Stati...	Cho, Kyunghyun and van Merrië...	2014	htt...	5,344
2	J93-2003	The Mathematics of Statistical Machine Translation: Parameter Estima...	Brown, Peter F. and Della Pietra...	1993	htt...	5,047
3	P07-2045	Moses: Open Source Toolkit for Statistical Machine Translation	Koehn, Philipp and Hoang, Hieu...	2007	htt...	4,581
4	J03-1002	A Systematic Comparison of Various Statistical Alignment Models	Och, Franz Josef and Ney, Herm...	2003	htt...	4,040
5	N03-1017	Statistical Phrase-Based Translation	Koehn, Philipp and Och, Franz J...	2003	htt...	3,501
6	P03-1021	Minimum Error Rate Training in Statistical Machine Translation	Och, Franz Josef	2003	htt...	3,023
7	J03-4003	Head-Driven Statistical Models for Natural Language Parsing	Collins, Michael	2003	htt...	2,271
8	J90-2002	A Statistical Approach to Machine Translation	Brown, Peter F. and Cocke, John...	1990	htt...	2,102
9	P05-1033	A Hierarchical Phrase-Based Model for Statistical Machine Translation	Chiang, David	2005	htt...	1,288
10	P02-1038	Discriminative Training and Maximum Entropy Models for Statistical M...	Och, Franz Josef and Ney, Herm...	2002	htt...	1,240



NLP Scholar Dashboard

Clicking on 'neural' in the
Title Unigrams treemap (F2)



C. Papers

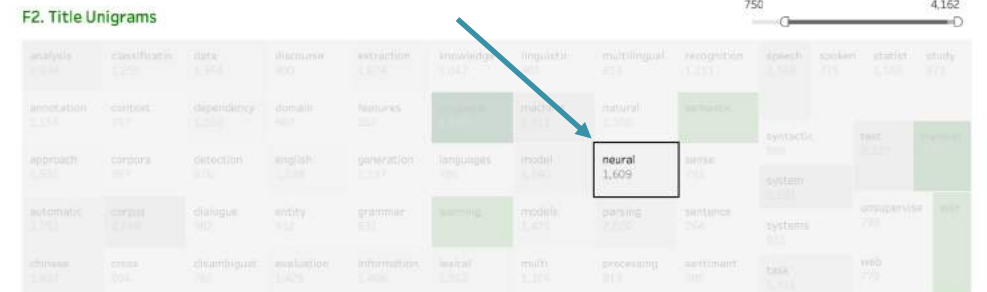
Row	Paper-id	Paper-Title	Author(s)	Year	Url	#citations
1	D14-1181	Convolutional Neural Networks for Sentence Classification	Kim, Yoon	2014	htt...	4,362
2	D15-1166	Effective Approaches to Attention-based Neural Machine Translation	Luong, Minh-Thang and Pham, ...	2015	htt...	1,961
3	P14-1062	A Convolutional Neural Network for Modelling Sentences	Kalchbrenner, Nal and Grefenst...	2014	htt...	1,794
4	W14-4012	On the Properties of Neural Machine Translation: Encoder-Decoder App...	Cho, Kyunghyun and van Merri...	2014	htt...	1,673
5	D14-1082	A Fast and Accurate Dependency Parser using Neural Networks	Chen, Danqi and Manning, Chris...	2014	htt...	1,110
6	P16-1162	Neural Machine Translation of Rare Words with Subword Units	Sennrich, Rico and Haddow, Bar...	2016	htt...	1,028
7	N16-1030	Neural Architectures for Named Entity Recognition	Lample, Guillaume and Balleste...	2016	htt...	957
8	D15-1044	A Neural Attention Model for Abstractive Sentence Summarization	Rush, Alexander M. and Chopra,...	2015	htt...	910
9	C14-1008	Deep Convolutional Neural Networks for Sentiment Analysis of Short Te...	dos Santos, Cicero and Gatti, M...	2014	htt...	697
10	D15-1167	Document Modeling with Gated Recurrent Neural Network for Sentime...	Tang, Duyu and Qin, Bing and Li...	2015	htt...	606



