Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Chloe Lee April 14, 2020

NLP Rankings: Publication-based Ranking System and Platform for NLP Research

by

Chloe Lee

Jinho D. Choi
Adviser

Department of Mathematics

Jinho D. Choi

Adviser

Bree Ettinger

Committee Member

Jeremy Jacobson

Committee Member

2020

NLP Rankings: Publication-based Ranking System and Platform for NLP Research

By

Chloe Lee

Jinho D. Choi

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Department of Mathematics

2020

Abstract

NLP Rankings: Publication-based Ranking System and Platform for NLP Research

By Chloe Lee

NLP Rankings is a metric-based ranking system and platform which ranks academic institutions in the United States, as well as researchers of these academic institutions, focusing only on the field of Natural Language Processing (NLP). As existing rankings are either opinion-based or too generic to be useful for current and prospective researchers to gain knowledge about NLP programs, NLP Rankings aims to provide insights to prospective NLP students and current faculties by using publications from multiple venues on ACL Anthology, published between 2010 and 2019. The publication-based ranking scores provide information that the research community may be interested in, from publication advancement trend over the decade at each institutions, to a variation of h-index used to evaluate research achievement at an author-level. NLP Rankings is also publicly available, and user analysis suggests the usefulness of NLP Rankings in the research community.

NLP Rankings: Publication-based Ranking System and Platform for NLP Research

By

Chloe Lee

Jinho D. Choi

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Department of Mathematics

2020

Acknowledgements

I would like to thank my thesis advisor Dr. Choi, for giving me the opportunity to work on such an awesome project. As the head of Emory NLP, he is especially busy and has many advisees under him. However, over the year, he has given me many supports and encouragement that allows me to complete this project. I would also like to thank my committee members, Bree Ettinger and Jeremy Jacobson, for taking the time to attend my defense.

Table of Contents

List of Tables

List of Figures

1. Introduction

As the Information Age evolves with the wave of big data, demand to analyze unstructured textual data increases, which brought tremendous attention to the field of Natural Language Processing (NLP) and resulted in numerous emergences of higher education to NLP programs at academic institutions. Each university is unique and has its own strengths, making it difficult for prospective students and faculty candidates to choose the right programs to apply to. There have been several university rankings which attempt to provide such valuable insight to the quality of universities, but none dedicated specifically to the field of NLP. NLP Rankings, a publication-based ranking system and platform, is thus created to serve the research community in NLP, focusing on academic institutions in the United States. NLP Rankings[1] is also publicly available on the website to offer accessible information regarding the research environments of academic institutions in the United States.

This ranking system uses research publications on reputable NLP journals and conferences collected from ACL Anthology as the indicators to derive the ranking score. Focusing only on the aspect of academic research achievement, this ranking should be used in conjunction with traditional generic rankings to evaluate program performance and suitability, such that researchers can make informed decisions to best proceed with their careers in NLP.

A review of the benefits and limitations of the existing rankings is discussed in Section 2. Section 3 describes the data collection and data cleaning processes, as well as the scoring mechanisms used to derive the NLP Rankings. Section 4 presents a short demonstration of the publicly available platform and how users may customize by the weights of different publication venues and conferences to yield rankings that best fit their needs and preferences. Analyses on

[1] http://nlprankings.org

different levels are conducted in Section 5 to provide more supporting information that provides valuable insights to prospective students and faculty candidates, as well as an evaluation of the usefulness of such ranking to the research community. At last, the conclusion and discussion of findings, along with next steps to improve NLP Rankings are addressed in Section 6.

2. Related Works

2.1. Generic University Rankings

There is a substantial number of well-established rankings for universities in the United States and around the world. Most often, rankings are generic rankings that consider multiple indicators which the evaluators believe effectively capture university performances. A common indicator of generic rankings are the opinions of peers and professionals, which forces the rankings less objective and manipulatable.

2.1.1. U.S. News Rankings[2]

The most well-known university ranking, U.S. News, has been providing education ranking since 1983, serving as a valuable reference to students when deciding on their future. Similar to its undergraduate rankings which incorporates various indictors, U.S. News' Best Graduate Schools rankings are based on two areas:

1. expert opinions about the program excellence
2. statistical indicators that measure the quality of a school's faculty, research, and students

In particular, the data used to calculate U.S. News rankings comes from statistical surveys answered by academic professionals (i.e. deans, program directors and senior faculty), as well as statistical indicators of admission criteria, student-faculty ratio, job placement success, etc. To arrive at the final rank, U.S. News applies different and undisclosed weights to each indicator based on their judgment of relative importance. Because experts tend to praise the academic institutions they had studied at or worked for, U.S. News rankings are sometimes deemed unreliable because of this factor.

[2] https://www.usnews.com/

2.1.2. QS World University Rankings[3]

International university rankings also take similar generic approach to rank academic institutions. Published annually by Quacquarelli Symonds (QS), QS World University Rankings was first produced in 2004 as the response to the perceived need for an international ranking of universities. This ranking is derived according to six metrics with the respective weights:

1. Academic Reputation (40%)
2. Employer Reputation (10%)
3. Faculty/Student Ratio (20%)
4. Citations per faculty (20%)
5. International Faculty Ratio (5%)
6. International Student Ratio (5%)

As showed in the weightings of each metrics, 50% (*Academic Reputation* and *Employer Reputation*) of the university evaluation is opinion-based, as the scores for the two indicators come from *Academic Survey* and *Employer Survey* collected by QS. Institutional research quality only takes up 20% of the ranking, and is reflected via the *Citations per Faculty* metric, which may be susceptible to “citation cartels”[4].

2.2. Publication-Based University Rankings

On the other hand, as opposed to generic rankings that incorporates multiple factors, some university rankings focus on field-based ranking, reflecting the level of research

[3] https://www.topuniversities.com/

[4] The phenomenon of a group of publication authors citing each other disproportionately more than they do other groups of authors in the same field.

achievement to serve the population in the academia. Different from generic university rankings, these rankings rank universities by scientific publications volume and impact.

2.2.1. NTU Ranking[5]

Originally published by the Higher Education Evaluation and Accreditation Council of Taiwan (HEEACT) in 2007 and continued by the National Taiwan University (NTU) since 2012, the NTU Ranking uses bibliometric methodology to rank universities by its scientific paper performances. Other than an overall scientific publication ranking of all international universities, NTU Ranking publishes subject rankings for various academic fields.

The criteria and overall performance indicators with weightings for the rankings are as followed:

1. Research productivity (25%)
   - Number of articles in previous years[6] (10%)
   - Number of articles in the current year (15%)
2. Research Impact (35%)
   - Number of citations in previous years (15%)
   - Number of citations in the last 2 years (10%)
   - Average number of citations in previous years (10%)
3. Research Excellence (40%)
   - H-index[7] of the last 2 years (10%)
   - Number of highly cited papers in previous years (15%)

---

[5] Also known as Performance Ranking of Scientific Papers for World Universities; http://nturanking.lis.ntu.edu.tw/
[6] Starting from 2008 to current year
[7] A metric used to evaluate the cumulative scholarly impact of an author's by measuring measure both the productivity and citation impact of one’s academic publications

- Number of articles in the current year in high-impact journals (15%)

Indicated by the three major criteria, publication-based rankings reflect the university's publication quantity and quality, with a much greater emphasis on quality. However, in NTU Ranking, the quality of publications is represented by journals prestige and citation frequency, where the latter is less objective as mentioned before.

