

Grant Form Similarity Analysis

for the Technology Association of Grantmakers

November 10, 2021

Kwame Porter Robinson (kwame@kwamata.com)
Technology Association of Grantmakers (info@tagtech.org)

Executive Summary

The data science work of the grant similarity effort by the Technology Association of Grantmakers (TAG) achieved its technical goals through a practical summary of over 3,500 form fields across 20 multi-part summarizing questions. Far beyond the initial 38% similarity discovered the 20 summarizing questions provide evidence of nearly total form similarity because of the many to one relationships identified between sets of grant form questions and to summarizing questions. In this final report we review the data science methods utilized across four major tasks after justifying them through a set of principles and a process. Also, miscellaneous notes regarding grant form question trends and implications are provided as well.

Process Overview

Our process was divided into two conceptual stages and one methodological stage that implemented the prior conceptual stages. The problem of identifying similarity was first understood across a set of three principles related to similarity: 1) that similarity could be measured within pairs of form questions, or pairwise similarity, 2) that similarity could be clustered, and 3) that similarity could be summarized. The three principles together suggest a sequential process that constructs a many to one relationship. The process is carried out by selecting methods for each principle. Figure 1 illustrates the data science relationship between principles, processes, methods, and the resulting findings. Task 1, 2, 3 and 4 are detailed within the Method Details section. Their justification largely follows from the principles and processes but other relevant details are also mentioned in the Method Details section.

Principles → Processes → Methods → Results

Figure 1: Data science relationship between principles, processes, and methods. The resulting findings are an output.

Principle Details

The problem of finding similarity across grant form questions is an investigation of semantic similarity within a semi-structured socio-semantic organization, represented as a grant form, of varying levels. For example, many form questions have multiple parts that all ask the reader to provide written evidence (semantic) of social impact in restricted forms chosen by the foundation that are also aligned with their goals. This represents a complex socio-semantic similarity phenomena.

To simplify our investigation, we restricted the measurements and operations that we could undertake. We chose similarity principles widely assumed in other literature investigating semantic similarity phenomena. The principles were:

P1: Similarity can be pairwise

- Two questions can be similar or not similar to each other

P2: Similarity can be clustered

- Questions can be clustered with one or more similar questions.
- Those clusters of questions can include one or more other clusters.

P3: Similarity can be summarized

- Many questions can be summarized into a single question that is (very) similar.

A conceptual relationship illustrating inputs and outputs among these principles is given in Figure 2.

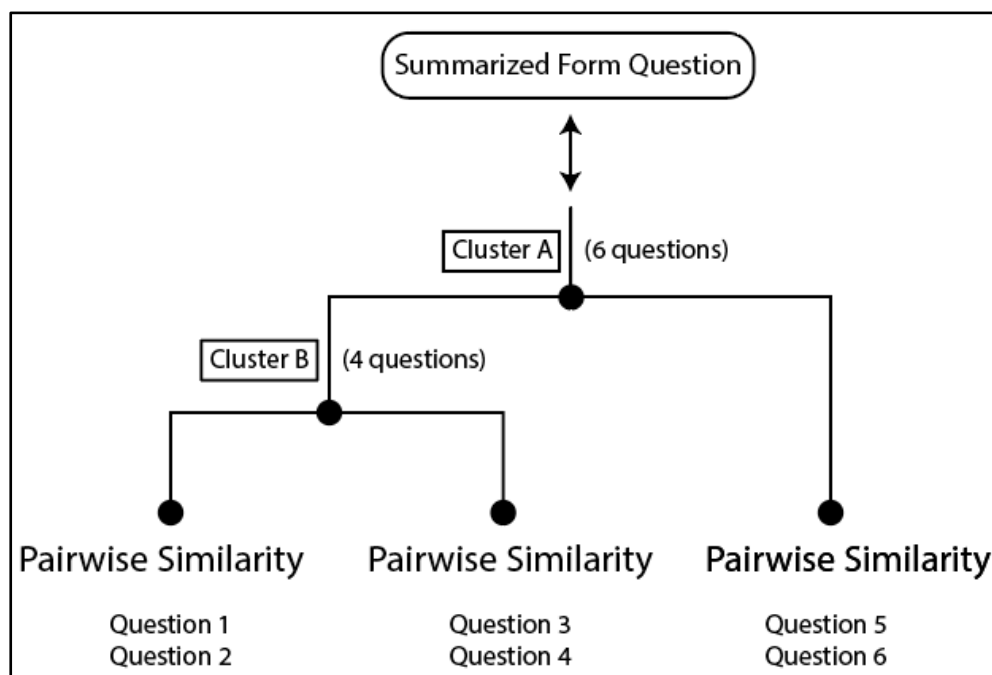


Figure 2: Conceptual relationships among principles. Reading from bottom up, pairwise similar questions are clustered. Clusters are clustered. Larger clusters can be summarized.

