Info Vis

Article

Information Visualization
2017, Vol. 16(4) 291–308

DOI: 10.1177/1473871616669193
journals.sagepub.com/home/ivi
SAGE

Map- or list-based recommender agents? Does the map metaphor fulfill its promise?

Rene F. Reitsma[1], Ping-Hung Hsieh[1], Anne R. Diekema[2], Robby Robson[3] and Malinda Zarske[4]

Abstract

We present a spatialization of digital library content based on item similarity and an experiment which compares the performance of this spatialization relative to a simple list-based display. Items in the library are elementary school, middle school, and high school science and engineering learning resources. Spatialization and visualization are accomplished through two-dimensional interactive Sammon mapping of pairwise item similarities computed from the joint occurrence of word bigrams. The 65 science teachers participating in the experiment were asked to search the library for curricular items they would consider using as part of one or more teaching assignments. The results indicate that whereas the spatializations adequately capture the salient features of the library's content and teachers actively use them, item retrieval rates, task-completion time, and perceived utility do not significantly differ from the semantically poorer but easier to comprehend and navigate list-based representations. These results put into question the usefulness of the rapidly increasing supply of information spatializations.

Keywords

Digital library, spatialization, experiment, map metaphor, Sammon map

Introduction: does the map metaphor fulfill its promise?

For quite some time, proposals have been formulated for map-like representations of inherently non-physical, non-geographic information. These so-called spatializations (Skupin[1]) are aimed at exploring, summarizing, searching, comprehending, and navigating such information visually rather than through traditional, nonvisual formats. Figure 1, for instance, represents a two-dimensional (2D) layout by Boyack et al.[2] of over 2 million biomedical articles.

Early explorations, mostly from the 1990s and early 2000s with some pioneering work conducted as early as 1974, were conducted by information and library scientists interested in the structure and dimensionality of co-citation and text corpora. Examples of this work are Small and Griffith,[3] Small and Garfield,[4] Everett and Pecotich,[5] Chalmers,[6] Wise et al.,[7] Chen et al.,[8,9] Small,[10] Noyons and Van Raan,[11] and Chen.[12] Compendia of those earlier works are provided by Chen,[13] Dodge and Kitchin,[74,15] Börner and Chen,[73] and Geroimenko and Chen.[16]

Due to the rapid growth of social media and linked/networked data, more recent studies and proposals exhibit a far richer mix of application domains including blog postings, social networks, Web-based

[1]College of Business, Oregon State University, Corvallis, OR, USA
[2]Southern Utah University, Cedar City, UT, USA
[3]Eduworks Corp., Corvallis, OR, USA
[4]University of Colorado Boulder, Boulder, CO, USA

Corresponding author:
Rene F Reitsma, College of Business, Oregon State University, Corvallis, OR 97331, USA.
Email: reitsmar@bus.oregonstate.edu

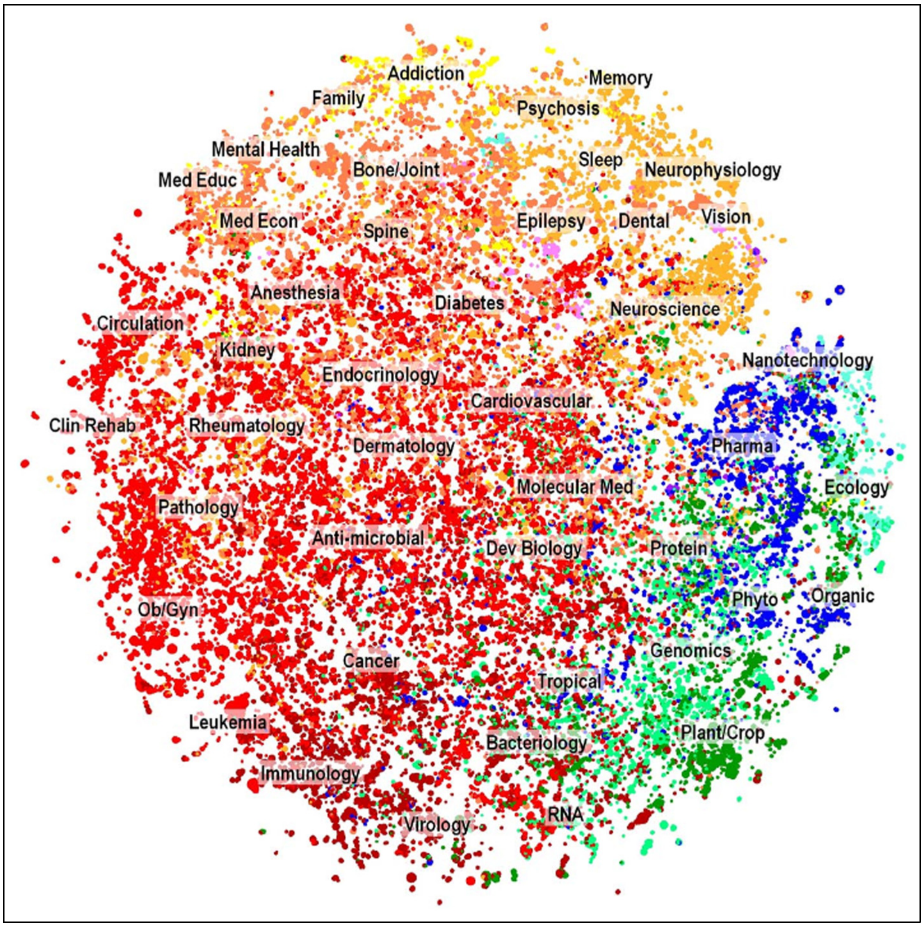

Figure 1. Spatialization of 2 million-plus biomedical publications (Boyack et al.[2]).

collaborations, and so on. Some examples from the academic literature from the first decade of the 2000s are Wang,[17] Boyack et al.,[18] Buzydlovky et al.,[19] Skupin,[20] Wang et al.,[21] Van Ham and Van Wijk,[22] Perer and Shneiderman,[23] Henry and Fekete,[24] Leydesdorff,[25] Burns and Skupin,[26] Zhu et al.,[27] Boyack et al.,[2] Lieberman et al.,[28] and Dörk et al.[29] For a mid-decennium overview, we refer to Börner.[30] Some more recent applications are Skupin et al.,[31] Burns and Skupin,[32] Chen et al.,[33] Liu et al.,[34] and Wang et al.[21] Another 80-plus recent examples, mostly from non-academic sources, can be found in Lima's[35] compendium on *Visual Complexity*.

Although these spatializations differ in their use of geometry, dimensionality, spatial layout methods, and overall interpretation and use of space, they have some common characteristics:

- They employ the well-known map metaphor, that is, the visualized information is projected into an *n*-dimensional (*n*D) space. In most, although not all applications, this space has a metric defined on it so that items' positions in it can be expressed in terms of that metric. The resulting information spatializations, especially the three-dimensional (3D) ones, are sometimes referred to as information "landscapes." Although we ourselves do not object to this casual use of the term, Fabrikant et al.[36] take exception to this use of the term, claiming "landscape" for the domain of geomorphology.
- Either explicitly or implicitly, the map makers espouse the theory that spatialization of information makes available means of comprehension and navigation which are absent from non-spatial representations. Most prominent are the notions of "landmark" (Sorrows and Hirtle[37]), "wayfinding" (Blades[38]), and "focus + context" (Mukherjea and Hara,[39] Baudisch et al.,[40] Cockburn et al.,[41] Liu et al.[34]).
- Another explicit or implicit assumption is that if the modeled information is inherently multidimensional, making at least some of that dimensionality explicitly may reveal aspects and characteristics

which would otherwise remain hidden (Gracia et al.[42]). Ordination techniques such as factor-analytical and multidimensional scaling (MDS) methods are commonly used to explore and model the dimensionality of the mapped information.

- Either implicitly or explicitly, mapping items in an information space is motivated by an appeal to our familiarity with 2D and 3D phenomena and indeed with traditional geographical maps as well as 2D pictorial and graphic representations of information (Fabrikant et al.[36]).
- The majority of maps are 2D. This is partly because of the limitation of display facilities for 3D information (Nielsen[43]) and partly because good evidence exists that the extra cognitive load of having to navigate additional dimensions in these abstract, non-physical spaces does not consistently translate into a corresponding gain in performance in terms of retrieval and usability (e.g. Sebrechts et al.,[44] Tory et al.,[45] Westerman et al.,[46] Ting et al.,[47] Johansson et al.,[48] Gracia et al.[42]).

Despite the great promise and the esthetic appeal of many spatializations—we speculate that it is not a coincidence that Lima's compendium (Lima[35]) was published in coffee-table format as many of the featured spatializations are quite striking—one may ask how effective they are when used in actual information search and exploration settings. Our non-inclusive survey of the literature seems to indicate that actual, task-based testing of these interfaces is quite rare. For instance, we were unable to find even a single use testing case for any of the 80-plus spatializations featured in Lima's book. A small portion of the older applications does contain user studies, but these typically have (very) small sample sizes, use within-subject experimental designs, rely on informally collected comments and anecdotal evidence, or juxtapose different types of spatializations instead of testing a spatialization versus a non-spatialized alternative. Some of the more carefully designed and executed studies are the ones by Chen et al.,[8,9] Sebrechts et al.,[44] Ebinger and Skupin,[49] Zhu et al.,[27] Borkin et al.,[50] and Buzydlovky et al.[51]

Although additional empirical evidence of efficacy of these spatializations may exist, from the scarcity of user studies we found in our search, we infer that the growing enthusiasm for generating these spatializations is not matched by a corresponding amount of user and efficacy studies. Others before us have observed this shortage as well. Fabrikant et al.,[36] for instance, observe that with few exceptions "Most others who present information landscapes do not even bother to outline why they think that the displays they create might be effective."