2.2.2. CSRankings[8]

CSRankings is a commonly used ranking for Computer Science programs compiled by Emery Berger, which takes multiple computer science research areas into consideration. With the intention to undermine influential rankings which are based entirely on reputation and rely heavily of surveys, the approach to CSRankings is entirely metrics-based, ranking universities by their presence at prestigious publication venues. The score for each university is calculated by summing the credits of its faculties. Aiming to construct an unbiased ranking, citation-based metrics is not included to avoid "citation cartels", and prestigious conferences and journals act as a proxy to avoid potential gaming of the ranking system.

Regardless of the popularity and usefulness of the previous mentioned university rankings, these existing rankings serve purposes different from what people interested in NLP research are looking for. U.S. News and QS ranking provide generic and subjective rankings, which are more suitable for students who are interested in the industry than in academia. NTU Ranking, though offers field-based ranking to different subjects, does not has a specific ranking dedicated only to

[8] http://csrankings.org/

NLP; and it reflects publication quality by using citation-based metrics, which may be deceiving given the widely seen "citation cartel" phenomenon in the academia. CSRankings, though very useful to people interested in different Computer Science programs, only considers publications from the venues ACL, EMNLP, and NAACL for its ranking of NLP programs. Furthermore, on CSRankings, publications from different journals and conferences carry equal value, whereas some people may hold different opinions.

3. NLP Rankings

Considering the benefits and limitations of existing rankings, NLP Rankings focus only on the field of NLP and are metrics-based such that the score of each author is determined by the number of one's publications and co-authors. The score of each university is calculated by summing the scores of all authors who have indicated that university as their primary appointment for the work. A large number of publications on NLP is collected to derive NLP Rankings (Section 3.1). Authors of those publications are then matched to their primary institutions using Levenshtein distance (Section 3.2). Finally, the score of each author is calculated and summed to obtain the final ranking score for each university (Section 3.3).

3.1. Data Collection

All publications used to derive the rankings are collected from the ACL Anthology[9], the largest open-source web-based platform that hosts publications from various venues focusing on NLP. For NLP Rankings, long and short papers published between 2010 and 2019 from the following venues are collected:

- Annual Meeting of the Association for Computational Linguistics (ACL)
- Computational Linguistics (CL)
- International Conference on Computational Linguistics (COLING)
- Conference on Computational Natural Language Learning (CoNLL)
- European Chapter of ACL (EACL)
- Conference on Empirical Methods in NLP (EMNLP)
- International Joint Conference on NLP (IJCNLP)
- North American Chapter of ACL (NAACL)

---

[9] https://www.aclweb.org/anthology/

- Transactions of the Association for Computational Linguistics (TACL)

Additionally, workshop and demonstration paper (WS) with lengths that are more than 4 pages are collected[10], which include workshops hosted by ACL events, student research workshops and system demonstrations hosted by the above venues, as well as conferences/workshops hosted by the Special Interest Groups (SIG*), Conference on Lexical and Computational Semantics and Semantic Evaluation (*SEMEVAL).

	ACL	CL	EMNLP	NAACL	TACL	COLING	CoNLL	EACL	IJCNLP	WS	Total
W	3	3	3	3	3	2	2	2	2	1	-
$\lvert V \rvert$	20	37	10	10	7	12	9	5	7	669	786
$\lvert P \rvert$	3,334	284	2,973	1,558	273	1,762	345	448	587	13,332	24,896
$\lvert A \rvert$	11,417	741	10,801	5,301	916	5,927	1,282	1,349	1,937	42,559	82,230
$\{A\}$	5,483	579	5,529	3,268	683	3,809	1,071	1,039	1,043	17,431	24,838
$\lvert P_u \rvert$	1,173	36	1,076	689	129	286	94	99	110	2,569	6,261
$\lvert A_u \rvert$	3,918	139	3,851	2,396	414	983	370	319	339	8,484	21,213
$\{A_u\}$	2,068	135	2,107	1,470	315	769	317	273	274	4,283	7,426

W: default weight; $\lvert V \rvert$: # of proceedings or issues; $\lvert P \rvert$: total # of publications; $\lvert A \rvert$: total # of authors; $\{A\}$: total # of unique authors; $\lvert P_u \rvert$: # of publications by academic authors in U.S.; $\lvert A_u \rvert$: # of academic authors in U.S.; $\{A_u\}$: # of unique academic authors in U.S.

Table 1: Statistics of the publications collected

All publications come with both a bibliography file (*.bib) comprising meta-information about the paper and the PDF file containing the publication contents. Although the bibliography files are structured, their formats are not necessarily consistent across different venues and years; thus, the information are converted into a consistent JSON format. All PDF files are converted into text files for further information extraction (Section 3.2). Note that some publications are disregarded if the PDF files are scan images that cannot be converted into text, which happens rarely for papers published in recent years.

[10] As most workshop/demonstration papers under 4 pages (including references) are found to be incomplete, they are discarded due to quality control

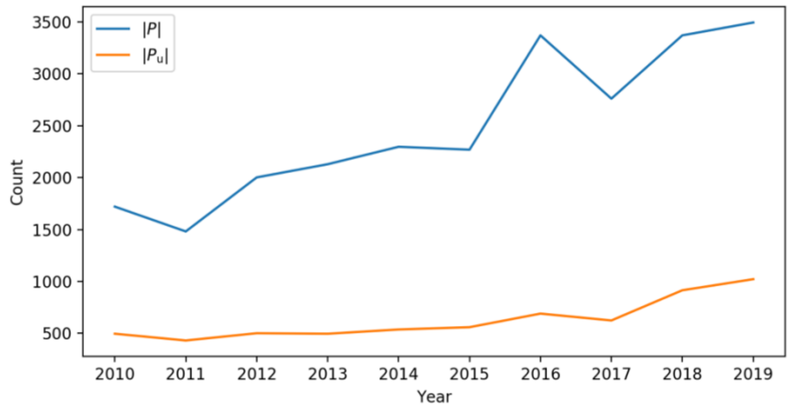


Figure 1.1: Number of NLP publications over the last 10 years

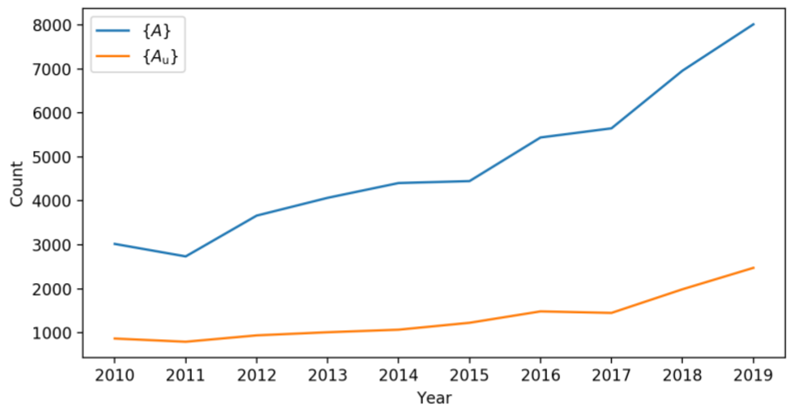


Figure 1.2: Number of NLP authors over the last 10 years

Figure 1.1 and 1.2 show the publication statistics on NLP over the last 10 years. As presented, the number of publications and the number of authors in NLP field are increasing over the past decade. Publications and authors from U.S. academic institutions consist of around 25-30% of the entire population.