Process Details

Because the principle inputs and outputs overlap it should be possible to convert many questions into one (or a few) using the conceptual relationship highlighted in Figure 2. A basic illustration that sequentially lays out the principles within a process is provided in Figure 3. The result is a many to one relationship between form questions and a summarized question.

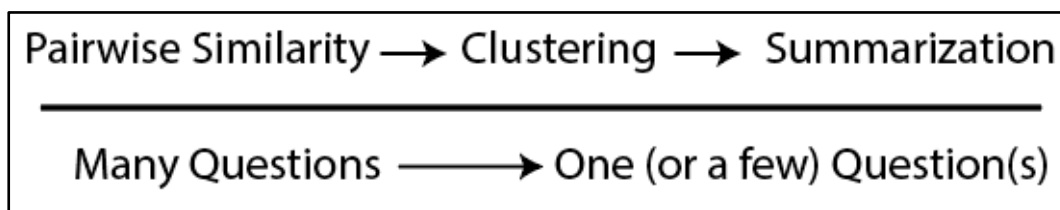


Figure 3: Process derived from similarity principles

We can implement this process by selecting methods that implement P1, P2, and P3. In Method Details we discuss our choice of Google’s University Sentence Encoder for P1, Dendrogram clustering for P2, and data triangulation for P3. Task 2 and 3 implement P1 and P2. Task 4 implements P3. Task 1 is a data sanitization phase.

Method Details

Removing Sensitive Information: Task 1

Data sanitization and de-identification is not a principle but was required to protect participant anonymity. Data was sanitized by removing geographic identifiers and personally identifiable information. Additional and specific references to foundations, such as foundation named grant making IT systems or grant types, were also sanitized. Additionally, to improve question comprehension as deemed appropriate the non-question content prior and after the field was included, or *the context*, through a semi-manual process. Some fields were consolidated in rare cases where the shared field topic was extremely similar. The context will greatly support the summarizing and field reduction tasks. Finally, our deeper look at the context raised several concerns of how foundation forms are actually used in practice. A process overview and list of concerns are provided. 210 pieces of sensitive data were removed.

Sanitization Process Overview

1. Using [Prodigy](#), sensitive content was identified and labeled. This resulted in a set of sanitized annotations.

2. The sanitized annotations underwent a deeper, second level of review by Kwamata LLC. Adjustments were made where necessary. Additional context was collected where deemed appropriate.
3. Light whitespace removal was carried out in the question context. The fields were typically left as is due to variation in field text structure.
4. A final column separated value (csv) file was generated.

Miscellaneous: Concerns raised regarding how foundations use forms

- Questions often ask for direct CEO information, possibly devaluing CEO time through repeated direct contact
- Questions often ask for opinions on areas outside of expertise, training or cause conflicts of interest:
 - Risk assessments for financial or reputational challenges.
- Questions sometimes ask for foundation-specific portfolio information that is already available within foundation IT systems but not easily available to nonprofits IT systems:
 - Prior grant id number (e.g. investment id, award #)
 - Date of prior applications
 - Amount of funds requested in prior application(s)
 - Contact information for references already submitted within grant portal
- Several applications include a form completion time estimate despite continued use of lengthy forms
 - This suggests that foundations already appear to care about form length but do not act on concerns related to form length
- Invisible labor caused by sequences of questions where the last question largely overlaps the first in intent or even overall nuance
 - For example:
 - Q1: WHAT'S YOUR IDEA AND WHAT PROBLEM DOES IT SOLVE?
 - Q2: IF YOUR IDEA WORKS, HOW WOULD IT BE TRANSFORMATIVE?
 - The answer to Q2, something working and being transformative, would include the problem solved and what the idea is; this largely includes Q1
 - For example:
 - Q1: Foundation Funding Amount (USD)
 - Q2: Has the amount changed since the proposal or last report?
 - But asking Q2 would answer Q1 if the amount was also given
- Questions asking for percentages creates extra demographic work because software or foundation staff can calculate percentages from count data
 - Furthermore, combining percentage data can be problematic, for example averaging percentages is methodological unsound because you know nothing about the sample size that determines the final percentage. [See this for an easy to read explanation](#). We suggest asking for counts only.
 - e.g. Both percentages are a half but when you consider sample size you get different demographics across nonprofits
 - Case 1: $(1/2 + 1/2) / 2 = 1/2$

