Searching to Measure the Novelty of Collected Ideas

Rhema Linder
Interface Ecology Lab
Department of Computer Science and Engineering Texas A&M University
rhema@ecologylab.net

Andrew M. Webb
Interface Ecology Lab
Department of Computer Science and Engineering Texas A&M University
andrew@ecologylab.net

Andruid Kerne
Interface Ecology Lab
Department of Computer Science and Engineering Texas A&M University
andruid@ecologylab.net

ABSTRACT
Visual representations of ideas are valuable for creative thinking and expression. Prior research on design and information-based ideation has assessed novelty in creative products as the inverse of the frequency that an idea or visual element occurs in the complete space of responses. In controlled experiments, frequency has previously been calculated in reference to the set of ideas collected by all participants (corpus). Experimental conditions restricting the space of possible elements resulted in overlap between participant responses, yielding a range of frequencies. Alas, in field investigations the space of possible elements is unrestricted, resulting in little overlap of ideas, and thus mostly a single frequency $(1/N)$ of collected elements.

We introduce a new method that uses web search to measure the novelty of individual ideas. Instead of using the local corpus directly to calculate frequency, we use the number of results from web searches generated from elements in the corpus. Our implementation uses Googles reverse image lookup to determine the popularity of images. We compare results with those derived via prior methods.

Author Keywords
novelty, information-based ideation, evaluation metrics

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION
Visual representations of ideas are valuable for creative thinking and expression [1, 4]. Prior research on design and information-based ideation has assessed novelty in creative products [8, 5, 9]. In researching creativity support environments, researchers often label user authored media by hand to extract data from experiments and investigations in the field. As the amount of user authored media increases, researchers have more impetus to invent automated techniques for measuring components of creativity.

Information-based ideation (IBI) is the process of having new ideas while working with information [5, 11]. In *information-based ideation tasks*, a person searches, collects, organizes, and thinks about information to answer open-ended questions, such as planning a vacation, deciding on a thesis topic, or designing an innovation. In IBI tasks, people author collections of information and media elements from sources including the Web to represent ideas relevant to the task at hand. In this methodology, each collected element is an individual "answer" or idea developed in response to an IBI task, which one can use to measure components of creativity.

Novelty, one component of creativity, is the uniqueness of an idea. Researchers have computed novelty as statistical infrequency in laboratory experiments [8, 9, 11]. We refer to the prior method for calculating elemental novelty as *corpus-based novelty*. In corpus-based novelty, a collected element has high novelty if it appears rarely over the set of all collections in an experiment (the corpus). However, this metric only produces a rich range of values when the number of possible elements an author can collect is constrained by experimental conditions. When users collect media from the internet using self authored queries, the number of possible elements to choose can include any online content. An element collected in such unconstrained conditions is likely unique among participants, but may be commonly or uncommonly found online.

Search-based novelty uses the number of search results from a query generated from a media element to measure its novelty, producing a rich range of values. Computing search-based novelty is an algorithmic process, which is faster than manual solutions that require human raters. Once a set of sample of media is used to create a novelty function, it not require on a corpus of experimentally collected data to calculate novelty.

We begin with a discussion of prior work. Next, we introduce search-based novelty. We show empirical evidence to validate this metric in the context of images and Google Image search. We conclude by discussing implications for design and future work.

PRIOR WORK
Prior work has addressed measuring creativity, ideation process, and the experiential ratings of creativity for evaluating creativity support environments.

CHI 2013, April 27–May 2, 2013, Paris, France. Evaluation Methods for Creativity Support Environments Workshop.

Carroll et al. produced a range of questions for evaluating creativity support tools [2]. Self reported data provides insight about participant experiences. Our approach measures creative products, which contain attributes that we can measure, rather than a creative processes or experience.

Dow et al. used a variety of metrics to evaluate the efficacy and creativity of ad prototypes [3]. Instead of coding attributes of the creative products as a measure for novelty, they used Mechanical Turk to measure similarity for each ad in the corpus of created ads. To verify, Dow et al. also employed a panel of experts and used the click through rate of ads to measure ad value. Instead of measuring similarity as pairwise, both corpus-based and search-based metrics described in this paper use statistical infrequency to describe novelty.