E. Search by year of publication, title term (unigram, bigram), or author name

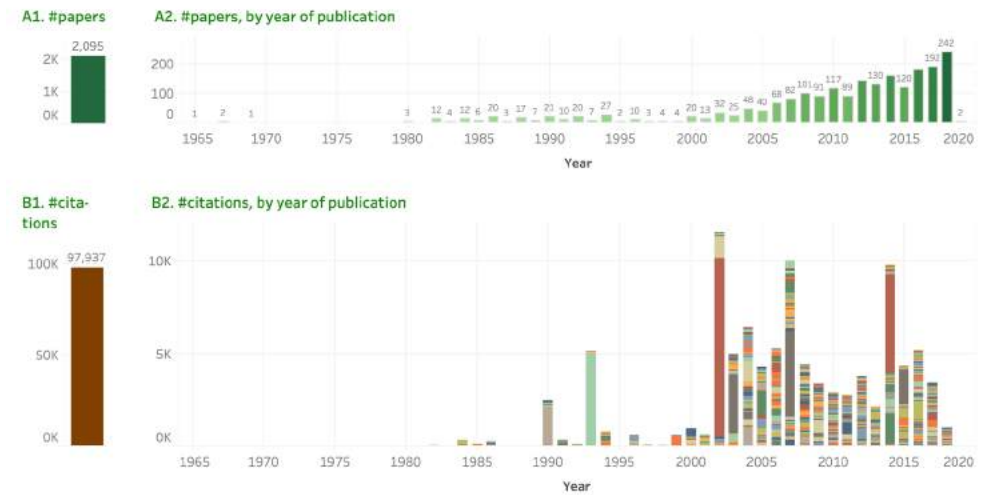
Year of publication: 1965 to 2019

Unigram: [] Bigram: [] Author Name: []



NLP Scholar Dashboard

Clicking on `machine translation` in the Title Bigrams treemap (F3)



C. Papers

Row	Paper-id	Paper-Title	Author(s)	Year	Url	#citations
1	P02-1040	Bleu: a Method for Automatic Evaluation of Machine Translation	Papineni, Kishore and Roukos, S.	2002	htt...	9,098
2	D14-1179	Learning Phrase Representations using RNN Encoder-Decoder for Stati...	Cho, Kyunghyun and van Merrie...	2014	htt...	5,344
3	J93-2003	The Mathematics of Statistical Machine Translation: Parameter Estima...	Brown, Peter F. and Della Pietra...	1993	htt...	5,047
4	P07-2045	Moses: Open Source Toolkit for Statistical Machine Translation	Koehn, Philipp and Hoang, Hieu...	2007	htt...	4,581
5	P03-1021	Minimum Error Rate Training in Statistical Machine Translation	Och, Franz Josef	2003	htt...	3,023
6	J90-2002	A Statistical Approach to Machine Translation	Brown, Peter F. and Cocke, John...	1990	htt...	2,102
7	D15-1166	Effective Approaches to Attention-based Neural Machine Translation	Luong, Minh-Thang and Pham, ...	2015	htt...	1,961
8	W14-4012	On the Properties of Neural Machine Translation: Encoder-Decoder App...	Cho, Kyunghyun and van Merrie...	2014	htt...	1,673
9	P05-1033	A Hierarchical Phrase-Based Model for Statistical Machine Translation	Chiang, David	2005	htt...	1,298
10	P02-1038	Discriminative Training and Maximum Entropy Models for Statistical M...	Och, Franz Josef and Ney, Herm...	2002	htt...	1,240

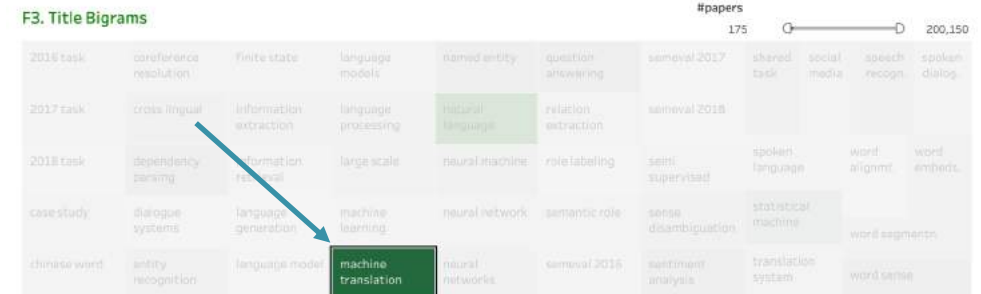
D. Authors

Row	Author-name	#citations
1	Koehn, Philipp	12,705
2	Papineni, Kishore	9,295
3	Roukos, Salim	9,217
4	Ward, Todd	9,101
5	Zhu, WeiJing	9,098
6	Och, Franz Josef	8,947
7	CallisonBurch, Chris	8,865
8	Cho, Kyunghyun	8,526
9	Bengio, Yoshua	8,060
10	Brown, Peter F.	7,375