- Case 2: $(1/2 + 5/10) / 2 = 1/4$ ← sample size matters, but foundations overly ask for percentages.
- Multiple deal breaker provisions are asked after significant amounts of information already provided
 - For example, if an institutional research board (IRB) reference is required but the question is encountered after the nonprofit has filled out half of the form then they have wasted that time.
 - Suggest that forms put their deal-breakers up front, on page 1.

Pairwise Similarity, and Clustering Pairwise Similarity: Task 2 and 3

Task 2 and Task 3 together use pairwise similarity and dendrogram clustering to address P1 and P2. For pairwise similarity, the distributional hypothesis [1, 2], that states that words or language occurring together are also inherently similar, allows us to work with form questions. For dendrogram clustering, corpus-based similarity methods [3] justify the use of clustering to assess grant form similarity. Because corpus-based similarity methods incorporate global as well as local semantic information they provide greater nuance in similarity whereas direct comparison methods do not.

To calculate pairwise similarity, Google's Universal Sentence Encoder [4] was used to provide a neural network approximation of classical cosine text similarity in sentences. For clustering, the dendrogram ward algorithm with optimal ordering was used [5]. This method of clustering is simpler to explain and uses fewer computational resources than T-SNE and other highly non-linear methods.

Some form questions are not similar to others. Presuming that they should be assessed differently, we separate the set of similar form questions from the set of non-similar form questions. Non-similarity was retro-actively assessed from dendrogram characteristics at a given threshold such as a form question having no other associated cluster, indicating that it was independent of other questions. High (0.89) and low (0.11) thresholds were used. This allowed us to create a list of fields and clusters at low thresholds for similar form questions and at high threshold for non-similar form questions. At high thresholds the increased clustering of non-similar was helpful for the final qualitative summarization in Task 4.

NOTE: Conducting similarity analysis on fields that have been sanitized produced slightly different results than the unsanitized fields used in [earlier analysis from August 2021](#) because the content of those fields was slightly different with identifiers removed. But the fundamental similarity relationships remained the same. The sanitized analysis results in ~38% similarity at a low threshold with clusters becoming more unintelligible past that cutoff, just as in prior work from August 2021.

Summarizing Similarity: Task 4

The question set was reduced to 20 multi-part questions through a multi-method and multi-analyst data triangulation approach that started with roughly 3,500 form fields organized across

several thousand question groups in a dendrogram. Question groups were divided between (a) those having many similar questions and (b) those having very few or no similar questions, i.e. non-similar. Data triangulation started from initial dendrogram organization using high and low thresholds and was manually performed through concurrent top-down and bottom-up content summarization processes. The manual summarization processes elucidated both structural question themes and additive specifics that together largely reflected many question groups despite only containing several paragraphs or several bullet points. The process resulted in 163 parts across 20 multi-part question groups, reduced significantly from the original 3,577 form fields in Task 1.

At this point in the process, we were confronted with the choice of using quantitative or qualitative methods for summarization. Examples of quantitative methods include PEGASUS [8] or BART Summarization [9]. Large-scale summarization natural language processing methods have advantages and disadvantages. They are fast to run, requiring minimal input and are routinely used by large data science companies. However, these methods can “hallucinate” text [9; pg 8], produce demonstrably racist results [10], as well as require manual reliability assessments. Alternatively, qualitative methods include thematic analysis or qualitative content analysis [11]. These methods are foundational for multiple academic research disciplines and allow the user to cognitively reflect on their choices to account for biases (an example of this is recognizing positionality). However, these methods are slow, not strictly replicable, although external validity can be assessed by inter-rater reliability.

Given that many form questions deal with questions of race, gender and sexual identity we did not want to introduce racist or gendered biases from quantitative summarization algorithms that would require individually rechecking over 3,500 form questions. Furthermore, either method requires some kind of manual review and the method is made more efficient if the review is also part of the summarization work itself. Therefore we chose data triangulation [12, 13] as our qualitative summarization method.