In summary, we observe that there is a long-standing and rich tradition of the use of spatializations, that is, map-based visualizations of inherently non-spatial information, and that the recent proliferation of linked data has reinvigorated this use with a surge of new forms and variations in these spatializations. In addition, Lima's compendium (Lima[35]) provides evidence that these visualizations are no longer the explicit domain of academics but can nowadays be found across public media. For instance, Scientific American, in collaboration with the Office of Creative Research (OCR), recently published a map of articles on relativity theory (OCR[52]). We also observe, however, that although some assessment of the usefulness and efficacy of these visualizations for real-world information searching and retrieval tasks has been conducted, additional assessment is needed.

In this article, we develop a 2D spatialization of digital library content. After validating the spatialization against the structure of the library, we submit it to a between-group experiment in which we compare the efficacy of the spatialization to help real-world users search the library with a traditional, non-spatialized list-based alternative. We find that although subjects choose to use the spatialization and express appreciation for it, the results in terms of retrieval rates and confidence in having adequately explored the search space do not differ significantly between the two conditions. These results challenge claims on usefulness and efficacy of this type of visualizations.

Application context and expected benefits

Following the promise of spatializations as discussed above, we are interested in whether such a spatialization can be successfully used in actual user settings. Because of our work in Kindergarten through grade 12 (aka "K-12") science and engineering digital libraries, we are interested in exploring how and whether spatialization can help practicing science teachers find curriculum in on-line Science, Technology, Engineering, and Math (STEM) libraries. A typical use case for this context is science teachers using such a library in search of curricular materials which they would consider using as part of their teaching assignments. For example, if a given teaching assignment is that "third grade students can plan and conduct an investigation to provide evidence of the effects of balanced and unbalanced forces on the motion of an object" would a spatialization of a STEM library content help them find curriculum which they would consider using in teaching to this assignment?

We care to point out that this success is not to be interpreted in terms of "recall" and "precision" as commonly used in information retrieval experiments (Baeza-Yates and Ribeiro-Neto[53]). That is, success is not measured in terms of the percentage of the existing resources found (recall) or the accuracy of item retrieval (precision). Convenient as it would be to have a so-called gold-standard set of vetted and validated assignment/curriculum alignments, previous work by Devaul et al.,[14] Marshall et al.,[53,54] Reitsma et al.,[55] and Wetzler et al.[56] has shown that such alignment is multidimensional, context-sensitive, and quite variable between individual teachers. Although this has not prevented curriculum collection builders from offering users predetermined alignments, we are quite hesitant to accept this notion of objective and "true" alignment. Instead, in accordance with the literature in this area as well as with actual system use (Diekema and Olsen,[57] Kelly[58]), we leave this decision to the individual teacher. Hence, we define success as the average number of curricular items retrieved across one or more teaching assignments for a given amount of search time:

Hypothesis 1: Science teachers having access to the spatialized version of library content will find more curricular items than those supported by a non-spatialized version of that content.

In one common class of spatializations, the distances between items represent the (dis)similarity between them. Items which are in close proximity to each other are similar to each other and different from the ones farther away. Hence, once a user decides that an item is attractive or suitable, items in the immediate vicinity or neighborhood of the chosen item become likely candidates for being chosen as well and items farther away may be dismissed. Given a spatialization's explicit affordance of this notion of "neighborhood" or proximity, we expect that if the search space's geometry corresponds to that of the spatialization:

Hypothesis 2: Science teachers having access to the spatialized version of library content will be more confident that they have adequately explored the search space than those supported by a non-spatialized version of the content.

Finally, we expect that users of a spatialization who are informed about it, have been trained to use it, and have made a serious attempt at using it, can express whether they consider that spatialization helpful in finding curricular items or not. Hence,

Hypothesis 3: Science teachers having access to a spatialized interface will consider the spatialization more helpful for finding curriculum than those having access to a non-spatialized interface.

Spatialization of a K–12 STEM digital library

Our application domain consists of www.TeachEngineering.org, a digital library of K–12 STEM content (Sullivan et al.[58]). The library contains 1400-plus curricular items loosely organized in a hierarchical structure: "curricular units" contain one or more "lessons" and "lessons" contain one or more "activities." Although the hierarchy allows the skipping of tranches, for example, activities can be part of curricular units or indeed exist just by themselves, it is strictly unidirectional; for example, lessons cannot be contained by activities and units cannot live on activities or lessons. The project started in 2003 as part of the US National Science Foundation's National Science Digital Library (NSDL) project (Zia[59]). To date, 35 universities have contributed materials to the library. Its patronage consists of 3.5 million-plus users and has an annual growth rate of 35%.

Borrowing from the literature on educational recommender systems (Neumann,[60] Santos and Boticario[61]), we opted for a content-based spatialization of the library. Since TeachEngineering users are neither tracked nor identified, a spatialization based on navigational patterns, user reviews, or user ratings was deemed impractical.

As the method for spatialization, we chose a standard approach consisting of a combination of pairwise item similarity computation followed by ordination, that is, we compute the similarity between all pairs of curricular items and use these similarities as input for a 2D ordination. Obviously, reliance on pairwise item similarities implies that this method is not suited for collections containing more than a few thousand items since the number of pairs grows quadratically. However, most K–12 STEM collections, including TeachEngineering, are well within this range, and we therefore considered this approach acceptable.

Similarity measure

Our measure of similarity (1) expresses the coincidence of bigrams, that is, two-word terms which co-occur in documents (Tan et al.,[62] Bekkerman and Allan[63]), weighted by the root mean square of their inverse document frequencies (IDFs). More precisely, for any document D, define $\beta(D)$ to be the set of bigrams in D. Let $S = \{S_1, \ldots, S_n\}$ be a corpus of documents to which a document D is to be compared and let S_i be a document in S. For each $b \in \beta(S_i)$, let

w_b be the IDF of b, that is, the reciprocal of the number of $S_i \in S$ containing b. For any $b \in \beta(D)$ let $\delta_i(b) = 1$ if $b \in \beta(S_i)$ and 0 otherwise and let $N = |\beta(D) \cap \beta(S_i)| = \sum_{b \in \beta(D)} \delta_i(b)$ be the number of bigrams shared between D and S_i. Then the measure of similarity μ between D and S_i is defined as N times the root mean square of the IDFs of the shared bigrams, that is, as

$$\mu(D, S_i) = N\sqrt{\frac{1}{M}\sum_{b \in \beta(D)} \delta_i(b) \times w_b^2} \tag{1}$$

where $M = |\beta(D)|$ is the number of bigrams in D.

We note that whereas it is standard procedure to count the number of times that b appears in S_i, that is, the term frequency (Baeza-Yates and Ribeiro-Neto,[53] Manning et al.[64]), the measure in equation (1) only registers whether it appears or not. The reason for our choice is that adding term frequency over-weights larger documents in S and larger source documents D. Our measure is not entirely insensitive to size, however, since a larger document has more terms and therefore a larger random probability of having a bigram shared with another document. However, it avoids most of the over-weighting problem. Furthermore, although our measure is not normalized, it produces a scale for each corpus S in which higher numbers empirically indicate more similarity; smaller numbers indicate less similarity; and unrelated resources typically produce exceedingly small similarity values. Our measure can therefore be used effectively for ranking similarity against all documents in a single corpus and is faster to compute than one that uses term frequency weighting. We also note that in our bigram extraction, we ignore common words—so-called stop words—and include only those terms which are nouns, verbs, adjectives, or adverbs. Stemming and/or lemmatizing further distills the term set.

We opted for the extraction of word bigrams rather than single terms or n-grams for $n > 2$ for several reasons. Heuristically, bigrams represent a balance between not needing word sense disambiguation to detect semantic similarity and being able to consistently detect similarity when it is present. Whereas single-word terms often suffer from semantic ambiguity, for example, "solar" has a different meaning in "solar energy" than in "solar system" or "solar radiation," trigrams (and other n-grams) over-reduce the density of matches, leading to false negatives, for example, "fast operating system" versus "slow operating system." Moreover, and perhaps critically important, is that bigrams which carry very special meaning are ubiquitous in a STEM collection such as TeachEngineering. Bigrams such as "solar energy," "solar system," "kinetic energy," and "Pythagorean theorem" are both common and much more distinguishing than the single-word terms comprising them. Similarly, a term such as "laws of motion" reduces to a bigram ("laws motion") after removal of the stop word ("of").

In Reitsma et al.,[65] we provide a detailed analysis of the performance of this similarity measure in the context of the TeachEngineering collection. The analysis shows precision of 0.90, 0.85, 0.78, 0.73, and 0.70 at ranks 1–5, respectively. We considered this performance sufficiently good to use the similarity measure as the basis for a spatialization.

Spatialization

In computing a spatialization from these similarities, we seek an ordination with the following characteristics:

- It must adequately represent the configuration of library items in two dimensions, that is, the inverse of the item similarities must adequately represent the inter-item distances in the spatialization.
- TeachEngineering curricular units and their lessons and activities should spatially cluster, that is, on average, the distances between units' items should be less than that between items across units since items grouped in a unit are more likely to be similar than items outside this unit.
- Clustering should repeat at different levels of the collection hierarchy. For instance, TeachEngineering contains two curricular units on so-called "simple machines." Although these units should ideally form distinct clusters, their contents are sufficiently similar to each other yet dissimilar from other units that they themselves should again cluster and be spatially separated from other units.
- On average, lessons should cluster more than activities, that is, since activities comprise widely varying applications of a common concept, they can be expected to differ more from each other than a unit's lessons differ from each other.
- Exceptions to these patterns will exist. One of the most promising aspects of the spatialization is that some items belonging to different curricular units will be placed in each other's proximity because their contents are in fact very similar. For instance, an activity from a unit on "absorption" could be related to a lesson on "environmental cleanup" which uses absorption technologies for oil spill mitigation. Whereas the above-stated patterns would simply reify the collection's hierarchy in its use, these exceptions point users to unexpected yet potentially useful cross-curricular relationships between items.