3.2. Author-University Matching

Although not explicitly required, authors tend to include email addresses from their primary organizations in the publications as a contact method and a way of indicating institutional affiliation. It is common nowadays where an academic author belongs to multiple organizations, in which case, the author chooses which email address to include in the publication heading by considering the authorship of the work. For instance, if a professor worked at an industrial company during the summer and published a paper, one may choose to use the company's email address instead of the one from the university; in this case, the work is credited to the company.

Because email addresses are always presented on the first page in the original PDF files, which is roughly equivalent to the first 2,000 lines in the converted text files (Section 3.1), email addresses are extracted from the first 2,000 lines of text using a comprehensive group of regular expressions that consider almost all possible forms of email addresses. Special cases such as emails presented in groups where the IDs are quoted in brackets and separated by certain delimiters (e.g. {id1,id2}@institute.edu) as a substitution of listing individual emails are also captured by the expressions. The set of regular expressions yields an 85.8% coverage of the publication authors.

Since emails are not always provided in the same order of the authors listed in the bibliography files, they are matched by minimizing the Levenshtein distance between the actual email addresses extracted and pseudo-generated email addresses with the authors' name.

Academic institutions typically use similar naming conventions for emails as follows (*f*/*m*/*l*: the initial of the first/middle/last name, *(m)* is optional):

- *firstname lastname*
- *f (m) lastname*
- *lastname f (m)*
- *firstname*

- *lastname*
- *f (m) l*

To match emails and authors accurately, the matrix $M \in \mathbb{R}^{e \times c}$ is created for every publication, where e is the number of extracted email addresses, c is $n \cdot a$, $n = 6$ is the number of naming conventions above, and a is the number of authors in the corresponding bibliography file. Essentially, an email address is pseudo-generated per column by substituting the corresponding author's first name, last name, and initials if applicable. Each cell in M is then filled with the Levenshtein distance between the corresponding row and column. Finally, the rows and columns are matched by taking the $argmin$ of the matrix one after another so that its corresponding author is matched to the email represented by the row by the most similar to the least. A publication may have more authors than emails, in which case, the contribution of the unmatched authors are discarded from scoring (Section 3.3).

3.3. Scoring Mechanism

Given the advantages and disadvantages of existing rankings mentioned in Section 2, NLP Rankings ranks universities in the United States on a metric-based scoring based on academic publications, which best serves the needs of prospective NLP students and current researchers in the field. Distinguished from generic university rankings (Section 2.1), NLP Rankings does not consider experts opinions, but only focuses on research achievements reflected by academic publications. Distinguished from NTU Ranking (Section 2.2.1) which measures research impact by citations, NLP Rankings avoid such measurement as it is a gameable metric. Distinguished from CSRankings (Section 2.2.2), NLP Rankings is dedicated only to the field of NLP and have several different scoring features.

First, unlike CSRankings where different journals and conferences carry equal weights, in NLP Rankins, each publication is weighted by its venue and publication type. By default, papers from major venues (CL, TACL, ACL, NAACL, EMNLP) are credited with the weight of 3, other conferences with the weight of 2, and workshops/demonstrations with the weight of 1. Since the weighting can be subjective, weights are customizable on the NLP Rankings platform, which allows users to personalize the weights by their needs and preferences.

Following CSRankings, the credit of each publication is evenly distributed to all authors such that each author receives the score of $\frac{w}{a}$ where w is the weighted credit and a is the total number of the authors in the paper. For each publication, the score of each institution is measured by summing the scores of all authors from that institution using the matching algorithm described in Section 3.2.

Different from CSRankings, which reflects only the scores of faculty members to measure university score, NLP Rankings accounts for contributions from students as well. For instance, if there are 4 authors in the paper where 2 students and 1 professor are from the institution I_1 and 1 professor from the institution I_2, NLP Rankings scoring gives 75% of the credit to I_1 and 25% to I_2, whereas CSRankings gives 25% to both institutions, completely neglecting contributions from students, which may consequently yield misleading comparisons.

The third distinctive aspect of NLP Rankings' scoring mechanism is that it is sensitive to institutional authorship, such that scores earned by an author from one institution will not be transferred to another institution upon the author's move. Although a reputable author with numerous high-quality publications is very likely to continue a high performance at another institution, such expectation cannot be guaranteed because the research environment and student quality vary by institutions. As a substitution, NLP Rankings indicates how active each author is (or the presence of each author) in every institution by the year of the author's last publication dedicated to that particular institution.

4. Demonstration

NLP Rankings is publicly available online on http://nlprankings.org. Figure 2 shows the homepage of the platform, and the features and components are explained as follows.

4.1. Rankings

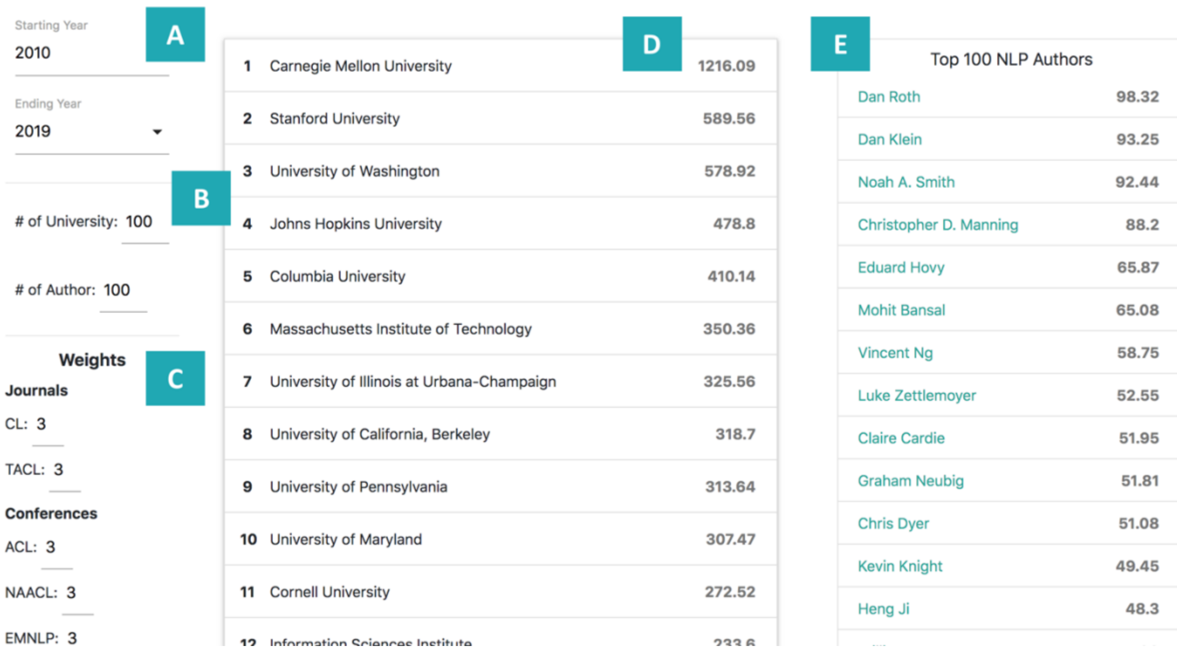


Figure 2: NLP Rankings Homepage User Interface

A. **Time Range** *Starting Year* and *Ending Year* filter the NLP publications that are used to calculate the rankings. Rankings will refresh immediately after selection.