Data triangulation is a method of developing a qualitative sense of data through multiple methods of data measurement, where the recorded emergent meaning is the output. We chose a multi-method data triangulation involving two reviewers and two methods. The author and a trusted contractor were the reviewers. Top-down summarization, to elicit and understand themes, was the first method. Bottom-up summarization, to elicit bottom up summarization of specifics was the second method. External validity was justified by assessing inter-rater reliability, which was extremely high. The methodological iteration of data triangulation for TAG's grant question similarity analysis is illustrated in Figure 3.

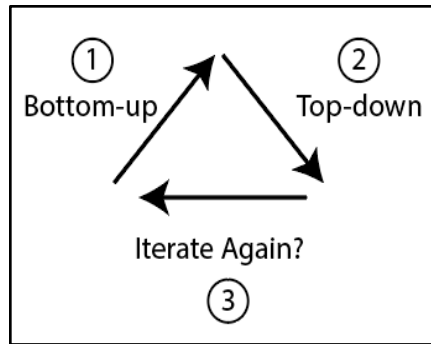


Figure 4: Data triangulation in TAG Grant Similarity Analysis

Findings

Overview

13 high level question themes emerged from data triangulation. These themes were largely and roughly equally distributed across two dimensions. The relationship of information to the nonprofit, as inside of it or outside of it, is the first dimension. What the information largely regards, as who or, instead, what (and often how), is the second dimension. Table 1 presents a tabulation of the themes according to these dimensions.

<i>Rough* theme tabulation</i>		Relationship of Information to Nonprofit		
		Inside of	Outside of	
Information largely regarding	Who	1, 3	4, 5, 7, 10	6 themes
	What (and How)	2, 8, 9, 13	6, 11, 12	7 themes
		6 themes	7 themes	13 themes

Table 1: Rough tabulation of 13 themes defined by both information and its inside or outside relationship to the nonprofit as well as what it largely regards.

Figure 6 reflects theme counts and prevalence as percentages across form questions. The top three themes by size were: What the Organization Does (796; 22%), Requested Grant Funding related (715; 20%), and Organizational Biographical and General Information (642; 18%).

Question Group "Theme"	Counts (Percentage)
Organizational Biographical and General Information	641 (18%)
Miscellaneous	119 (3%)
Corporate Delegation and Oversight, Organizational Structure	186 (5%)
Data Handling, Overview, Measurement, Evaluation and Reporting	129 (4%)
Project Demographics/Orientation/Status	84 (2%)
Alternative Supports	5 (<1%)
How Did You Hear of Us	13 (<1%)
How has COVID-19 Impacted Your Work	8 (<1%)
Organizational Budgeting, Revenue Practices, and Forecasts	699 (20%)
Collaborative Partnerships and Community Support	170 (5%)
Requested Grant Funding Related	719 (20%)
Time Spent Filling Out the Form	5 (<1%)
What the Organization Does	802 (22%)

Figure 6: Theme representation in counts and percentages across the form questions.

Shared Questions Within Each Theme

The specific themes and their derived summarizing questions are presented in Table 3. For data scientists, a .csv file of the sanitized form questions and their theme associations is provided at <https://github.com/TAG-repo/grantform-similarity/releases/tag/v1.0.0-alpha>.