After reviewing and testing a variety of ordination techniques (principal components analysis (PCA), metric and nonmetric MDS, and various force-directed methods), we opted for a so-called Sammon map (Sammon;[66] Sun et al.[67]) using *1/similarity* as our measure of distance. Statistical fit results are discussed in section "Spatialization results." We furthermore opted to render the ordination using the Google Maps JavaScript API (Google[68]). This API makes it straightforward to embed the spatialization in the Web pages of TeachEngineering items and provides basic and familiar means of map navigation such as panning and zooming.

Collection preparation

As is the case for most Web-based digital library collections, TeachEngineering curricular items are amalgamations of content and metadata. Since some of the metadata, for example, stated copyrights, author names, internal references to other items, and publication data, can easily bias any text-based similarity estimator, all metadata were removed from the items prior to computing the similarities and the subsequent Sammon maps. Moreover, to assess the correspondence of the spatialization with the collection's hierarchical structure, we removed all activity items which were not part of the collection hierarchy, that is, all activities which were not part of either a lesson or curricular unit. This resulted in a test collection of 958 items. We care to stress that, as far as we know, all information relating to item hierarchy were removed from all items prior to similarity estimation, that is, similarity estimator (1) was given no information about the collection's structure.

Spatialization results

Rather than generating a single map covering the entire collection, we opted to give each curricular item its own map consisting of the item itself, the five items most similar to the chosen one, and any and all of their "relatives," where a "relative" is defined as an item which is part of the same curricular unit as the item in question. In other words, for each curricular item in the collection, we display all its family members, the items in its immediate semantic neighborhood, and all their family members.

Figure 2 displays the Sammon map of a *simple machines* curricular unit. The chosen item is rendered in magenta and items are labeled as curricular units (C), lessons (L), or activities (A). Lines between items indicate family (hierarchy) relationships. We observe that although none of the hierarchical item information was used to compute the item similarities, the desired "family" clusters are expressed in the map. The map contains two curricular units: the chosen one (magenta) and another *simple machines* unit which, although not related by means of the item hierarchy, is "pulled into" the map because of its items' semantic similarities. We furthermore observe that as expected, lessons cluster more than activities. This is particularly shown in the top cluster where activities are pushed to the map's periphery because their content differentiates them from their lessons and from each other. We also note that the bottom cluster of the items is quite concentrated, indicating close semantic resemblances of its items. However, when we zoom in on it (Figure 3), we once again see the pattern of its lessons showing a higher degree of clustering than its activities.

The map in Figure 4 for the activity *Biomimicry: Natural Designs*, once again shows the pattern of activities radiating out from the center of a curricular unit. We also observe the cluster separation between curricular units.

Figure 5 shows an example of curricular "crossovers." In this case, the chosen *Sugar Spill* activity's map pulls in two additional curricular units and their family members. Once again, the units and the curricular items comprising them are spatially clustered and easily recognizable. Still, we see two activities and one lesson from another unit associated with oil spill cleanup located in the immediate vicinity and to the "north" of the *Sugar Spill* activity. Spatial clustering, once again, appears to follow the expected pattern.

Formal assessment

To formally assess the degree to which the spatial clusters correspond to the collection's curricular units, we followed a two-step procedure:

1. Cluster the Sammon results in as many clusters as there are curricular units in the corpus, forcing each Sammon cluster to contain as many items as there are present in a corresponding curricular unit, and center each Sammon cluster on the mean coordinates of its members.
2. Compute the associative statistics (χ^2, Cramer's V) for the contingency table of the clusters computed in Step 1 versus the corresponding curricular units in the corpus. Ideally, all frequencies are on the diagonal of this table, indicating perfect Sammon-curricular unit matching and Cramer's V will be 1.0 and χ^2 will be highly significant. Off-diagonal frequencies indicate cluster mismatches, but as long as these are rare, V and χ^2 will still be statistically significant.

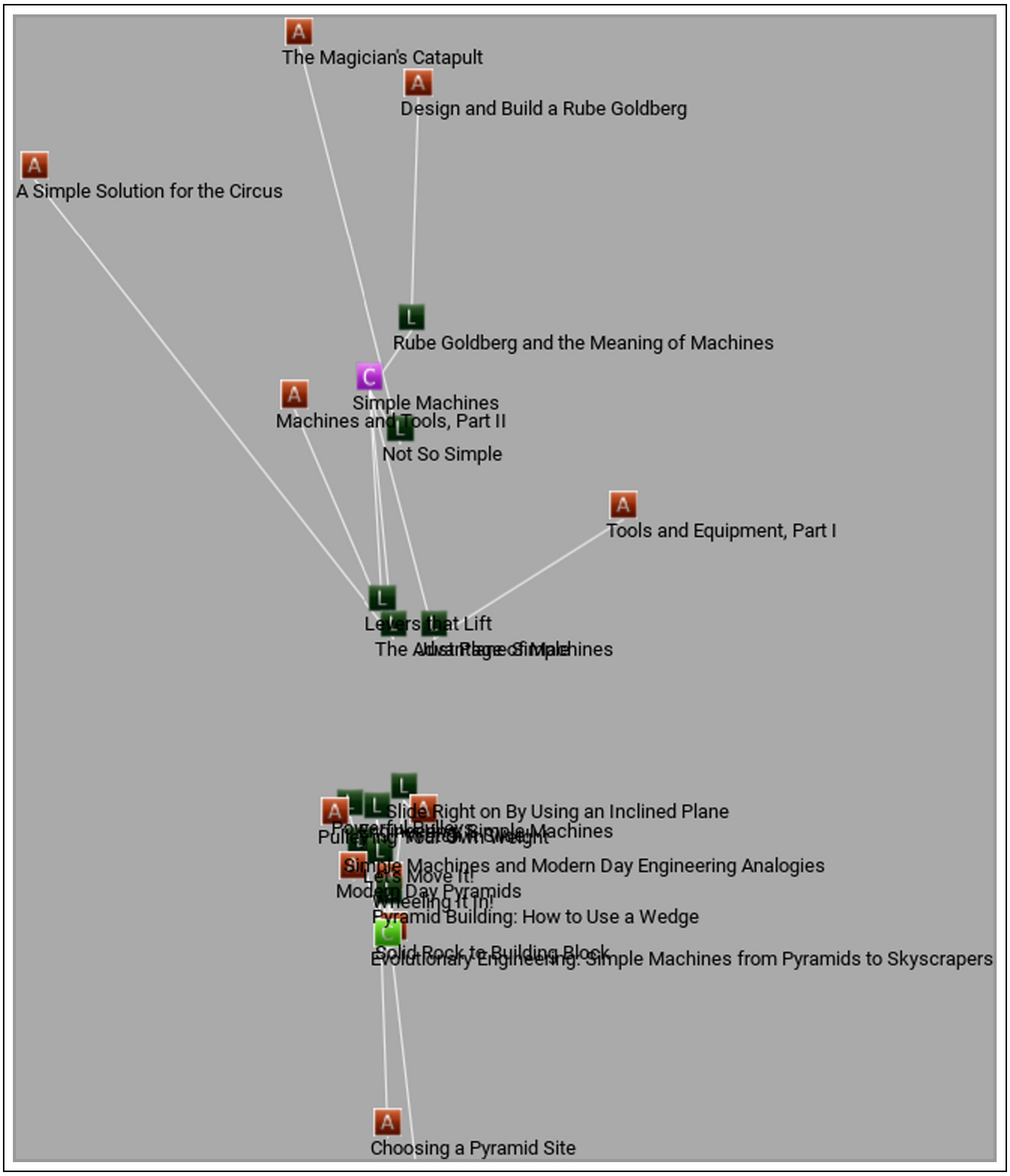

Figure 2. Sammon map of two *simple machines* curricular units. For a detailed, zoomed-in view, see Figure 3.