B. **Display** As default, only the top 100 academic institutions and authors are displayed. Users may choose the number of academic institutions and researchers that are displayed on the platform.

C. **Weights** NLP Rankings is designed to allow users to weight publications based on their preference and understanding of the values of different publication venues and types. Users may increase or decrease the default credits, which updates both the institution and author rankings with customized weights.

D. **Academic Institution Ranking** By clicking on each institution name, the drop-down menu contains all the authors who had published for the institution, their respective scores achieved within the selected time range, the latest year each author published for the institution, as well as the total number of publications written by each author within the selected time range for the institution.

E. **Author Ranking** This ranking shows the top academic researchers with their respective scores.

4.2. Visualizations

The visualizations tab includes several interactive graphs that allow users to compare up to 5 different universities in different aspects.

1. **University Ranking Score Timeline** This graph shows the selected universities' ranking score from 2010 to 2019 in a stacked bar chart, where the lighter shading represents the amount of score contributed by top 10% of the authors in the university in terms of ranking scores.
2. **Number of Authors in Various Publication Amount** This stacked bar chart shows the number of authors in the university who have published only one, two, or more than three publications.
3. **Average Publication Percent Contribution Overtime** This line graph shows the average publication contribution percentage from 2010 to 2019, which indicates how likely universities co-author with different universities overtime.

4. **Average Number of Authors per Publication Overtime** This line graph shows the average number of authors per paper of each university from 2010 to 2019, which reflects how independent the researchers at each university are.

5. Analysis

5.1. University-Level Analysis

5.1.1. Top 50 Universities in the United States

Appendix A presents the top 50 universities from NLP Rankings between 2010 to 2019 based on the default weights mentioned in Section 3.3. Carnegie Mellon University, ranking first among the 216 universities with NLP publication in the United States, has a research population and an NLP ranking score which are double that of University of Washington, which ranks second. There are also a few other significant difference in ranking scores between ranks – Stanford University (3rd) and Johns Hopkins University (4th) has a score difference of 97.54, which is roughly equivalent to 33 long papers; Johns Hopkins University (4th) and Columbia University (5th) has a difference of 72.01, which is roughly equivalent to 24 long papers; Columbia University (5th) and Massachusetts Institute of Technology (6th) has a difference of 57.62, which is roughly equivalent to 19 long papers. The gaps between consecutive ranks decreases until the sixth rank, and the ranks below are closer in scores such that they are only a few long papers different from the previous.

The overall rank presents how universities had performed in the past ten-year period. However, to understand how universities have changed and grow, Appendix B presents the ranking of these top 50 universities (Appendix A) at each given year over the past decade, which shows several notable points.

First, top tier universities remained largely the same over time. Carnegie Mellon University, University of Washington, Stanford University, among other highly recognizable universities remained competitive in each of the given year. Carnegie Mellon, for instance, remained first for all ten years, indicating its strong academic achievement in NLP in the past decade. Top 10 universities such as University of Washington and Stanford University began lower in rank, but showed an upward movement in NLP Rankings year over year, and ended with ranks second and forth respectively in 2019.

However, some top tier universities have been decreasing in ranks in the recent years. For example, University of California, Berkeley used to rank third in 2010, but fell out of top 20 universities in 2019. Columbia University also did not perform as well as it has started off with. Nevertheless, among the top 50 universities, there are still more universities with rank jumps than falls. 29 of the 50 universities have a rank jump comparing between 2010 and 2019; 20 universities have a lower rank than it started with; 1 university remained the same. The average rank change between year 2010 and year 2019 is 15.52, indicating an overall improvement in ranking on average.

5.1.2. University Trend Clustering

Although ranking is a popular reference when students apply to graduate NLP programs, there are more factors that influence application choices than the rank of each programs. For instance, as mentioned in Section 5.1.1, the trend of each individual university's ranking score is an important indicator of current and future performances.

Since students generally apply to more than one university, either with a combination of similar programs or sets of different programs, a hierarchical cluster analysis is performed to cluster universities by their similarity in trends. Using agglomerative hierarchical clustering algorithm, universities are first grouped into individual clusters, and as the hierarchy moves up, the pairs of clusters are merged as a larger cluster. Eventually, all universities are merged into one group.

The purpose of using hierarchical clustering to group university, instead of other clustering methods, is to accommodate for the unknown k, which is number of clusters, as well as to have sub-clusters that potentially reveals universities with similar research interests and areas in NLP.

The cluster analysis uses Ward variance minimization algorithm to calculate the distance between the newly formed clusters and each untouched university (or cluster). The distance function is defined as followed:

$$d(c_i \cup c_j, c_k) = \sqrt{\frac{n_i + n_k}{n_i + n_j + n_k} d(c_i, c_k)^2 + \frac{n_j + n_k}{n_i + n_j + n_k} d(c_j, c_k)^2 - \frac{n_k}{n_i + n_j + n_k} d(c_i, c_j)^2}$$

where c_i, c_j, c_k are disjoint clusters with sizes n_i, n_j, and n_k.

Appendix C shows the dendrogram of the clustering result, which groups the 216 universities into 3 clusters on a high-level. The red cluster is the high tier which includes universities with high ranks and scores, the green cluster is the mid-low tier which includes universities that have relatively lower ranks and less NLP publications, and the blue cluster is the outlier Carnegie Mellon University, which outperforms all other universities in NLP research.

Of the 26 universities which are in the high tier group, all of them are top 30 universities in the overall NLP Ranking presented in Appendix A. The 3 universities that are top 30 but not in the high tier group are University of North Texas (26th), Brandeis University (28th), and Stony Brook University (29th). Based on the ranking trend in

Appendix B, these universities have a downward sloping trend, which may be the reason why they are not included in the high tier group.

There is a total of 189 universities in the mid-lower tier, with most of the university with very little NLP publications. Though some of the universities have substantial research advancement and are highly recognizable, for instance Ivy-Leagues like Brown University, they have been inactive recently as showed in Appendix B, causing them to be clustered to the mid-lower tier.

The sub-clusters in the high tier is also worth exploring, considering that most students wish to be admitted to top programs and thus apply to a few similar universities. Figure 3 shows the score trend of a sub-cluster of 4 universities in high tier. All 4 universities started strong in 2010, but did not perform as well until around mid 2010s, and showed an improvement between 2018 and 2019. Similar trends in publication score may infer similar research interest, where a given topic is more popular in a certain year, thus having more papers approved and published. Mid 2010s is when neural network in NLP became popular, and therefore NLP labs focusing on this area showed an increase in ranking score. In 2016, all 4 universities have several publications on topic modelling or neural learning in NLP.

However, the clustering result occasionally fails as an indicator to project future outlook of research quality, as it is still sensitive to past performance. For instance, University of Texas Dallas is grouped among the high tier, yet as shown in Appendix B, the recent three-year performance is not as optimal, where it dropped from rank 13 in 2010 to 64 in 2019.