Shah et al. review and posit measures and processes for evaluating the effectiveness of methods for generating ideas in the context of engineering design: the intersection of utility and novelty [8]. To assess novelty in the context of a highly controlled experiment, they prescribe aggregating ideas present in experiments into categories. The more often an idea occurs across participants, the less novel it is. Calculating novelty with this approach is easy, but requires many hours of work. Similarly, we use uncommonness to measure novelty, but our analysis is automated.

Webb and Kerne et al. designed a laboratory experiment evaluating components of creativity of participant authored information compositions [5, 11]. *Information composition* is a medium that affords collecting and organizing ideas as text and image bookmarks, helping people perform IBI tasks. Participants collected images from a preselected set of search queries. Under these conditions, nearly all of the images collected were collected by multiple participants, creating significant overlap. We report a field study where each participant collected images with little overlap.

Webb and Kerne use Equation 1 to calculate the image novelty of an individual element image e in a set of all image collections C is 1 divided by the number of sets in C where e is present:

$$enov_i(e, C) = \frac{1}{|c \in C : e \in c|} \tag{1}$$

As elements in a corpus become more common, the value of $enov_i(e, C)$ becomes smaller, indicating that the element is less novel. Also note that the novelty of an element e is dependent on the corpus of participant respondent collections C. Calculating the inverse popularity of elements among a corpus closely follows principles that make IDF useful in information retrieval.

Inverse Document Frequency (Equation 2) highlights rare terms in documents by giving high weights to terms that occur in fewer documents over a corpus [7]. The IDF of a term t depends on the number of times that term appears in documents in D.

$$idf(t, D) = \log \frac{|D|}{|d \in D : t \in d|} \tag{2}$$

Figure 1. Example images from least to most novel. Images $inov$ scores ranging from 0 to 1 and e_n from $135,000$ to 1. The Facebook logo is the least novel. Next, an image of Apple TV2, then a stock photo, a photo of a self checkout system, a us quarter back, and a diagram of power grid.

SEARCH-BASED NOVELTY

Search-based novelty measures the uniqueness of a media element with the number of pages it appears on the Web. We show an example implementation using images as media elements in the next section. The process for implementing a search-based novelty metric includes four steps:

1. Collect a set of sample of media elements.

2. Generate search queries for each element algorithmically.

3. Perform searches with each generated query, collecting the number of search results for each element.

4. Create a function that maps novelty inversely to the number of search results produced by an element.

Search-Based Image Novelty

We present our implementation, *search-based image novelty*, which uses Google Image searches to measure the novelty of images found on the Web.

Google Images indexes almost all images on the Web, allowing one to search for images using text or image queries. Google Images combines very similar images, making the search invariant to resolution and small visual perturbations.

First, collect a set of sample images C (1). We used a set of 3,579 images from the information compositions from the field study described in the next section. For each image, generated a search query (2) using each image's url as the query. For each image in C, perform a Google Image search to get the number times the images appears on the Web e_n (3). The number of returned search results determines popularity (e_n).

To create a function that normalizes e_n (4), take the sum of the logarithms of the popularity of images from all compositions, deriving average popularity ($\bar{c}$). Then double average popularity ($\bar{c}$) to estimate a maximum, and calculate the normalized image novelty ($inov$) of an element (e). This effectively sets a maximum popularity.

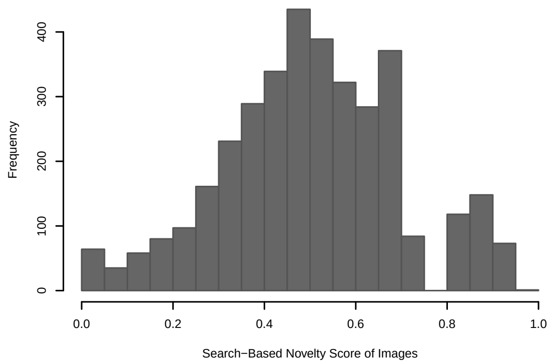

Figure 2. Distribution of the search-based image novelty ($nov_i(e)$) for 3,579 image queries.