Although Step 1 of this procedure does not correspond to either a standard hierarchical or K-means cluster analysis, it can be expressed as a mixed integer linear programming (ILP) problem of the following kind:

Constants:

- d_{ij} = Sammon space distance between item i and the center (mean x–y coordinates) of its curricular unit j
- c_j = number of items in curricular unit j

Decision variables:

- x_{ij} = 1 if item i is part of curricular unit j; 0 otherwise

Objective function

$$\min \sum_{i,j} d_{ij} \times x_{ij}$$

Constraints

$$\forall j \sum_{i} x_{ij} = 1$$

$$\forall i \sum_{i} x_{ij} = c_j$$

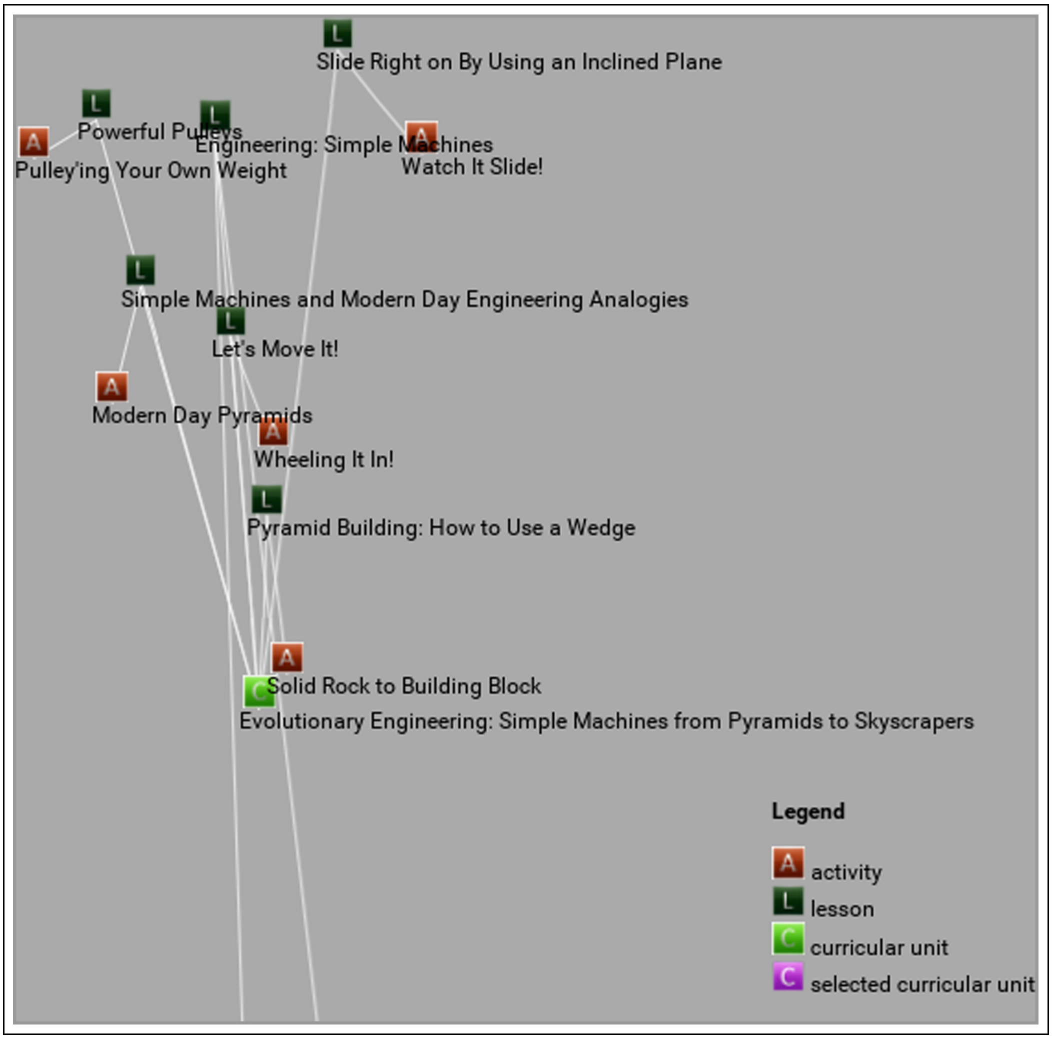

Figure 3. Zoom-in on the second *simple machines* unit.

Running this model for the entire collection implies solving for almost 100,000 variables. Instead, we took a 15% random sample containing 11 (of 75) curricular units containing 107 items. This represents a problem comprising 107×11 = 1177 variables which is readily solvable. Applying Step 2 to this sample results in a contingency table with an almost perfectly filled diagonal ($\chi^2 = 449.80$, degree of freedom (df) = 100; $p < 0.001$; V = 0.943). Since it is possible that the solution is sensitive to the constraint that clusters must have the same number of items as their corresponding curricular units, we recomputed the ILP after dropping that constraint. (We note that dropping the constraint renders the problem equivalent to computing the space's ordinary Voronoi sets (Okabe et al.[69]).) This resulted in only nine off-diagonal cells with non-zero frequencies. The goodness-of-fit statistics ($\chi^2 = 431.60$, df = 100; $p < 0.001$; V = 0.914) indicate that the constraint does not affect the statistical significance of the association between membership of a Sammon cluster and membership of a curricular unit. Having good interpretative as well as statistical evidence that the spatialization encapsulates the structure (hierarchy) of the collection well, we proceed with comparative analysis.

Comparison experiment

We designed a simple between-group experiment in which subjects are asked to use the TeachEngineering system equipped with either the map interface or a simple list-based interface to find curricular items which they deem supportive in teaching to one or more teaching assignments. Since K–12 instruction, at least in the United States, is strongly driven by curriculum standards, teachers tend to look for resources that specifically relate those standards (Diekema and Olsen[57]). In designing the task for the experiment, we chose a

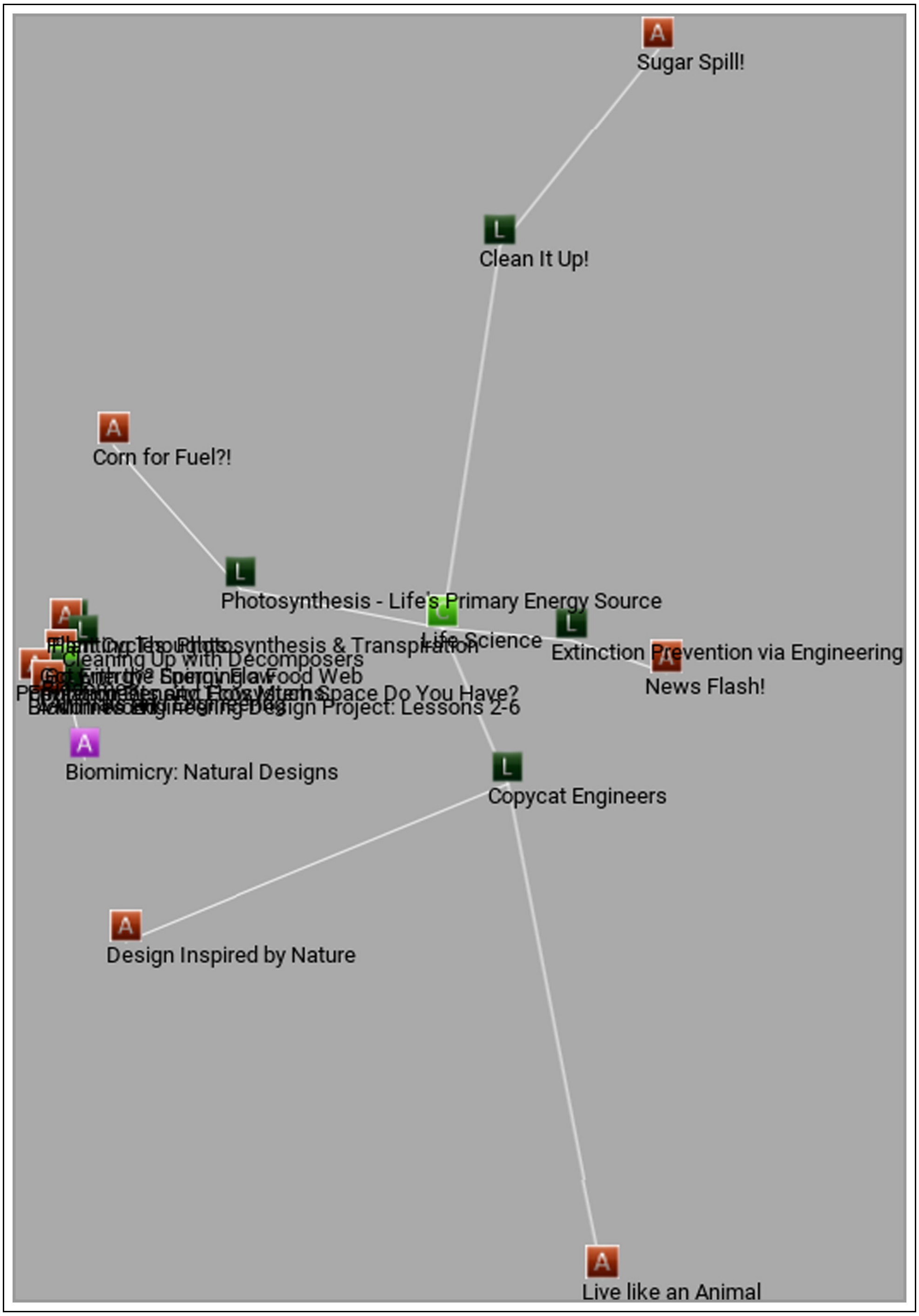

Figure 4. Biomimicry example.

task that participants commonly carry out, namely, finding curriculum which they would consider using in teaching to a standard. The set of assignments consisted of 11 (real-world) K–12 science standards (Table 1) used in an earlier experiment by Reitsma et al.[55] to uncover the multidimensionality of educational standard alignment. These standards were selected to represent appropriate task context in terms of grade level, topic, content versus inquiry, and digital library content (Borlund[70]). Although most of them have since been superseded by newer ones—the US educational science standards' "landscape" is complex. It contains 65,000-plus different standards, about 10% of which change every year (Sumner et al.[71])—we