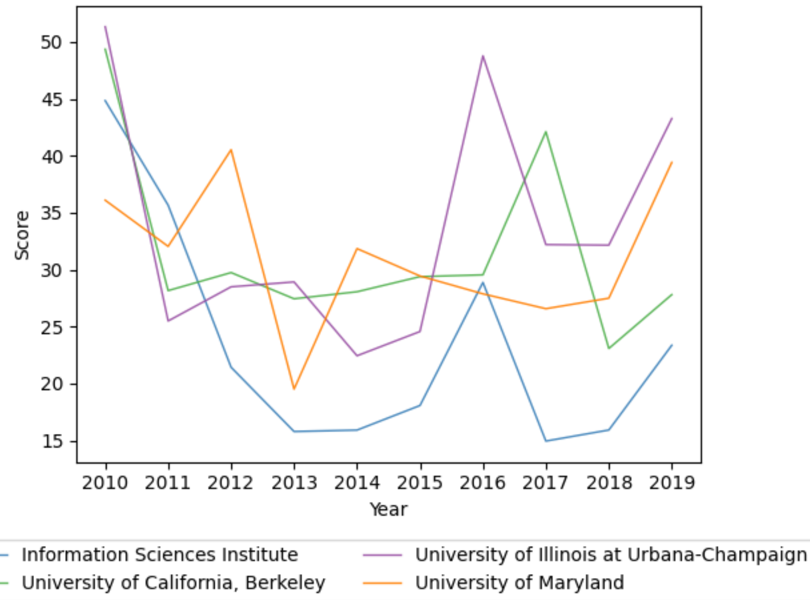


Figure 3: Score Trends of Clustered Universities

5.2. Author-Level Analysis

5.2.1. Top Universities Attended by Top 100 NLP Authors

In a young researcher's perspective, being in a research lab with famous researchers is beneficial in several ways. By following the lead of reputable researchers, young researchers can learn how to think like excellent scientists and how to publish in high impact journals, along with publication collaboration experiences with major innovators in the field. Although universities may influence these researchers only to an extent, the universities that top researchers had worked for or are currently working at indicates research environments of high standards.

The Author Ranking in NLP Rankings ranks top academic researchers with their respective cumulative publication scores, which sums all the publications scores one had published for an academic institution over the past decade. Out of the 7,426 authors who had published for a university in the United States, the universities that the top 100 authors have published for are showed in Figure 4.

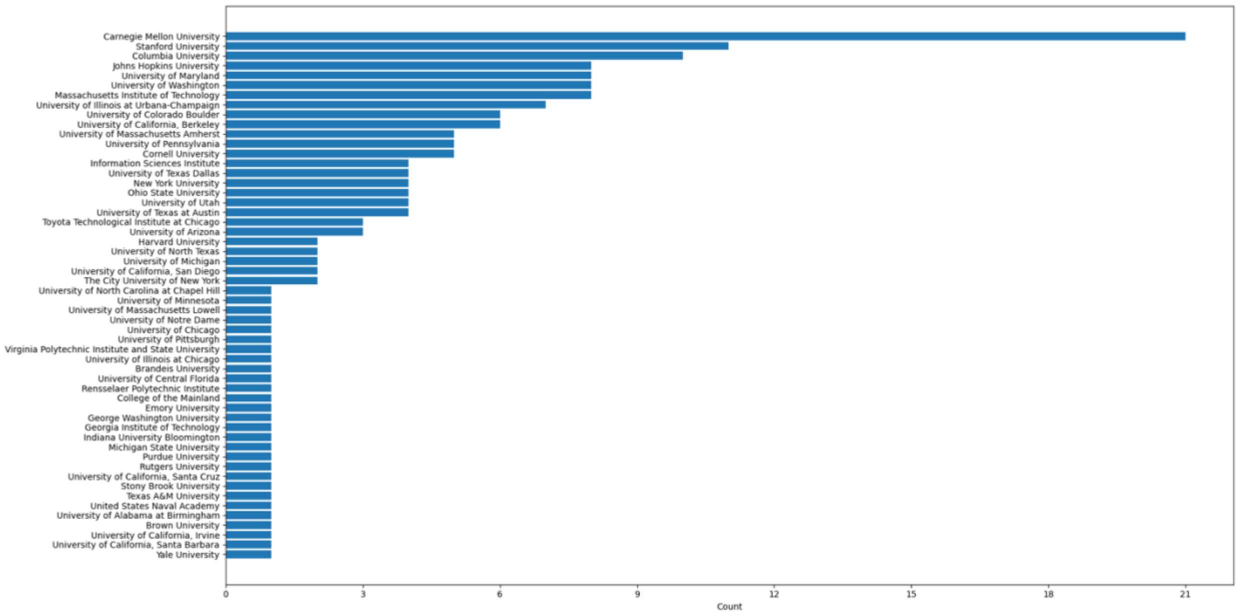


Figure 4: Universities Attended by Top 100 NLP Authors

The author ranking and the university ranking in NLP Rankings are correlated, as university ranking score is the sum of the author's score who have contributed one's publication to the university. A university will have higher ranking score if it has strong researchers. Therefore, not surprisingly, most top authors have worked at Carnegie Mellon University, followed by Stanford University and Columbia University.

The top universities that top NLP authors published for matches the top universities in the university ranking. However, the universities that have only one or two top NLP authors worked for also signify a potential rising research university in NLP. Although the overall ranking for such universities are relatively lower, the research environment may be favorable, in which case top NLP authors are hired to establish the new lab, and greater attentions will be given to the young researchers who freshly joined the program.

5.2.2. Authors Success Evaluation: weight-contribution index

University rankings are popular because academic researchers rely on such ranking information to assess the quality of their prospective research environment and gauge the expected research advancement they will experience. However, rankings may be deceiving because it is on a university-level, yet researchers care most about individual performance on an author-level.

To evaluate the publication excellence of individual authors, a common index is the h-index, suggested by Jorge E. Hirsch In 2005, which is defined as the number of papers with $citation\ number \geq h$ as an index to characterize the scientific output of a researcher in both aspects of publication productivity and citation impact. In short, achieving an h-index of h indicates that of the N_P publications of the author, h of them have at least h citations each, and the rest $N_P - h$ publications have less than h citations each.

This index has been commonly used in the academia as an indicator of individual researcher's research achievement, and widely accepted for applying to research fellowships and positions at research universities. As a robust indicator, it is a mathematically simple index and it encourages large amount of high-quality publications. However, such index has the same disadvantages as all indicators that use citations. A citation-based indicator is field-dependent and is easily manipulatable.

A proposed variation of the h-index for NLP Rankings, namely weight-contribution index (wc-index), is to revise the citation impact to publication journal prestige, while keeping the productivity aspect. In NLP Rankings, publication journals and conferences are assigned with a weight (Section 3.3) based on publication type and venue. Particular journals and conferences are viewed to be more prestigious than others. The fact that the author's paper is chosen by the journal or conference committee represents a high-level research achievement and academic quality.

Following the scoring mechanism of NLP Rankings where the credit of each publication is evenly distributed to all authors, the percent contribution to the publication is also considered. An author can have multiple publications in a year where none of the work is led by the author or conducted independently. The contribution percentage is calculated as $c = \frac{1}{a}$ where a is the total number of authors of the publication.