E. Search by year of publication, title term (unigram, bigram), or author name

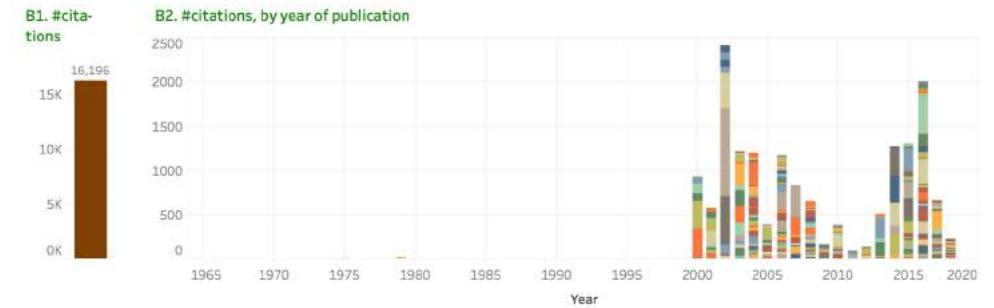
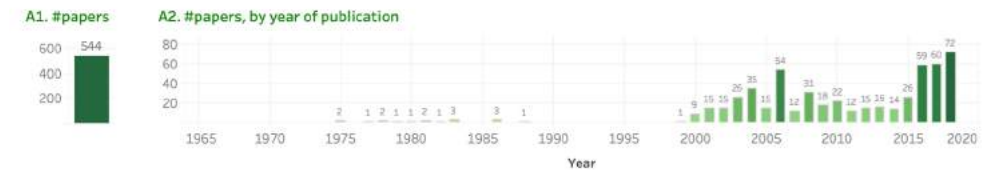
Year of publication: 1965 ————— 2019

Unigram: [] Bigram: [] Author Name: []



NLP Scholar Dashboard

Clicking on 'question answering' in the Title Bigrams treemap (F3)



C. Papers

Row	Paper-Id	Paper-Title	Author(s)	Year	Url	#citations
1	P02-1006	Learning surface text patterns for a Question Answering System	Ravichandran, Deepak and Hov...	2002	htt...	1,001.0
2	D16-1044	Multimodal Compact Bilinear Pooling for Visual Question Answering an...	Fukui, Akira and Park, Dong Huk...	2016	htt...	430.0
3	W02-1033	An Analysis of the AskMSR Question-Answering System	Brill, Eric and Dumais, Susan an...	2002	htt...	384.0
4	D07-1002	Using Semantic Roles to Improve Question Answering	Shen, Dan and Lapata, Mirella	2007	htt...	354.0
5	P02-1005	Performance Issues and Error Analysis in an Open-Domain Question Ans...	Moldovan, Dan and Pasca, Mari...	2002	htt...	350.0
6	D14-1067	Question Answering with Subgraph Embeddings	Bordes, Antoine and Chopra, Su...	2014	htt...	318.0
7	P14-1090	Information Extraction over Structured Data: Question Answering with ...	Yao, Xuchen and Van Durme, Be...	2014	htt...	256.0
8	N16-1181	Learning to Compose Neural Networks for Question Answering	Andreas, Jacob and Rohrbach, ..	2016	htt...	255.0
9	D14-1070	A Neural Network for Factoid Question Answering over Paragraphs	Iyyer, Mohit and BoydGraber, J...	2014	htt...	255.0
10	D15-1237	WikiQA: A Challenge Dataset for Open-Domain Question Answering	Yang, Yi and Yih, Wentau and M...	2015	htt...	250.0

D. Authors

Row	Author-name	#citations
1	Harabagiu, Sanda	1,800
2	Moldovan, Dan	1,204
3	Hovy, Eduard	1,204
4	Ravichandran, Deep...	1,045
5	Yih, Wentau	985
6	Pasca, Marius	939
7	Meek, Christopher	695
8	Rohrbach, Marcus	685
9	Darrell, Trevor	685
10	Molla, Diego	605