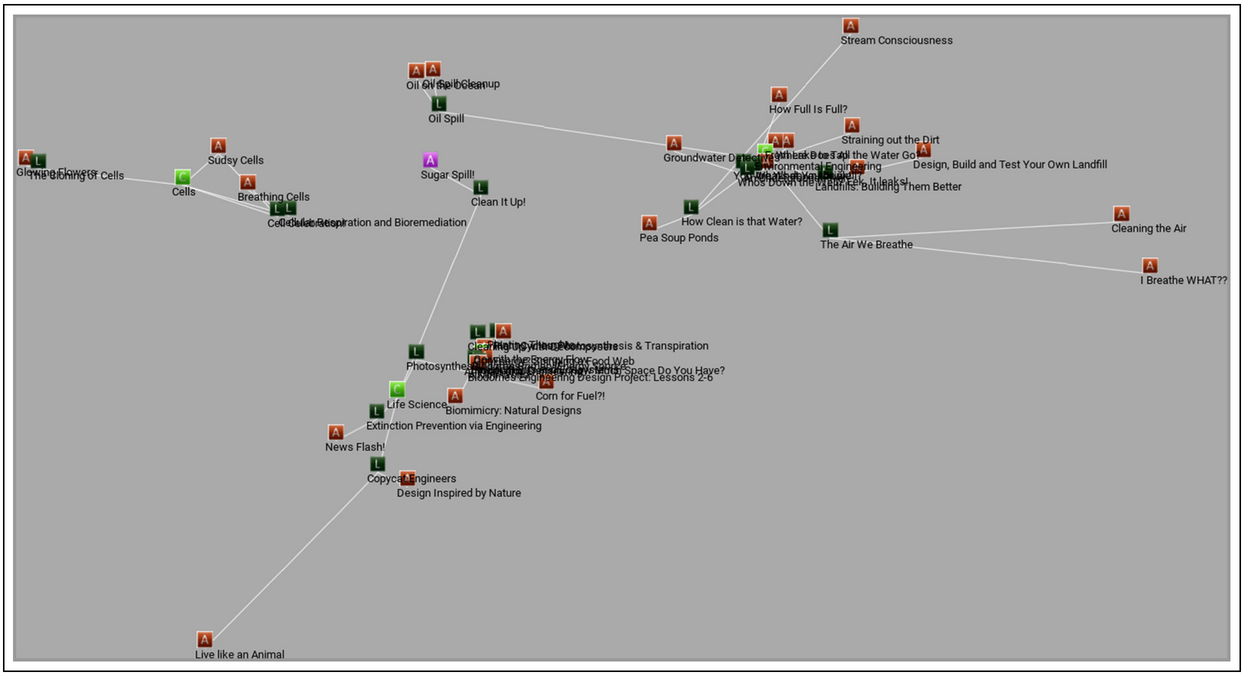

Figure 5. *Sugar Spill* activity and its semantic neighborhood. Three items from a different curricular unit on *Environmental Engineering* are nearby.

Table 1. 11 science educational standards used as teaching assignments in the experiment.

Teaching assignment	US state	Grade	Assignment text
T01	Virginia	10	Student will investigate and understand characteristics of sound and technological applications of sound waves. Key concepts include resonance
T02	Virginia	9	Student will investigate and understand the organization and use of the periodic table of elements to obtain information. Key concepts include classification of elements as metals, metalloids, and nonmetals
T03	Virginia	10	Student will investigate and understand basic principles of electricity and magnetism. Key concepts include static, current, and circuits
T04	Maryland	4	Life Science—the students will use scientific skills and processes to explain the dynamic nature of living things, their interactions, and the results from the interactions that occur over time
T05	Maryland	10	Student will explain changes in Earth's surface using plate tectonics
T06	Massachusetts	10	Recognize the purposes of zoning laws and building codes in the design and use of structures
T07	Massachusetts	7	Given a prototype, design a test to evaluate whether it meets the design specifications
T08	Massachusetts	11	Reactants, products, and purposes of photosynthesis and respiration
T09	North Carolina	8	Learner will conduct investigations, use models, simulations, and appropriate technologies and information systems to build an understanding of cell theory
T10	North Carolina	7	Learner will conduct investigations and utilize appropriate technologies and information systems to build an understanding of the atmosphere
T11	New Jersey	7	All students will integrate mathematics as a tool for problem-solving in science, and as a means of expressing and/or modeling scientific theories

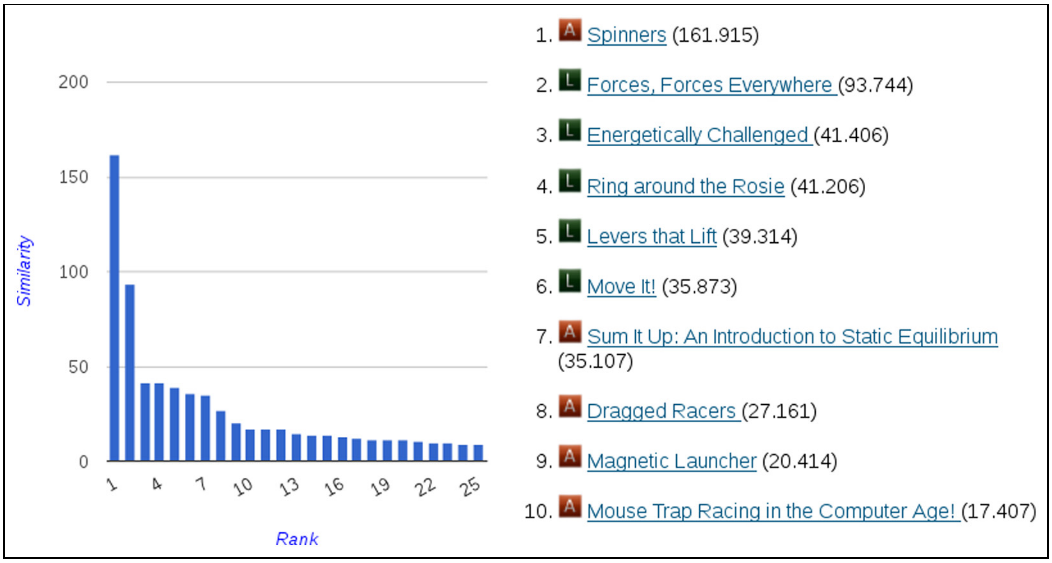

Figure 6. List-based interface.

Table 2. Dates, locations, participation, and conditions of the various experiment sessions.

Date (yyyy-mm-dd)	Location	No. of subjects	Condition
2014-09-11	Boulder, CO	7	Map
2014-11-07	Logan, UT	1	Map
2014-11-24	Corvallis, OR	4	List
2015-04-28	Portland, OR	3	List
2015-04-30	Portland, OR	3	List
2015-05-28	Portland, OR	3	Map
2015-06-08	Boulder, CO	12	List
2015-06-08	Boulder, CO	13	Map
2015-06-14	Seattle, WA	5	Map
2015-07-09	Boulder, CO	6	List
2015-08-17	Boulder, CO	6	Map

consider them as validated by the Reitsma et al.[55] work and hence, chose to reuse them for this experiment.

Subjects were provided with only one of the two versions of the TeachEngineering system. The "map" version included the spatializations discussed earlier. The maps were embedded as Google Maps in the Web version of each of the curricular items. Both page-embedded and full-screen versions of the maps were available. In contrast, the "list" version did not include any maps, but instead embedded a simple list of the 10 most similar items in each curricular item, complemented with a bar chart of the rank-ordered similarities of the 25 most similar items (Figure 6). Although both the list and the bar chart contain the same ordinal information, the bar chart offers a way to estimate the differences in similarity between curricular items. The items displayed in both the map and list versions consisted of Web links. Clicking these links would take subjects to a rendering of the content of the associated items.

We recruited 65 professional science teachers and/or science teachers in training—trainees had to be actively teaching in science classrooms—in the Portland and Corvallis, OR; Logan, UT; and Boulder, CO areas. Five additional teachers participated in the experiment at a TeachEngineering workshop during the 2015 ASEE Annual Conference in Seattle, WA. All subjects participating in a session would run the identical condition. Determination of whether to run a map or list session was entirely determined by the existing counts of subjects per condition (different numbers of subjects volunteered at different locations and on different dates). This resulted in more or less equal numbers of subjects for both conditions and essentially random allocation of subjects to conditions. Dates, locations, participation, and conditions of the various sessions are listed in Table 2.

Procedure and experimental task

Although almost all participants were familiar with the production version of the TeachEngineering system (which in its current state contains neither the map nor the list interfaces), each experimental session was preceded by a 30-minute training during which participants were familiarized with using the system and searching the collection. Participants were trained only in the condition to which they were subjected in the experiment, either list or map, but not both. The training consisted of a short demonstration followed by a series of written challenge questions that subjects were asked to answer while navigating the system. Answers to the questions were provided so that subjects could check their work.

Following the training, participants were asked to assume a random identity and to register for the experiment so that all their interactions with the system could be recorded. Participants were then given 45 minutes to select teaching assignments—written on index cards and placed on a table in the center of the experimental facility—and to search TeachEngineering for curricular items which they deemed to support the assignments. For each item deemed supportive, subjects were asked to score a short series of 7-point Likert scales, most of which were once again taken from Reitsma et al.[55] Two of the scales, however, were new and addressed subjects' (perceived) utility of the similarity tool and the role of item similarity in their searches:

- *I found this document with the help of the semantic similarity tool.*
- *I selected this document because its content is similar to that of another document I selected.*

When finished with a teaching assignment—no specific criteria for finishing an assignment were provided; the decision to move on to a new assignment was entirely left to the participants—participants would score the following two scales relating to the assignment:

- *I am confident that I found all the documents that support this teaching task.*
- *The similarity tool helped me to find documents which support teaching this standard.*

Participants would then put their assignment card back on the table, select another one, and search the collection again. Participants were paid $25 in cash for their participation in the experiment.