Assigned weights and contribution percentage are then combined to derive the wc-index, where if the product of the two is less than 1, the publication is discarded and not added toward the index. For instance, author A published an ACL paper, which accounts to 3 credits by default, along with 2 other authors. A receives a contribution percentage of $\frac{1}{3}$, and consequently the product of the weight and contribution percentage is 1. This ACL paper will be added towards the wc-index. On the other hand, if author A has another ACL paper which one published along with 3 other authors, the product is then 0.75, which is less than 1, this second paper will not be counted towards the wc-index. This awards researchers who publish on prestige journals and conferences independently, which translate to the success of an academic researcher.

The wc-index also shows the behavior and current status of researchers. With an updated wc-index that shows the cumulative wc-index in different time ranges (e.g.

2011 wc-index: 2010-2011, 2012 wc-index: 2010-2012, …), it shows how individual researchers have performed over the past decades.

Appendix D shows a comparison between the h-index and the wc-index since 2015 for top 30 NLP authors. By observing the trend of the updated wc-index year over year, it is clear that such index serves a better purpose for prospective students to understand the current faculty than the h-index. First of all, because wc-index includes the contribution component, prospective students can understand how involve the current faculties are with their students' researches. The current status of researchers can also be inferred by observing the trend of wc-index. For instance, trends that stayed flat over the past year are probably faculties who are no longer active in the academia, and trends that are upward sloping are young researchers who are actively involved in the research community.

5.3. User Analysis

Because NLP Rankings is an interactive platform that allows users to customize the settings to find rankings based on desired weights and between selected timeframes, the choices that users made when customizing the ranking is collected to analyze user's interests and behaviors.

5.3.1. Log Data Statistics

Over a period of 46 days, since the date it was launched on February 12, 2020, to March 28, 2020, there is a total of 3,913 accesses to NLP Rankings, coming from 1,219 distinct IP addresses. Of the 3,913 accesses, 97.3% only viewed the rankings during the default timeframe from 2010 to 2019. Given the short running time period of only 46 days, this shows that most users are checking NLP Rankings without exploring the site.

More recent years starting from 2015 appear to receive more attention than earlier years. Other than the default timeframe, 2015 and 2016 are the start years that are checked the most, followed by 2018 and 2017. Under the assumption that NLP research users probably check the years when they started their academic career, the result suggests that the year 2015 is the beginning of the emerging interest in NLP.

Start Year	End Year	Count
2010	2010	2
	2016	1
	2017	1
	2019	3806
2011	2011	1
	2019	1
2012	2019	4
2013	2019	3
2014	2019	4
2015	2016	1
	2019	27
2016	2016	1
	2019	24
2017	2019	11
2018	2010	1
	2019	13
2019	2019	2

Table 2: Time frame choices on NLP Rankings platform

5.3.2. Weight Customization

As described in Section 3.3, users are allowed to change the weights for different venue and publication type. However, 99.2% of the total accesses use the default weights to calculate the NLP Rankings. This suggests that users agree with the proposed values that different journals and conferences hold.

5.3.3. Re-Visit Frequency

To assess the usefulness of the platform, the logs from the same IP address but on different days over this 46-day period is analyzed. Of the 1,219 unique IP addresses, 73.9% of the users only viewed the site once, and never used the site again. 18.7% used it twice, and 3.0% checked it on three different days.

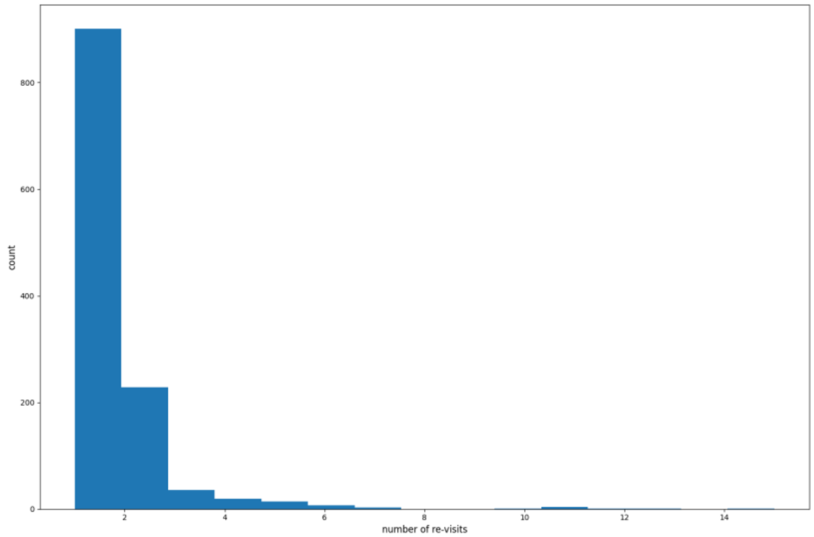


Figure 5: Histogram of unique IP re-visit frequency

The observations suggest that NLP Rankings is useful to NLP researchers to some extent. However, given the short running time and time of the year when NLP Rankings is launched, NLP Rankings is expected to receive more usage during application seasons later at the end of the year.

6. Conclusion and Discussions

Although numerous institutional rankings provide valuable information to current and prospective researchers when they are deciding on their future, NLP Rankings is a ranking system and platform dedicated specifically for the field of NLP, which provides more useful insights that the NLP research community requires.

To derive the rankings for university and NLP authors, publications on the open-source ACL Anthology published between 2010 to 2019 are collected and evaluated. Focusing only on universities in the United States, a total of 216 universities have publications on NLP over the past decade. Universities are then ranked accordingly based on the number of publications and weights assigned to each venue and issue type.

Different from other academic fields that have longer history and more well-established programs, Natural Language Processing is relatively a newer field and thus have limited references that can be used to evaluate the institutions objectively. NLP Rankings contain valuable information that may be useful to NLP researchers and students interested in the field to infer something about the quality of faculty and research productivity at different academic institutions. Furthermore, besides the research community, industry employers may also refer from the university ranking to evaluate candidates.

The analysis of university's ranking finds that although top tier universities remained competitive overtime, there are still a few notable programs that ascend into or descent out of the top tier. Still, improvements in rankings are more common in general. This phenomenon may very likely be due to the fact that NLP is a relatively new academic field, and academic institutions are only starting to establish the program in recent years. Strong institutions remain competitive as they have always been the lead in the field, thus attracting more competent young researchers to the program. Even though the time frame used to evaluate the program is short with only ten years of observations, given the ranking score trends of different programs, students may project the prospective research advancement at each institution.

The hierarchical clusters also grouped universities by the scoring trends over the past decade. With a clear dividing scheme that clusters 216 universities into 3 main groups, students have more information than the ranking. Because ranking is simply a summation of scores, it fails to provide future outlooks to the research environment. The clustering results appear to be reasonable, with Carnegie Mellon University alone as an outlier, most high-ranking universities clustered as one, and other lower-ranking universities as another.

The universities that famous NLP authors worked at also indicate the potentiality of a good research program for young NLP researcher, adding values beyond the rank so that newly established but emerging programs can be identified. The proposed weight-contribution index also allows trend analysis to researcher's publication quality and quantity, giving more information about the research achievement at an individual level.