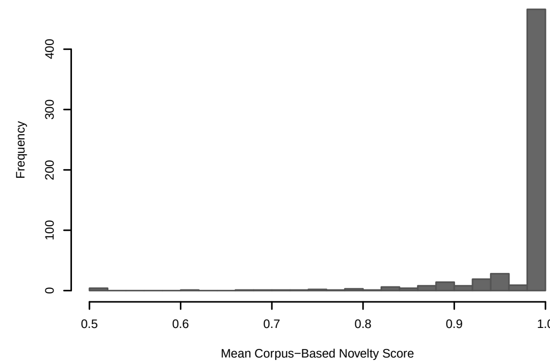

Figure 3. Distribution of corpus-based image novelty scores of 682 student authored compositions from DPCE. Even with the large number of student made compositions, the lack of overlap creates a small range of image novelty scores.

$$\bar{c} = \frac{\sum_{\forall e \epsilon C} log(e_n)}{|e \epsilon C|} \quad (3)$$

$$inov(e) = 1 - \max(1, \frac{\log e_n}{2 \cdot \bar{c}}) \quad (4)$$

Thus, $inov(e)$ is a function for novelty that does not depend on a corpus of collected responses. $\bar{c}$ does not need to be calculated for every experiment. We found $\bar{c} = 5.02$, to get the resulting function $inov(e) = 1 - \max(1, log(e_n)/10.04)$. $inov$ will always be between 0 (not novel) to 1 (very novel). We show the histogram for $inov$ scores over C in Figure 2. Figure 1 shows example images such as the Facebook logo, which is the least novel.

FIELD STUDY

Laboratory experiments help researchers focus on factors that contribute to creativity by providing a controlled environment. In contrast, investigations in the field provide ecological validity. We conducted a field study in The Design Process: Creativity and Entrepreneurship (DPCE), an interdisciplinary undergraduate course.

DPCE students used the creativity support tool, InfoComposer [10], to author information compositions on soft innovations – new ideas formed from combining and extending existing ideas. Students searched the Internet, collecting relevant image and text bookmarks to represent ideas about their soft innovations. Students' search queries and potential information sources were *unrestricted*.

To calculate the novelty of a composition, which contains a set of images, we first calculate novelty for each of its images and then calculate a mean. We compare the prior corpus-based image novelty (Equation 1, Figure 3) and the new search-based image novelty (Equation 4, Figure 4).

Corpus-Based Image Novelty Results

Corpus-based image novelty yielded a small range of values, leaning heavily toward high novelty. Figure 3 illustrates how

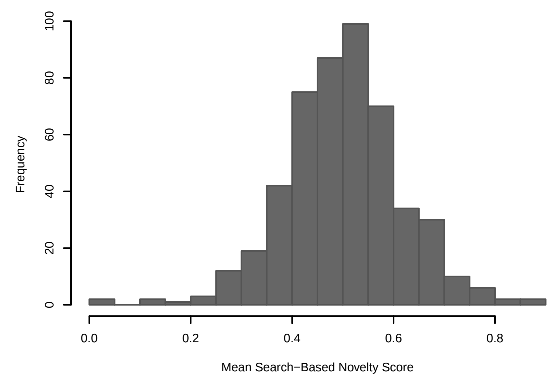

Figure 4. Distribution of search-based image novelty scores of 682 student authored composition from DPCE. Using the same images used in Figure 3, search-based novelty produces novelty scores with more granularity.

poorly the metric is suited for non-fixed queries, showing a value of 1 for most compositions. Such extreme values have do not contain enough granularity for valid comparison.

Search-Based Image Novelty Results

Search based image novelty scores provided a higher range of values. These values are what one would expect for novelty, showing a range of values. Avoiding ties in values makes this metric better suited for statistical test that require interval values, which are useful for comparing creativity support environments in field studies and laboratory experiments.

DISCUSSION

In this section, we discuss how search-based methods for measuring the novelty of elements could be applied to other kinds of media. We discuss benefits of using automated metrics to analyze products made with creativity support environments.

More Media

We have shown search-based novelty more effectively describes image novelty than corpus-based novelty when images can be collected from anywhere on the Web. While we have only implemented and tested search-based image novelty, we suspect that similar implementations would work with other kinds of media. Once one can generate queries and perform searches for an element of an arbitrary media type, then implementation of a search-based novelty score is straight forward. We envision this metric being used to assess elemental novelty for collections of various media including, audio, text, or video. For text, one could perform searches using samples of extracted text. For audio, one could search using queries from tags.