Theme and Summarizing Questions

Theme	Summarizing Question(s)
Organizational Biographical and General Information	<p>Provide the following information about the organization and staff. Where contact information is requested, provide (if applicable) Name, Title, Address, Email Address, Phone Number.</p> <ul style="list-style-type: none"> ● Professional References/Testimonials (others who can speak on the organization's behalf) ● Qualitative Staff Characteristics (project-related experience, shared backgrounds, executive biographies, etc.) ● Applicant Contact Information ● Contact Person Contact Information ● Intern Roles (if any) ● Key Contacts Contact Information ● Key Program Staff Resumes/CVs ● Organization's Contact Information ● Organization's Donation Website ● Organization's DUNS Number and/or GuideStar profile ● Organization's Fax Number ● Organization's Founding Date/History ● Organization's Legal Status ● Organization Social Media (Facebook page, Instagram page, Twitter page, etc.) ● Primary Applicant Contact Information ● Staff Contact Information ● Staff Demographics (Age, Ethnicity, Gender, Immigration Status) ● Staff Qualifications ● Total Number of Part-time Workers ● Total Number of Full-Time Workers ● Total Number of Paid Staff ● Total Number of Staff Hours
Miscellaneous	<ul style="list-style-type: none"> ● To submit this grant please provide your signature, the date of signature, and indicate your title and position. Also confirm that you understand and acknowledge eligibility criteria and the project's satisfaction of these criteria. Your agreement includes permission to share information with other donor advised funds as well as acceptance of terms and conditions. ● If you are applying to any one (or more) of our supporting organizations please list them and provide an explanation of your involvement. For example, some organizations offer thought partnership while others might offer in-kind expert staff support.
Corporate Delegation and Oversight, Organizational structure	<p>Provide the following information:</p> <ul style="list-style-type: none"> ● Board/CEO Contact Information ● Board/CEO Demographics (Age, Ethnicity, Gender, Sexual orientation, Immigration Status) ● Board Compensation/explanation thereof ● Does the board contribute financially to the organization? If so, generally how much?

Theme	Summarizing Question(s)
	<ul style="list-style-type: none"> ● Qualitative Board/CEO Characteristics that are relevant and important for us to know, e.g. professional affiliations, national origin, and/or length of time in role ● Total Number of Board of Directors ● Total Number of Executive Staff ● Strategies for ensuring continual improvement within the organization/project staff
<p>Data Handling, Overview, Measurement, Evaluation and Reporting</p>	<p>Share the following information:</p> <ul style="list-style-type: none"> ● Briefly describe what kind of data will be collected or generated (if relevant). What outcomes, deliverables and evaluations are important to this work? ● Data to be collected in order to monitor progress/impact on served populations ● Planned policies/procedures/practices to ensure data security and regular data monitoring ● Whether or not collected data can be shared ● School-related data (including persistence/enrollment/completion rates for K-12, Bachelors, Associates, and/or Short-Term Credentials) ● Whom the desired data will be collected from and how often ● Methods for calculating volunteer hours ● Methods for calculating projected populations served ● Methods and theory applicable for data measurement ● If this work requires institutional review board (IRB) approval (e.g. research and experiments involving human beings) then please indicate whether you have obtained approval and explain why.
<p>Project Demographics / Orientation / Status</p>	<p>Provide the following information about populations served by the project:</p> <ul style="list-style-type: none"> ● Annual counts of population served (including number of people served) ● Specific race and ethnic populations served ● Served populations by geography/location ● Served populations by income/socioeconomic status ● Served populations by work/occupational status ● Served populations by gender identification ● Served populations by sexual orientation ● Served populations by disability status ● Served populations by veteran status ● Specific age groups served ● Total population counts (including number of people served)
<p>Alternative Supports</p>	<ul style="list-style-type: none"> ● Aside from financial assistance, what other resources, partnerships, or skill sets can be offered to support this work?
<p>How Did You Hear of Us</p>	<ul style="list-style-type: none"> ● Explain how you heard about the Foundation and why you have chosen to apply for funding here. Were you referred by another non-profit organization?
<p>How has COVID-19 Impacted Your Work</p>	<ul style="list-style-type: none"> ● Provide information on the ways in which COVID-19 has had an impact on your organization, your current projects, and

Theme	Summarizing Question(s)
	<p>anticipated future work.</p>
<p>Organizational Budgeting, Revenue Practices, and Forecasts</p>	<p>Provide information on the following:</p> <ul style="list-style-type: none"> ● Chief Finance Officer (CFO) and finance controller certifications, degrees, certifications, and/or any professional affiliations. ● The organization's IRS history/records, including 501c3 determination/EIN, 990, if applicable ● The organization's GuideStar Exchange participant level ● Any instances of fraud, malfeasance, or financial mismanagement in the past five years (including details on how they were managed and resolved, if applicable) ● Available non-monetary resources (e.g. in-kind support) ● Financial health (rate the condition of your finances, including strengths and challenges) ● Financial management structure (including information about both staff and processes that make this organization a good steward of resources) ● Financial expenses during the last fiscal year ● First and second most recently completed internal audits ● Fiscal sponsor details, including fees, contact information, and letter ● Funding allocation methods (e.g. overhead, staffing, or to project) ● Funding from external sources (e.g., federal, state, or city funding) ● Income and balance sheet ● Liability amount and insurance ● Number of trustees ● Recent major changes to financial circumstances (if any) ● Records of the organization's most recent budget and year-to-date budget ● Revenue ● Start and end dates for the organization's fiscal year ● Statement of revenue and expense for the most recently completed fiscal year ● The remaining number of months worth of cash reserves still on hand ● The organization's current operating budget and expected annual operating budget for the next fiscal year ● The organization's long-term fundraising plan/plans for project sustainability ● Whether or not the organization is a registered Living Wage Employer ● Any other pertinent product budget details
<p>Collaborative Partnerships and Community Support</p>	<p>Share the following information about served communities and similar organizations:</p> <ul style="list-style-type: none"> ● Beneficiaries of the project/services, also noting who are the primary beneficiaries ● Community involvement in project planning/implementation and/or decision-making ● Current collaborating organizations, contributions from these organizations, and which organizations are considered the most