By analyzing the user log information collected over a period of 46 days, the platform seems to provide information that the research community are interested in, based on the number of users who checked out the site as well as the frequency of re-visits. Because the time period when log information is collected is during application season, the site received limited views. However, further analysis should be conducted during application seasons to reevaluate the usefulness of the NLP Rankings platform.

NLP Rankings only rank universities based on their research in Natural Language Processing as a whole. However, there are still multiple research areas within this field. Different research interest and focus are also important factors to students who would like to pursue a research career in the academia. As a result, NLP Rankings can be further developed by conducting cluster analysis to identify main NLP research interests at each institution, as well as performing trend analysis and topic modeling to identify trending research topics over the past decade, to provide more valuable information that the NLP research community desires.

7. References

A. F. M. A. Al-Juboori, Y. Na and F. Ko, "University ranking and evaluation: Trend and existing approaches," *The 2nd International Conference on Next Generation Information Technology*, Gyeongju, 2011, pp. 137-142.

Hirsch, J. E. (2005), An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences* Nov 2005, 102 (46) 16569-16572.

Isidro F. Aguillo, José Luís Ortega & Mario Fernández (2008) Webometric Ranking of World Universities: Introduction, Methodology, and Future Developments, Higher Education in Europe, 33:2-3, 233-244.

Jin, B., Liang, L., Rousseau, R. *et al.* The *R*- and *AR*-indices: Complementing the *h*-index. *CHINESE SCI BULL* 52**,** 855–863 (2007).

McPherson, Michael A. (2012), Ranking U.S. Economics Programs by Faculty and Graduate Publications: An Update Using 1994–2009 Data. *Southern Economic Journal*: July 2012, Vol. 79, No. 1, 71– 89.

Morse, Robert. "How U.S. News Calculated the 2021 Best Graduate Schools Rankings." *U.S. News*, https://www.usnews.com/education/best-graduate-schools/articles/how-us-news-calculated-the-rankings.

"NTU Ranking – Indicators." *NTU Ranking*, http://nturanking.lis.ntu.edu.tw/methodoloyg/indicators.

"QS World University Rankings – Methodology." *QS World University Rankings*, https://www.topuniversities.com/qs-world-university-rankings/methodology.

8. Appendix

A. NLP Rankings: Top 50 Universities in the United States (2010 – 2019)

Rank	Institution	# of Authors	Score
1	Carnegie Mellon University	504	1223.18
2	University of Washington	226	581.59
3	Stanford University	248	577.31
4	Johns Hopkins University	143	479.77
5	Columbia University	160	407.76
6	Massachusetts Institute of Technology	193	350.14
7	University of Illinois at Urbana-Champaign	163	337.72
8	University of California, Berkeley	102	314.76
9	University of Pennsylvania	140	311.49
10	University of Maryland	132	310.90
11	Cornell University	102	269.52
12	Information Sciences Institute	70	234.86
13	University of Texas Dallas	50	221.22
14	New York University	113	210.85
15	University of Texas at Austin	86	209.71
16	University of Massachusetts Amherst	100	199.32
17	University of Michigan	96	188.27
18	Ohio State University	87	180.00
19	University of Colorado Boulder	78	157.54
20	University of Pittsburgh	74	129.60
21	Harvard University	54	114.43
22	University of Southern California	91	106.38
23	University of North Carolina at Chapel Hill	34	106.03
24	University of California, Santa Barbara	37	103.51
25	Georgia Institute of Technology	53	100.58
26	University of North Texas	41	93.31
27	Toyota Technological Institute at Chicago	20	92.98
28	Brandeis University	50	92.72
29	Stony Brook University	66	92.23
30	Rensselaer Polytechnic Institute	43	90.04
31	University of Illinois at Chicago	47	88.55
32	University of Utah	57	87.49
33	University of California, San Diego	56	81.35
34	University of Rochester	58	80.23
35	University of Arizona	56	74.13
36	University of California, Santa Cruz	45	66.05
37	University of California, Los Angeles	51	64.19
38	Indiana University Bloomington	54	63.66
39	Brown University	17	60.55
40	University of Wisconsin, Madison	33	56.74
41	George Washington University	24	55.86
42	Yale University	49	54.61
43	Northwestern University	40	52.48
44	The City University of New York	33	52.48
45	Georgetown University	39	51.28
46	Temple University	18	50.23
47	University of Notre Dame	27	49.76
48	University of Massachusetts Lowell	19	48.49
49	Penn State University	35	48.36
50	Michigan State University	17	46.88

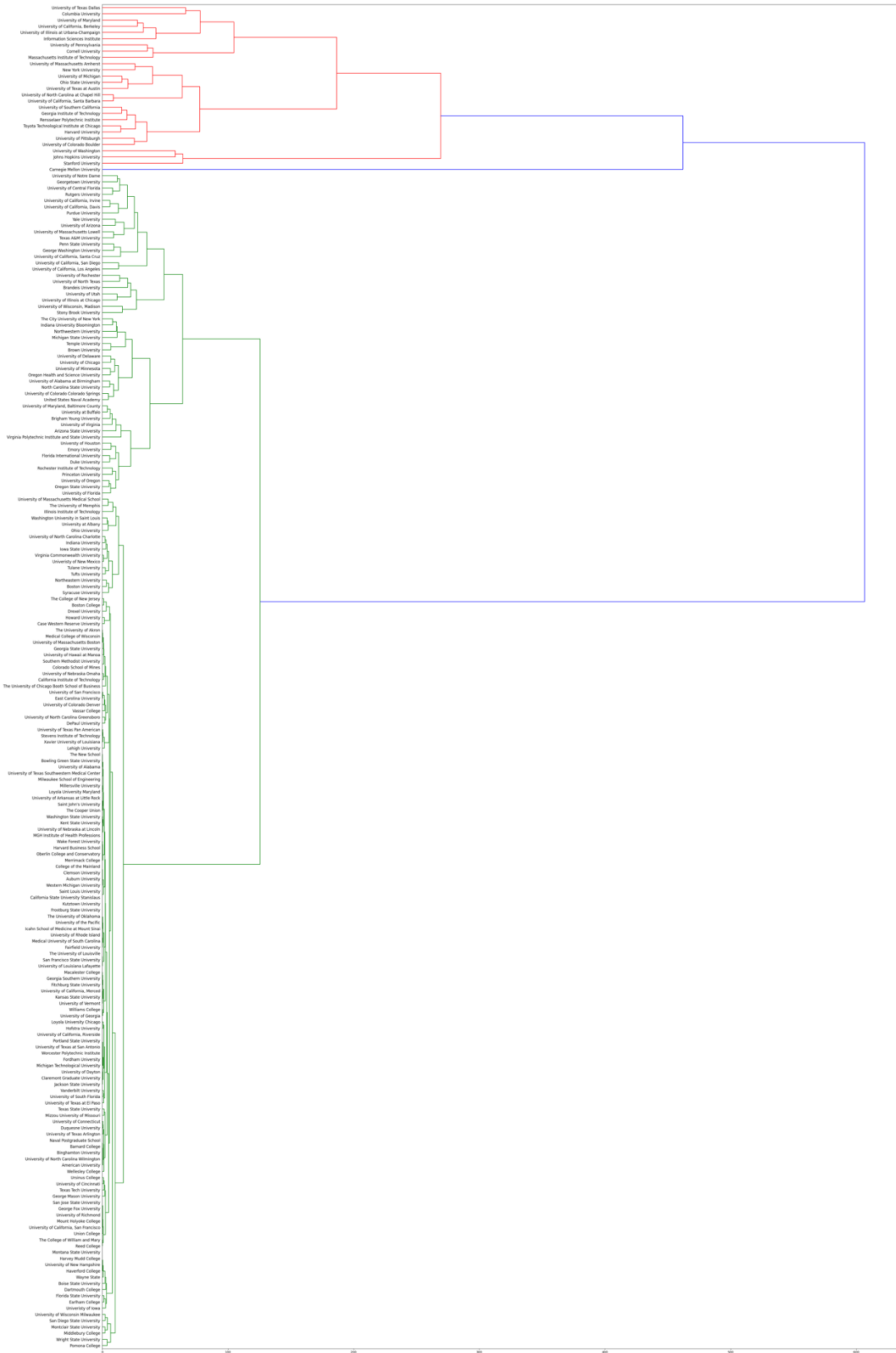
B. NLP Rankings: Top 50 Universities in the United States Rank Change (2010 – 2019)