E. Search by year of publication, title term (unigram, bigram), or author name

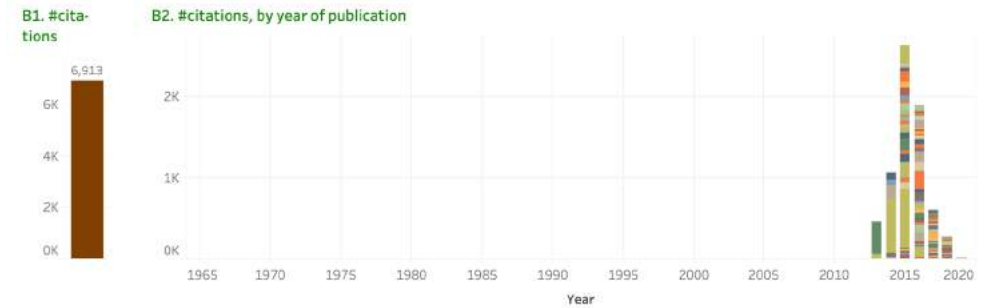
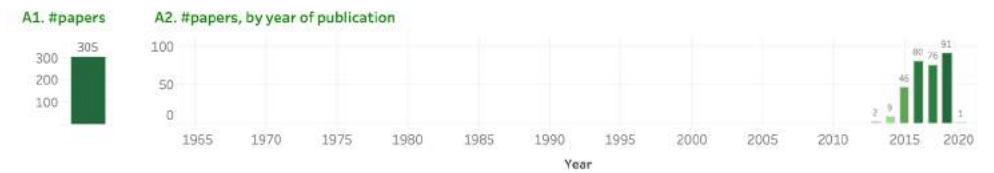
Year of publication: 1965 2019

Unigram Bigram Author Name



NLP Scholar Dashboard

Clicking on 'word embeddings' in the Title Bigrams treemap (F3)



C. Papers

Row	Paper-id	Paper-Title	Author(s)	Year	Url	#citations
1	Q15-1016	Improving Distributional Similarity with Lessons Learned from Word E...	Levy, Omer and Goldberg, Yoav	2015	htt...	718.0
2	P14-2050	Dependency-Based Word Embeddings	Levy, Omer and Goldberg, Yoav	2014	htt...	673.0
3	D13-1141	Bilingual Word Embeddings for Phrase-Based Machine Translation	Zou, Will Y. and Socher, Richard	2013	htt...	395.0
4	D15-1036	Evaluation methods for unsupervised word embeddings	Schnabel, Tobias and Labutov, I.	2015	htt...	233.0
5	P16-1141	Diachronic Word Embeddings Reveal Statistical Laws of Semantic Chan...	Hamilton, William L. and Leskov...	2016	htt...	215.0
6	P15-1077	Gaussian LDA for Topic Models with Word Embeddings	Das, Rajarshi and Zaheer, Manz...	2015	htt...	152.0
7	P14-1113	Learning Semantic Hierarchies via Word Embeddings	Fu, Ruiji and Guo, Jiang and Qin...	2014	htt...	152.0
8	D15-1168	Fine-grained Opinion Mining with Recurrent Neural Networks and Word...	Liu, Pengfei and Joty, Shafiq an...	2015	htt...	130.0
9	P16-1035	Query Expansion with Locally-Trained Word Embeddings	Diaz, Fernando and Mitra, Bhas...	2016	htt...	127.0
10	P15-2070	PPDB 2.0: Better paraphrase ranking, fine-grained entailment relations...	Pavlick, Ellie and Rastogi, Push...	2015	htt...	120.0

D. Authors

Row	Author-name	#citations
1	Goldberg, Yoav	1,438
2	Levy, Omer	1,422
3	Dagan, Ido	718
4	Manning, Christoph...	427
5	Dyer, Chris	401
6	Zou, Will Y.	395
7	Socher, Richard	395
8	Cer, Daniel	395
9	Labaka, Gorka	244
10	Artetxe, Mikel	244