Aggregating Novelty

Justifying the methods for aggregating elemental novelty into a single score is difficult. We have used the mean of each element to represent the novelty of a collection, but we do not have a clear understanding of which aggregation would be best. As researchers, we hope to find significant differences between conditions in experiments. Transformations designed to accentuate difference could be used on these metrics, but should not be used without theoretical justification.

Elemental novelty does not consider the context of a collection, ignoring relationships between elements. If a collection includes a very popular image among moderately novel images, the overall novelty of the collection may be low. If an author creates novel combinations of ideas with common elements, they may score low on elemental novelty.

We continue to develop human rated metrics that measure holistic components of creativity [6]. While human rated metrics take longer to calculate than automated processes, they can be used in small experiments. Using crowdsourcing services can help mitigate the time any one individual has to spend to rate aspects of collections.

Implications for Design

We have shown that metrics that work well in more constrained contexts can perform poorly without fixed search queries. Big data is a resource that can be used as a baseline for the originality of content. If researchers can uncover properties that leverage publicly available web services to analyze media, they can use those processes to measure the effectiveness of tool that help people create media rich products. Combining empirical results with experiential data ensures that both the experience and the product of creativity are analyzed.

Statistical methods on computational processes for evaluating creative products help researchers analyze more data because they do not require human analysis. Computational methods can be easier to verify and repeat across research groups. Evaluation at web scale is of interest, because it involves larger sets of users, which are essential to investigating the validity of at least some creativity support environments, such as large university courses, and massively open online courses. Such large scale deployments of creativity support environments necessitate techniques that enable equally large scale automatic assessment of ideation metrics.

Conclusion

We used observations and data from a field study to show that search-based image novelty measures are more effective when the corpus of collected images yields few commonly collected images. By leveraging Google Images, we computed a baseline for image novelty to construct a useful normalized novelty metric. We presented directions for future work that extend this metric to other media types that leverage search to measure novelty.

REFERENCES

1. Arnheim, R. *Visual Thinking*. UC Press, 1972.
2. Carroll, E. A., Latulipe, C., Fung, R., and Terry, M. Creativity factor evaluation: towards a standardized survey metric for creativity support. vol. In Proc. Creativity and cognition (2009).
3. Dow, S., Fortuna, J., Schwartz, D., Altringer, B., Schwartz, D., and Klemmer, S. Prototyping dynamics: sharing multiple designs improves exploration, group rapport, and results. vol. In Proc. CHI (2011).
4. Glenberg, A. M., and Langston, W. E. Comprehension of illustrated text: Pictures help to build mental models. *Journal of Memory and Language 31*, 2 (1992), 129 – 151.
5. Kerne, A., Koh, E., Smith, S., Choi, H., Graeber, R., and Webb, A. Promoting emergence in information discovery by representing collections with composition. vol. In Proc. Creativity and cognition (2007).
6. Kerne, A., Webb, A. M., Linder, R., Smith, S. M., Moeller, J., Damaraju, S., Lupfer, N., and Qu, Y. Components of information-based ideation productivity. *submitted to ToCHI* (2013).
7. Salton, G., and Buckley, C. Term-weighting approaches in automatic text retrieval. *Information processing & management 24*, 5 (1988), 513–523.
8. Shah, J. J., Smith, S. M., and Vargas-Hernandez, N. Metrics for measuring ideation effectiveness. *Design Studies 24*, 2 (2002), 111 – 134.
9. Shah, J. J., Smith, S. M., Vargas-Hernandez, N., Gerkins, D., and Wulan, M. Empirical studies of design ideation: Alignment of design experiments with lab experiments. In *Proc ASME Conf on Design Theory and Methodology* (2003), 1–10.
10. Webb, A., Linder, R., Kerne, A., Lupfer, N., Qu, Y., Poffenberger, B., and Revia, C. Promoting reflection and interpretation in education: Curating rich bookmarks as information composition.
11. Webb, A. M., and Kerne, A. Integrating implicit structure visualization with authoring promotes ideation. In *Proc. JCDL* (2011), 203–212.