Theme	Summarizing Question(s)
	<p>essential. Also indicate their roles in supporting your work.</p> <ul style="list-style-type: none"> ● Planned/hoped for future collaborations ● Donor base details ● Other organizations involved in the same or similar work ● Prior collaborating organizations and their supporting roles. ● Served population experiences/issues and how they inform community outreach ● Specific community needs addressed by the organization, the root causes of these needs, and how they were identified ● The organization's engagement with volunteers ● Total number of volunteer hours
Requested Grant Funding Related	<p>Provide the following information relating to your application:</p> <ul style="list-style-type: none"> ● Briefly and then in more detail describe why the grant is needed. For example, how did the project come to be and how, if funded, how might the work progress as it continues? What communities are involved and what opportunities exist? ● What of the foundation focus areas would this grant proposal be the best first for and why? ● Amount/funding tier applied for ● Type of grant requested ● Anticipated overall impact of the grant regarding your organization capabilities, if approved ● Requested grant start/end date; number of years applied for ● Descriptions of the types of things that the grant will be spent on/an itemized budget of planned grant spending ● Identify any grant funds intended for lobbying activities; if funds are identified then detail lobbying activities. ● Foundation priority groups that are the best match for the request ● Planned alternatives if the grant application is not approved ● Prior contact/interaction with the Foundation/specific Foundation staff; specify the expiration date if this application is related to replacing or renewing an existing grant ● Project-generated revenue (if applicable) ● Purpose of the application: project-specific or for general operating funding ● Resources dedicated to project support ● Sufficiency or insufficiency of the grant alone to support the project ● Any major changes in the organization over the past three years ● Things you hope to learn from this initiative if you receive funding ● If applicable, provide a signed cover letter from the CEO/Executive Director regarding this grant request. ● Project impacts/achievements within the past year ● Two main successes achieved for the project overall ● Strategies for meeting current and future organization/project goals ● Standards for gauging project success ● Any IP rights to be filed/generated for the project ● The timeline for the project

Theme	Summarizing Question(s)
Time Spent Filling Out the Form	<ul style="list-style-type: none"> Approximately how many hours did it take to complete this grant application?
What the Organization Does	<p>Along with any optional video uploads, share the following information about the organization and the project:</p> <ul style="list-style-type: none"> Areas in which the organization conducts policy and advocacy efforts (if applicable) How the organization’s work is planned to address both internal inequalities within the organization and external disparities within the served community The mission of the organization Project classification (e.g., arts & culture, community development, education, health, housing and shelter, public safety, etc.) The relationship of the project to the organization’s mission The organization’s geographic location The organization’s current programs The organization’s principles Racial equity approaches and economic impact guidelines followed for the project As an organization what populations and areas are a main focus? Do the organization policies align with Foundation nondiscrimination requirements? How work done thus far demonstrates a commitment to racial equity Identified project impact areas Identified project challenges/risks (social, legal, financial, etc.) and mitigation efforts Anticipated scope of the project (including the size of the population served) Project impacts/achievements within the past year Two main successes achieved for the project overall Strategies for meeting current and future project goals Standards for gauging project success Strategies for ensuring continual improvement within the organization/project staff Any IP rights to be filed/generated for the project The timeline for the project

Table 3

Dissimilar Questions Within Each Theme

Additionally, within each question group, there were fields that do NOT share a high degree of similarity but were present in the dataset.