E. Search by year of publication, title term (unigram, bigram), or author name

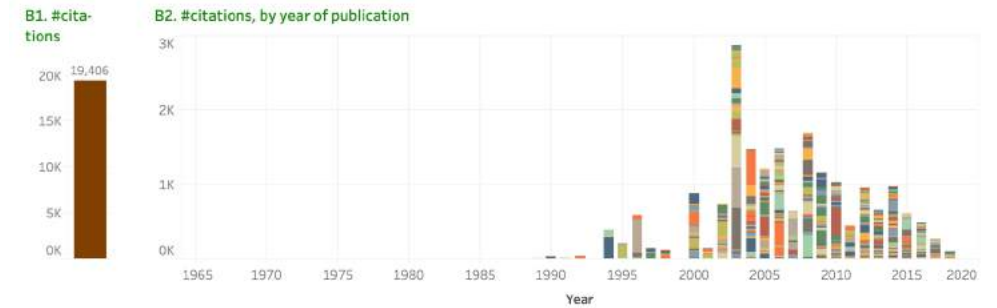
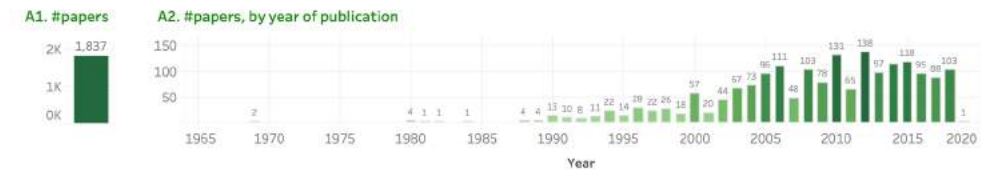
Year of publication: 1965 2019

Unigram Bigram Author Name



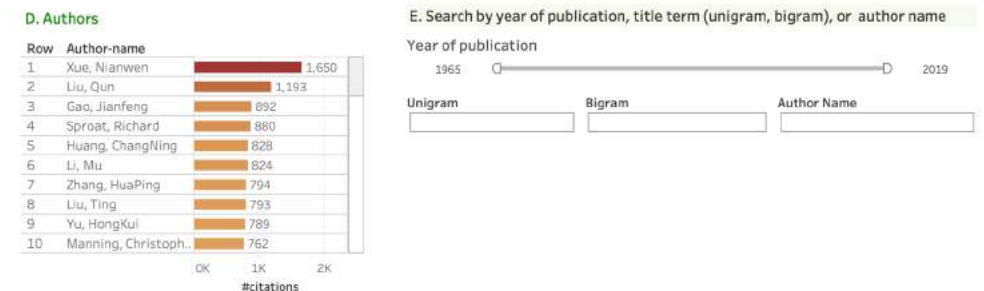
NLP Scholar Dashboard

Clicking on 'Chinese' in the Languages treemap (F4)