Reflection on procedure

In this section, we reflect on some of the experiment design and procedural choices:

- We executed a traditional between-group design in that each subject was exposed to and trained in only one of the two interface conditions, either list or map. We are aware that some experimenters prefer within-group designs where subjects are exposed to all experimental conditions. However, given a time requirement of training and experimenting with just one condition of over 1 hour, and given that our n (65) is large enough to conduct adequate significance testing and that subjects were essentially randomly assigned to a condition, we considered a between-group design adequate.
- The 45-min duration for the search task was chosen to represent a balance between providing teachers sufficient time to search for items given several teaching assignments and keeping the overall task fresh and interesting. Finding curricular materials which support science teaching assignments is not easy. The work of Reitsma et al.[55] indicates that standard alignment is a context-sensitive and multidimensional concept for which there are no right/wrong answers. Whether teachers can meaningfully use a curricular document to support their teaching assignment or not is up to each individual teacher and can be quite different across teachers because of differences in educational contexts (Diekema and Olsen[57]). Making that decision requires real effort by teachers exploring those documents in their search for content which is both adequate from a content perspective and which fits their teaching style. The average length of TeachEngineering documents is 2047 words and much of this content is technical or detailed in nature. As we will see, the mean number of items retrieved by teachers during the 45-minute time span was only 5.5 (list condition) and 5.74 (map). Ignoring some overhead time for selecting and reporting on teaching assignments, this indicates an average search and decision time per document of over 8 min.
- Whereas we exposed list subjects to the top ten most similar items, we seeded the maps with only the five most similar ones, supplemented with all the items from the curricular units to which these five most similar items belonged. Although we are quite convinced that this, in effect, pulled into the map more than just the five most similar items, it is possible (although unlikely) that this difference

Table 3. Total and average click-throughs by experimental condition.

Actions	List	Map	t-Test
Number of click-throughs	216	111	
Number of click-throughs per teaching assignment	3.72	1.37	$p = 0.001$
Number of click-throughs per subject	7.71	3.17	$p = 0.001$
Number of click-throughs per subject (subjects who did not click-through at all excluded)	7.71	4.26	$p = 0.001$

might have had an effect on the experimental outcomes.

- We opted to allow participants to freely select teaching assignments from a pool of 11 such assignments which covered a broad area of topics, grades, and coverage in TeachEngineering. This allowed participants to work on assignments with grade levels and topics with which they were most comfortable. Although ideally all subjects conduct the identical task across modalities, limiting the number of the available assignments would have made it very much harder to recruit a sufficient number of participants and would have biased the experiment toward just a few topics and grades. Across both modalities, all teaching assignments were conducted, just not by everyone.

Participant data

In all, 65 participants were recruited for the experiment. Of these, 52 (80%) were identified as professional science teachers. 11 (20%) were identified as science teachers in training with classroom teaching experience. (Two of the 65 subjects were removed from the subject pool since they did not complete even a single teaching assignment.) The average length of teaching experience was 11.58 years. Among the participants, all K–12, grades, were covered. The average teaching grade—measured as the midpoint between the lowest and highest grades participants taught—was 7.29. The average reported teaching grade band width was 2.29 grades.

Results

Table 3 shows item click-through counts, both totals and averages per teaching assignment and subject, for each of the two experimental conditions. As mentioned, both the list and map interfaces allowed participants to click-through to the items in which they were interested and both the map and list trainings covered the click-throughs. A total of 327 such click-throughs were recorded. Whereas only 26 of the 35 map subjects (74%) used item click-throughs, all 28 (100%) of the list subjects did.

The click-through data indicate that list subjects had a significantly higher propensity to click-through than map subjects for both the number of click-throughs per subject and per teaching assignment. This result is contrary to what we would expect if maps provide a superior browsing and searching opportunity. We expect that although both the lists and the maps contained the items as click-throughs, and despite the fact that clicking-through was covered in both the list and map training, the maps still represented a significantly higher barrier to direct interaction than the more familiar list format.

Comparing the distributions of number of click-throughs for the two conditions (Figure 7), we observe that they are quite different. Whereas the map distribution is bi-modal, mostly low frequency, and with only two or three high-frequency subjects, the list distribution is much smoother and approaches normality when we ignore its three most active click-through participants. What this suggests is that the map condition is indeed a more difficult condition which generates utility only for those who succeed to unlock its value. This interpretation would agree with that of Borgman's[72] observation that "Novices tend to rely on the most basic features …, and rarely take advantage of sophisticated search refinement capabilities." Yet, even for the three map subjects who clicked-through most, the average score on the "The similarity tool helped with this task" scale is quite low (2.85), that is, between "agree" and "somewhat agree."

In an attempt to uncover the source of this relative low usage of the maps, we investigated the trajectory of all so-called map actions, that is, clicking-through, panning, zooming, and item summary retrieval, of all map participants (Figure 8). We observe the following:

- Most map participants accessed (used) the maps throughout the experiment, although a few, for instance, participants 6, 10, and 22, try it for a while and then stop using it.
- Map use is clustered in time. This was expected—one might even say that it should be this way—as in between usage of the map, participants spend

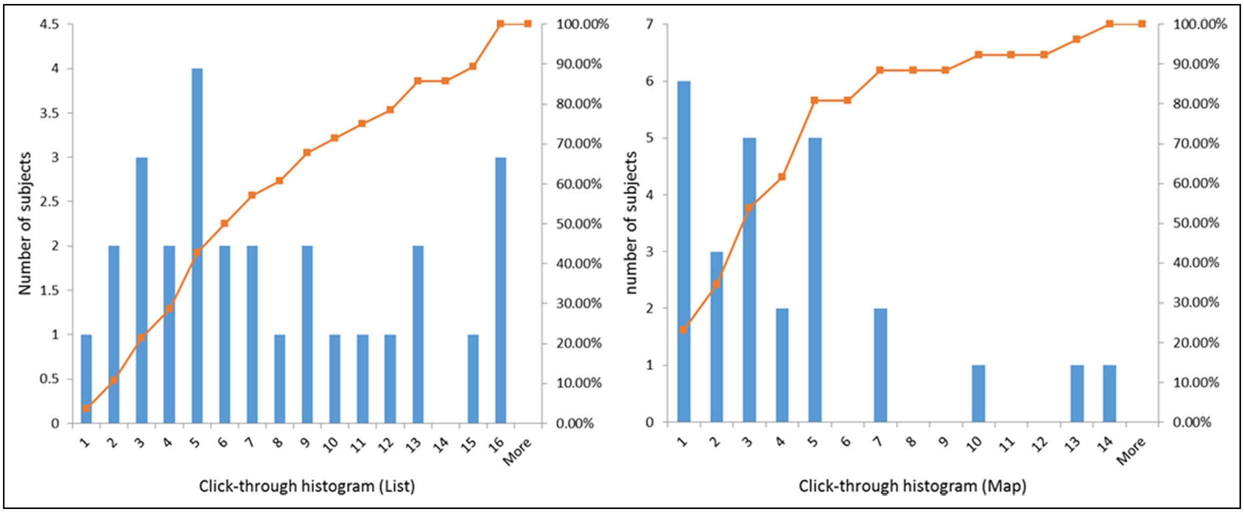

Figure 7. Distribution of click-throughs by subjects for both list and map conditions.

Table 4. Curricular items retrieved by condition.

	List	Map	t-Test
Number of participants	28	35	
Number of curricular items	154	201	
Minimum number of curricular items retrieved per participant	2	1	
Maximum number of curricular items retrieved per participant	11	11	
Average number of curricular items retrieved per participant	5.50	5.74	$p > 0.10$
Average number of curricular items retrieved per participant (single-item teaching assignments excluded)	129 Documents/ 23 subjects = 5.61	177 Documents/ 30 subjects = 5.90	$p > 0.10$
Average number of curricular items retrieved per teaching assignment (single-item assignments excluded)	129 Documents/ 37 tasks = 3.48	177 Documents/ 58 tasks = 3.02	$p = 0.07$

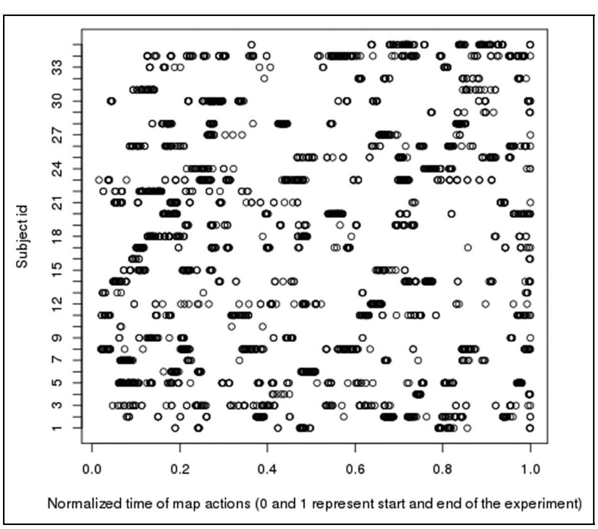

Figure 8. Time distribution of subjects' map actions (each map action is represented by a circle).

time reading and exploring a library item (the blank periods in the record).

- Some participants, such as participants 4, 25, 29, and 35, only seem to "warm up" to the map interface after some time or even quite late in the experiment.

Overall, we detect no systematic map dismissal or "refusal to use" patterns. Table 4 provides data on the number of curricular items found for each of the experimental conditions and hence, tests Hypothesis 1. We observe that the average number of curricular items retrieved for the map condition is higher than that of the list condition, but that this difference is not statistically significant.

Since the expected utility from similarity-assisted searching goes to 0 when a curricular item is unique, that is, no other documents are quite like it, we repeated the comparison excluding all teaching

Table 5. Average Likert scale scores.