Question Group Theme	Summarizing Question(s) for <i>Dissimilar</i> Fields in This Theme Group
Organizational Biographical and General Information	<ul style="list-style-type: none"> ● Please describe any relationships between your executive staff as well as board members and the communities you serve. Any roles and relationships to your projects or initiatives and how staff and community members typically work together are especially helpful for us to know.
Miscellaneous	<ul style="list-style-type: none"> ● Please provide an accompanying narrative or attachments for additional explanation if warranted. Is there anything else you would like to tell us about your organization, and work that you were unable to do so elsewhere? ● Do you have any recommendations for improving the application process? ● Would you like to be added to our mailing list or otherwise be notified of future funding opportunities or changes to the foundation's work?
Corporate Delegation and Oversight, Organizational structure	<ul style="list-style-type: none"> ● If they exist, please attach organizational bylaws. Also, please provide any board nomination and board voting procedures.
Project Demographics/Orientation/Status	<ul style="list-style-type: none"> ● Please indicate immigrant and family generation status of the communities that you serve (e.g. first or second generation immigrant).
Organizational Budgeting, Revenue Practices, and Forecasts	<ul style="list-style-type: none"> ● Disaggregate staff expenses to include fringe budgets, employer FICA, health, dental insurance, short and long-term disability, 401(k) plans, any offered matching, costs of unemployment insurance, and workers compensation. ● If you receive funding from indirect fund raising (e.g. United Way), government contracts, other foundations, individuals and/or transferred individual income please list amounts, sources and provide auxiliary documentation.
Requested Grant Funding Related	<ul style="list-style-type: none"> ● Although you may have described critical partnerships elsewhere please highlight and detail which partnerships are important to the proposed grant. ● If the work can be structured by milestones please provide a brief timeline of expected milestones for outcomes and impacts. How will you communicate impact outside of your immediate partners and community? ● What approaches inform and are incorporated in your proposed grant? Examples might include racial equity approaches, or asset based design.

Question Group Theme	Summarizing Question(s) for <i>Dissimilar</i> Fields in This Theme Group
	<ul style="list-style-type: none"> • Is in-kind support greater than 10% of the requested funding?
What the Organization Does	<ul style="list-style-type: none"> • Please note your organization's current GuideStar Seal of Transparency. For information on the benefits of updating your profile, visit GuideStar. • Has the organization experienced a recent merger and/or significant change in executive level staffing? If so, provide details and context for the changes. Are there any anticipated significant organization changes planned for the future? • Do you have a Theory of Change? If so, please attach any documents to help us understand the theory.

References

- [1] Firth, J.R. (1957). A synopsis of linguistic theory 1930-1955. In *Studies in Linguistic Analysis*, pp. 1-32. Oxford: Philological Society. Reprinted in F.R. Palmer (ed.), [Selected Papers of J.R. Firth 1952-1959](#), London: Longman (1968).
- [2] James Gorman and James R. Curran. 2006. Scaling distributional similarity to large corpora. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Meeting of the Association for Computational Linguistics*. 361–368.
- [3] Dhivya Chandrasekaran and Vijay Mago. 2021. Evolution of Semantic Similarity—A Survey. *ACM Comput. Surv.* 54, 2, Article 41 (February 2021), 37 pages. <https://doi.org/10.1145/3440755>
- [4] Cer et al (2018) “Universal Sentence Encoder” [arXiv:1803.11175v2](#) [cs.CL]
- [5] Dhivya Chandrasekaran and Vijay Mago. (2021) Evolution of Semantic Similarity—A Survey. *ACM Comput. Surv.* 54 <https://doi.org/10.1145/3440755>
- [6] Zhang et al (2020) “PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization” ICML 2020
- [7] Lewis et al (2019) “BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension” [arXiv:1910.13461v1](#) [cs.CL]
- [8] Bender et al (2021) “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜” FAccT ‘21 <https://doi.org/10.1145/3442188.3445922>
- [9] Mello (2021) “Qualitative Comparative Analysis: An Introduction to Research Design and Application” Georgetown University Press
- [10] Nightingale (2020) Triangulation. In: *International Encyclopedia of Human Geography: 2ed*
- [11] Fusch (2018) “[Revisiting Triangulation in Qualitative Research - ScholarWorks](#)” *Journal of Social Change*. DOI: 10.5590/JOSC.2018.10.1.02