		Rank in Year									
Rank	Institution	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
1	Carnegie Mellon University	1	1	1	1	1	1	1	1	1	1
2	University of Washington	5	6	10	4	7	3	3	3	2	2
3	Stanford University	7	8	5	3	2	2	2	4	3	4
4	Johns Hopkins University	10	5	6	6	6	4	5	2	4	3
5	Columbia University	4	2	2	2	3	5	6	15	23	12
6	Massachusetts Institute of Technology	9	11	13	10	10	6	6	6	7	5
7	University of Illinois at Urbana-Champaign	2	10	9	7	12	10	4	7	12	8
8	University of California, Berkeley	3	9	8	8	9	9	8	5	19	21
9	University of Pennsylvania	11	13	3	19	13	13	13	10	5	6
10	University of Maryland	8	7	4	12	5	8	10	9	14	13
11	Cornell University	14	22	14	11	4	11	21	8	6	10
12	Information Sciences Institute	6	4	12	15	15	15	9	19	28	24
13	University of Texas Dallas	13	3	7	5	8	7	18	49	44	64
14	New York University	18	15	29	32	15	12	11	11	8	11
15	University of Texas at Austin	12	16	15	9	17	26	14	17	13	17
16	University of Massachusetts Amherst	21	12	11	24	24	22	26	25	9	7
17	University of Michigan	15	14	17	14	28	21	16	12	17	16
18	Ohio State University	16	24	16	23	19	14	19	21	15	14
19	University of Colorado Boulder	35	17	22	31	20	16	12	13	16	23
20	University of Pittsburgh	27	23	25	27	11	19	15	28	29	59
21	Harvard University	44	146	32	28	38	30	17	26	24	19
22	University of Southern California	31	19	41	25	26	25	30	31	22	26
23	University of North Carolina at Chapel Hill	150	146	60	47	47	157	67	23	10	15
24	University of California, Santa Barbara	150	67	154	155	153	66	47	18	11	9
25	Georgia Institute of Technology	150	146	154	17	21	18	31	27	18	29
26	University of North Texas	32	29	28	41	22	44	20	16	32	46

(con't)

		Rank in Year									
Rank	Institution	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
27	Toyota Technological Institute at Chicago	77	146	43	76	29	24	24	14	33	20
28	Brandeis University	38	25	20	26	14	23	34	30	56	54
29	Stony Brook University	150	20	19	13	23	48	43	37	34	34
30	Rensselaer Polytechnic Institute	150	146	154	155	18	17	23	24	25	33
31	University of Illinois at Chicago	18	27	21	22	46	39	33	42	49	30
32	University of Utah	21	18	36	20	42	29	27	51	60	36
33	University of California, San Diego	28	30	46	30	43	91	61	34	43	18
34	University of Rochester	24	21	33	44	33	47	39	32	26	41
35	University of Arizona	64	37	50	56	48	35	49	33	21	25
36	University of California, Santa Cruz	64	65	38	39	57	20	35	20	42	49
37	University of California, Los Angeles	53	37	27	38	153	157	50	76	36	22
38	Indiana University Bloomington	26	26	23	33	36	55	41	39	59	69
39	Brown University	24	37	26	21	27	31	61	83	77	65
40	University of Wisconsin, Madison	64	32	18	16	56	72	104	161	38	62
41	George Washington University	150	146	154	29	31	32	25	48	50	44
42	Yale University	53	58	43	86	86	157	69	36	20	31
43	Northwestern University	21	35	154	36	30	41	91	46	58	52
44	The City University of New York	17	28	30	34	59	61	169	44	61	61
45	Georgetown University	64	146	24	55	153	36	37	53	37	40
46	Temple University	36	40	35	18	32	33	169	83	66	75
47	University of Notre Dame	150	146	154	155	63	28	66	40	30	35
48	University of Massachusetts Lowell	150	58	65	78	79	40	57	22	31	50
49	Penn State University	57	146	65	51	37	26	44	54	40	53
50	Michigan State University	21	49	38	42	45	157	29	59	51	75

C. Hierarchical Clustering Dendrogram based on Scoring Trends



D. Top 30 NLP Authors h-index and wc-index Comparison (since 2015)

			wc-index				
Rank	Name	h-index*	2015	2016	2017	2018	2019
1	Dan Roth	50	5	13	17	25	31
2	Noah A. Smith	53	7	8	11	17	23
3	Dan Klein	47	6	9	18	23	26
4	Christopher D. Manning	90	6	12	15	18	20
5	Eduard Hovy	54	2	5	6	7	12
6	Mohit Bansal	30	1	4	7	20	30
7	Vincent Ng	30	5	8	10	11	11
8	Luke Zettlemoyer	45	4	7	7	10	16
9	Claire Cardie	43	2	3	7	13	16
10	Garham Neubig	32	0	2	5	13	18
11	Chris Dyer	53	5	5	5	5	5
12	Heng Ji	38	0	2	3	3	5
13	Kevin Knight	42	3	7	9	11	11
14	William Yang Wang	24	3	3	6	13	19
15	Jason Eisner	32	4	9	12	16	19
16	Regina Barzilay		5	8	10	12	14
17	Mona Diab	33	0	1	1	4	4
18	Dan Jurafsky	63	3	5	7	7	7
19	Nizar Habash	33	2	3	4	7	9
20	Jordan Boyd-Graber	33	2	5	8	10	14
21	Kathleen McKeown	33	4	6	8	9	14
22	Mark Dredze	49	8	11	11	13	14
23	Percy Liang	45	4	10	12	15	18
24	Rada Mihalcea		2	4	7	8	10
25	Yejin Choi	38	0	2	4	5	5
26	Kevin Gimpel	27	1	1	4	6	9
27	Jacob Eisenstein	30	6	10	12	15	15
28	Tom Mitchell	54	4	6	9	11	13
29	Yang Liu	26	4	5	5	5	5
30	Bing Liu	69	2	3	5	5	5

** h-index data (since 2015) is collected from Google Scholar on at the end of March; some profiles are missing*
** the h-index is not limited to publications on Natural Language Processing and publications dedicated to academic institutions in the U.S., which may result in more publications than those collected for NLP Rankings*