Scale	List	Map	t-Test
Confidence in having exhausted the search (avg.)	3.04	3.28	$p > 0.10$
Similarity tool helped finding the item (avg.)	2.17	2.30	$p > 0.10$
Similarity tool helped finding the item (excl. single-item teaching assignments) (avg.)	2.06	1.98	$p > 0.10$

assignments which retrieved but a single curricular item. Again, no significant difference was found.

When comparing the retrieved curricular items per assignment rather than per subject under exclusion of assignments which resulted in the retrieval of only a single item, the difference is near significant but opposite expectations, that is, list-based searches result, on average, in more items returned than map-based searches. These results do not support Hypothesis 1.

Although teachers were free to decide when to stop searching for curricular items for a given teaching assignment—we implemented no controls on the searches other than providing 11 assignments to select from and limiting total search time to 45 minutes, we compared the number of searches conducted in each of the two conditions. Whereas the mean number of searches per subject in the map condition was higher than that in the list condition (2.31 versus 2.07), the difference was not statistically significant ($p > 0.10$).

Turning our attention to the attitudinal data and hence Hypotheses 2 and 3, Table 5 compares the results of the Likert scales between the two experimental conditions. Once again, no significant differences were found.

Table 6 explores some of the relationships which might exist between these scales. The only interesting difference found here is that between confidence in having exhausted the search and whether the similarity tool was considered helpful or not. For both the list and map conditions the relationship is opposite expectations in that list participants felt more confident that they found all relevant items.

On the curricular item level, however, the situation is different. First, we must realize that since the similarity tools only generate value once at least one item is found, the first-found item should always score very low on the "Similarity tool helped" and "Used the similarity tool to find this item" scales. Since our experiment administration did not allow us to retrace which curricular item was returned first for each teaching assignment, we instead (re)computed the average values for these scales after dropping the lowest item-specific value for all assignments which retrieved more than one item. Although this (obviously) improved the scores, once again no significant difference was found between the two conditions. Hypotheses 2 and 3, therefore, were not found to be supported by our data.

Discussion

We did not uncover any advantage of the maps over the lists. Although most participants used the maps throughout the experiment period and across teaching assignments, no advantage in terms of number of curricular items found, searches completed, confidence in having adequately explored the search space, or perceived utility of the map interface in finding standard-aligned curriculum was found. Several of the map participants did comment on the "neatness," "elegance," "novelty," and "cleverness" of the map interface. Although this is in line with the many novel and esthetically attractive spatializations that exist these days, we had hoped to find more tangible benefits. In this section, we offer some thoughts on why we did not find the expected benefits and where, if they exist, these benefits might reside.

- One possible reason for lack of success of the map view is that participants are firmly ingrained in traditional search engine paradigms where the results are generally provided in list-like formats. A single training session with the map interface to a document space might not have been enough for our participants to recognize the value of this still relatively new way of searching.
- Although we have good evidence that the maps properly represent the structure of the collection, being able to read and navigate a document space in terms of similarity distances and similarity neighborhood is likely to be more difficult than picking the top three items from a list of most similar items. Pleasing and intriguing as information spatializations may be, actually reading them and working with them might be quite difficult.
- One possible explanation for the observed lack of benefits of the spatialization is that the difference in information content between a list of similar items and a 2D map computed from those similarities is negligible, and that hence, the map cannot be expected to offer benefits over a list. However, this argument would put any and all spatializations in question. One could, of course, conduct an experiment where one condition is the map condition and the other is a condition with neither map nor similarity list. Indeed, some of the work to

Table 6. Correlations between attitudinal scales and number of curricular items found.

Correlation	List	Map
Similarity tool helped—confidence	0.39 ($p < 0.01$; $R^2 = 0.15$)	0.07
Similarity tool helped—curricular items found	0.11	−0.20
Similarity tool helped—curricular items found (for teaching assignments with curricular items found > 1)	−0.04	−0.10
Confidence—curricular items found	0.08	0.07

which we refer in the introductory sections of this article follows this approach. That approach, however, would spuriously attribute any observed benefits to the spatialization whereas these benefits might have been derived from the similarities.

- Perhaps we just designed a poor visualization. Maybe a differently laid-out one, using different colors, a different map projection, or perhaps even an extra dimension, would have made all the difference? Although it is quite possible that our choices for visualization were unfortunate, our underlying methods (other than the exclusive use of bigrams for the computation of item similarity) are all well known and well documented and have been used, in some variation or other, before. Also, our test of whether the Sammon maps properly reproduce the structure of the collection returned very good results, and not a single map participant communicated, formally or informally, that the maps were difficult to comprehend or navigate.
- Perhaps, spatializations such as the one presented here and the ones discussed earlier offer benefits other than improved searching and item retrieval. One can think of the discovery of spatial patterns indicating relationships in the data which may otherwise remain undetected. Other benefits might be educational and illustrative, that is, a map as an illustration of certain relationships and groupings of data. Exploring this line of inquiry suggests the use of explanatory theoretical frameworks such as task-technology fit (TTF) and technology acceptance modeling (TAM).

Conclusion

We designed a bigram-based similarity measure over the TeachEngineering corpus and used it to spatialize the library's collection with 2D Sammon maps. The resulting maps reproduce the hierarchical structure of the collection well although the structure information was not part of the similarity computations. Next, we conducted a between-group experiment in which real-world library users were asked to search the library and were provided with either the Sammon maps or an interface which simply listed the items most similar to the items they were displaying.

We compared the performance of the map and the list interfaces in terms of the numbers of items found by the participants, the participants' confidence that they had adequately explored the search space, and the utility of the interfaces as perceived by the participants and found no significant differences.

These results indicate that these spatializations, interesting, intriguing, pleasing, and novel as some of them are, might not be as useful as their creators, us among them, might have hoped.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

1. Skupin A. Spatialization. In: Kemp K (ed.) *Encyclopedia of geographic information science*. Thousand Oaks, CA: SAGE, 2007, pp. 418–422.
2. Boyack KW, Newman D, Duhon RD, et al. Clustering more than two million biomedical publications: comparing the accuracies of nine text-based similarity approaches. *PLoS ONE* 2011; 6: e18029.
3. Small H and Griffith BC. The structure of scientific literature I: identifying and graphing specialties. *Sci Stud* 1974; 4: 17–40.
4. Small H and Garfield E. The geography of science: disciplinary and national mappings. *J Inform Sci* 1985; 11: 147–159.
5. Everett JE and Pecotich A. A combined loglinear/MDS model for mapping journals by citation analysis. *J Assoc Inform Sci Technol* 1991; 42: 405–413.
6. Chalmers M. Using a landscape metaphor to represent a corpus of documents. In: Frank AU and Campari I (eds) *Spatial information theory: a theoretical basis for GIS* (lecture notes in computer science). Berlin: Springer-Verlag, 1993, pp. 377–390.
7. Wise JA, Thomas JJ, Pennock K, et al. Visualizing the non-visual: spatial analysis and interaction with information from text documents. In: *Proceedings of the information visualization (INFOVIS '95)*, Atlanta, GA, 30–31 October 1995, pp. 51–58. New York: IEEE.