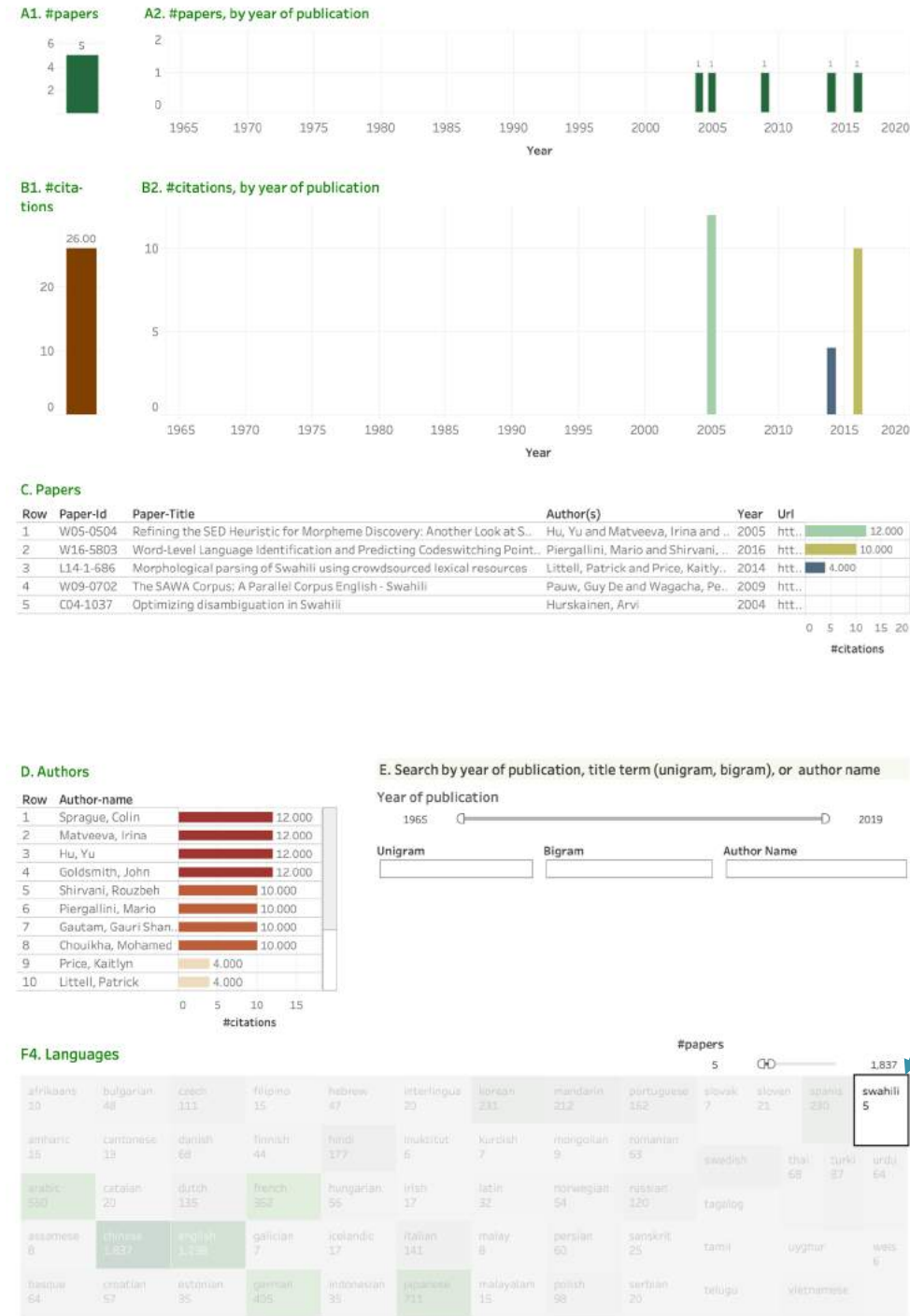
C. Papers

Row	Paper-id	Paper-Title	Author(s)	Year	Url	#citations
1	W03-1730	HHMM-based Chinese Lexical Analyzer ICTCLAS	Zhang, HuaPing and Yu, HongKu..	2003	htt...	545.0
2	W03-1728	Chinese Word Segmentation as LMR Tagging	Xue, Nianwen and Shen, Libin	2003	htt...	517.0
3	C04-1081	Chinese Segmentation and New Word Detection using Conditional Rand..	Peng, Fuchun and Feng, Fangfa..	2004	htt...	480.0
4	J96-3004	A Stochastic Finite-State Word-Segmentation Algorithm for Chinese	Sproat, Richard and Shih, Chilin ..	1996	htt...	450.0
5	W03-1719	The First International Chinese Word Segmentation Bakeoff	Sproat, Richard and Emerson, T..	2003	htt...	406.0
6	C10-3004	LTP: A Chinese Language Technology Platform	Che, Wanxiang and Li, Zhenghu..	2010	htt...	367.0
7	W06-3812	Chinese Whispers - an Efficient Graph Clustering Algorithm and Its Appli..	Biemann, Chris	2006	htt...	310.0
8	P94-1012	ALIGNING A PARALLEL ENGLISH-CHINESE CORPUS STATISTICALLY WITH...	Wu, Dekai	1994	htt...	295.0
9	W08-0336	Optimizing Chinese Word Segmentation for Machine Translation Perfor...	Chang, PiChuan and Galley, Mic..	2008	htt...	293.0
10	P03-1056	Is it Harder to Parse Chinese, or the Chinese Treebank?	Levy, Roger and Manning, Chris..	2003	htt...	272.0



NLP Scholar Dashboard

Clicking on ‘Swahili’ in the Languages treemap (F4)



Summary

NLP Scholar—an interactive visual explorer for the ACL Anthology

- access to citation information from Google Scholar
- interconnected interactive visualizations (multiple dashboards)
- quickly and efficiently search for relevant related work

Future Work:

- automatically identifying related papers
- analyze NLP papers that are published outside of the ACL Anthology
- compare patterns of citations in NLP with those in other fields
- **develop richer ways of capturing scholarly impact**

Project page for NLP Scholar: <http://saifmohammad.com/WebPages/nlpscholar.html>

- data
- Interactive visualizations
- limitations and ethical considerations



Created by Symbolon
from Noun Project



Contact:



Saif.Mohammad@nrc-cnrc.gc.ca



@SaifMMohammad



National Research
Council Canada

Conseil national de
recherches Canada

Canada