8. Chen H, Schuffels C and Orwig R. Internet categorization and search: a self-organizing approach. *J Vis Commun Image R* 1995; 7: 88–102.
9. Chen H, Houston AL, Sewell R, et al. Internet browsing and searching: user evaluations of category map and concept space techniques. *J Assoc Inform Sci Technol* 1998; 49: 582–603.
10. Small H. Update on science mapping: creating large document spaces. *Scientometrics* 1997; 38: 275–293.
11. Noyons ECM and Van Raan AFJ. Advanced mapping of science and technology. *Scientometrics* 1998; 41: 61–67.
12. Chen C. Visualizing semantic spaces and author co-citation networks in digital libraries. *Inf Process Manag* 1999; 35: 401–420.
13. Chen C (ed.) *Information visualisation and virtual environments*. London: Springer Verlag, 1999.
14. Devaul H, Diekema A and Ostwald J. Computer-assisted assignment of educational standards using natural language processing. *J Assoc Inform Sci Technol* 2011; 62: 395–405.
15. Dodge M and Kitchin R. *Atlas of cyberspace*. Addison-Wesley, 2001.
16. Geroimenko V and Chen C (eds) *Visualizing the semantic web: XML-based internet and information visualization*. London: Springer-Verlag, 2006.
17. Wang M. A significant improvement to clever algorithm in hyperlinked environment. In: *Proceedings of the 11th international conference on the world wide web (WWW2002)*, Honolulu, HI, 7–11 May 2002.
18. Boyack KW, Wylie BN and Davidson GS. Domain visualization using VxInsight for science and technology management. *J Assoc Inform Sci Technol* 2002; 53: 764–774.
19. Buzydlovky JW, White H and Lin X. Term co-occurrence analysis as an interface for digital libraries. In: Börner K and Chen C (eds) *Visual interfaces to digital libraries*. Berlin, Heidelberg: Springer, 2002, pp. 133–143.
20. Skupin A. A picture from a thousand words. *Comput Sci Eng* 2004; 6: 84–88.
21. Wang C, Reese JP, Zhang H, et al. Similarity-based visualization of large image collections. *Inform Visual* 2015; 14: 183–203.
22. Van Ham F and Van Wijk J. Interactive visualization of small world graphs. In: *Proceedings of the IEEE symposium on information visualization*, Austin, TX, 10–12 October 2004, pp. 199–206. New York: IEEE.
23. Perer A and Shneiderman B. Balancing systematic and flexible exploration of social networks. *IEEE T Vis Comput Gr* 2006; 12: 693–700.
24. Henry N and Fekete J. MatrixExplorer: a dual-representation system to explore social networks. *IEEE T Vis Comput Gr* 2006; 12: 677–684.
25. Leydesdorff L. *Environment and Planning B: Planning and Design* as a journal: the interdisciplinarity of its environment and the citation impact. *Environ Plann B* 2007; 34: 826–838.
26. Burns R and Skupin A. Visualization of attribute spaces involving places, people and utterances. In: *Proceedings of the 24th international cartographic conference*, Santiago de Chile, 15–21 November 2009.
27. Zhu B, Watts S and Chen H. Visualizing social network concepts. *Decis Support Syst* 2010; 49: 151–161.
28. Lieberman MD, Taheri S, Guo H, et al. Visual exploration across biomedical databases. *IEEE/ACM T Comput B* 2011; 8: 536–550.
29. Dörk M, Carpendale S and Williamson C. Visualizing explicit and implicit relations of complex information spaces. *Inform Visual* 2012; 11: 5–21.
30. Börner K. Making sense of mankind's scholarly knowledge and expertise: collecting, interlinking, and organizing what we know and different approaches to mapping (network) science. *Environ Plann B* 2007; 34: 808–825.
31. Skupin A, Biberstine J and Börner K. Visualizing the topical structure of the medical sciences: a self-organizing map approach. *PLoS ONE* 2013; 8: e58779.
32. Burns R and Skupin A. Towards qualitative geovisual analytics: a case study involving places, people, and mediated experience. *Cartographica* 2013; 48: 157–176.
33. Chen VY, Razip AM, Ko S, et al. Multi-aspect visual analytics on large-scale high-dimensional cyber security data. *Inform Visual* 2015; 14: 62–75.
34. Liu X, Shen H and Hu Y. Supporting multifaceted viewing of word clouds with focus + context display. *Inform Visual* 2015; 14: 168–180.
35. Lima M. *Visual complexity: mapping patterns of information*. New York: Princeton Architectural Press, 2011.
36. Fabrikant SI, Montello DP and Mark DM. The natural landscape metaphor in information visualization: the role of commonsense geomorphology. *J Assoc Inform Sci Technol* 2009; 61: 233–270.
37. Sorrows ME and Hirtle SE. The nature of landmarks for real and electronic spaces. In: Freksa C and Mark DM (eds) *Spatial information theory: cognitive and computational foundations of geographic information science*. Berlin, Heidelberg: Springer, 1999, pp. 37–50.
38. Blades M. Wayfinding theory and research: the need for a new approach. In: Mark DM and Frank AU (eds) *Cognitive and linguistic aspects of geographic space* (NATO ASI Series). Berlin, Heidelberg: Springer, vol. 63, 1991, pp. 137–165.
39. Mukherjea S and Hara Y. Focus + context views of world-wide web nodes. In: *Proceedings of the hypertext 97*, Southampton, 6–11 April 1997, pp. 187–196. New York: ACM Press.
40. Baudisch P, Good N, Bellotti V, et al. Keeping things in context: a comparative evaluation of focus plus context screens, overviews, and zooming. In: *Proceedings of the ACM CHI conference on human factors in computing systems*, Minneapolis, MN, 20–25 April 2002, pp. 259–266. New York: ACM.
41. Cockburn A, Karlson A and Bederson BB. A review of overview + detail, zooming, and focus + context interfaces. *ACM Comput Surv* 2007; 41(2): 1–2.31.
42. Gracia A, González S, Robles V, et al. New insights into the suitability of the third dimension for visualizing multivariate/multidimensional data: a study based on loss of quality quantification. *Inform Visual* 2016; 15: 3–30.

43. Nielsen J. 2D is better than 3D, 1998, https://www.nngroup.com/articles/2d-is-better-than-3d/
44. Sebrechts MM, Cugini JV, Vasilakis J, et al. Visualization of search results: a comparative evaluation of text, 2D, and 3D interfaces. In: *Proceedings of the 22nd annual international ACM SIGIR conference on research and development in information retrieval*, Berkeley, CA, 15–19 August 1999, pp. 3–10. New York: ACM.
45. Tory M, Kirkpatrick AE, Atkins MS, et al. Visualization task performance with 2D, 3D, and combination displays. *IEEE T Vis Comput Gr* 2005; 12: 2–13.
46. Westerman SJ, Collins J and Cribbin T. Browsing a document collection represented in two- and three-dimensional virtual information space. *Int J Hum-Comput St* 2005; 6: 713–736.
47. Ting S, Tan T, West G, et al. Quantitative assessment of 2D versus 3D visualization modalities. In: *Proceedings of the visual communications and image processing*, Tainan, Taiwan, 6–9 November 2011, pp. 1–4. New York: IEEE.
48. Johansson J, Forsell C and Cooper M. On the usability of three-dimensional display in parallel coordinates: evaluating the efficiency of identifying two-dimensional relationships. *Inform Visual* 2014; 13: 29–41.
49. Ebinger S and Skupin A. Comparing different forms of interactivity in the visualization of spatio-temporal data. In: *Kartographische Nachrichten*, 2007, vol. 57, http://www.kartographische-nachrichten.de/fileadmin/user_upload/Kartographische_Nachrichten/KN_Index_2007.pdf
50. Borkin M, Vo AA, Bylinksi Z, et al. What makes a visualization memorable? *IEEE T Vis Comput Gr* 2013; 19: 2306–2315.
51. Buzydlovky JW, Lin X, Zhang M, et al. A comparison of a hierarchical tree to an associative map interface for the selection of classification terms. In: *Proceedings of the ASIST 2013*, vol. 50, Montreal, QC, Canada, 1–5 November 2013
52. Office of Creative Research (OCR). Relativity's landscape: visualizing the impact of Einstein's general relativity theory, https://medium.com/memo-random/relativity-s-landscape-422d4f7ed20e#.68nis1jmo
53. Marshall B and Reitsma R. World vs. method: educational standard formulation impacts document retrieval. In: *Proceedings of the JCDL'11 2011*, Ottawa, ON, Canada, 13–17 June 2011. New York: ACM.
53. Baeza-Yates R and Ribeiro-Neto B. *Modern information retrieval*. Addison-Wesley, 1999.
54. Marshall B, Reitsma R and Samson C. Unraveling K-12 standard alignment: report on a new attempt. In: *Proceedings of the JCDL'16*, Newark, NJ, 19–23 June 2016. New York: ACM.
55. Reitsma R, Marshall B and Zarske M. Aspects of "Relevance" in the alignment of curriculum with educational standards. *Inform Process Manag* 2010; 46: 362–376.
56. Wetzler P, Bethard S, Leary H, et al. Characterizing and predicting the multifaceted nature of quality in educational web resources. *ACM Trans Inter Intell Syst* 2013; 3: 15–25.
57. Diekema AR and Olsen MW. Teacher personal information management (PIM) practices: finding, keeping, and re-finding information. *J Assoc Inform Sci Technol* 2014; 65: 2261–2277.
58. Kelly D. Methods for evaluating interactive information retrieval systems with users. *Found Trends Inform Retr* 2009; 3: 1–224.
58. Sullivan J, Cyr M, Mooney M, et al. The TeachEngineering digital library: engineering comes alive for K-12 youth. In: *Proceedings of the ASEE annual conference 2005*, Portland, OR, 12–15 June 2005. Washington, DC: ASEE.
59. Zia L. The NSF National Science, Technology, Engineering and Mathematics Education Digital Library (NSDL) program. *D-Lib Magazine*, 2004, vol. 10, number 3.
60. Neumann AW. *Recommender systems for information providers*. Heidelberg: Physica Verlag, 2009.
61. Santos O and Boticario J (eds) *Educational recommender systems and technologies: practices and challenges*. Hershey, PA: Information Science Reference, 2012.
62. Tan C, Wang Y and Lee C. The use of bigrams to enhance text categorization. *Inform Process Manag* 2002; 38: 529–546.
63. Bekkerman R and Allan J. *Using bigrams in text categorization*. Technical report IR-408 2003. Amherst, MA: Center of Intelligent Information Retrieval, UMass Amherst.
64. Manning CD, Raghavan P and Schütze H. *Introduction to information retrieval*. New York: Cambridge University Press, 2008.
65. Reitsma R, Hsieh PH and Robson R. Estimation and visualization of digital library content similarities. In: *Proceedings of the international conference of information systems 2015*, Fort Worth, TX, 13–16 December 2015.
66. Sammon JW. A nonlinear mapping for data structure analysis. *IEEE T Comput* 1969; 18: 401–409.
67. Sun J, Fyfe C and Crowe M. Extending Sammon mapping with Bregman divergences. *Inform Sciences* 2011; 187: 72–92.
68. Google. Google Maps JavaScript API, 2015, https://developers.google.com/maps/documentation/javascript/
69. Okabe A, Boots B, Sugihara K, et al. *Spatial tesselations: concepts and applications of Voronoi diagrams*. Chichester: John Wiley & Sons Ltd, 1999.
70. Borlund PA. Study of the use of simulated work task situations in interactive information retrieval evaluations. *J Doc* 2016; 72: 394–413.
71. Sumner T, Ahmad F, Bhushan S, et al. Linking learning goals and educational resources through interactive concept map visualizations. *Int J Digit Libr* 2005; 5: 18–24.
72. Borgman CL. Designing digital libraries for usability. In: Peterson Bishop A, Van House NA and Buttenfield BP (eds) *Digital library use: social practice in design and evaluation*. Cambridge, MA: The MIT Press, 2003.
73. Börner K and Chen C (eds) *Visual interfaces to digital libraries*. Berlin, Heidelberg: Springer-Verlag, 2002.
74. Dodge M and Kitchin R. *Mapping cyberspace*. London: Routledge, 2000.