Estimation and Visualization of Digital Library Content Similarities

Completed Research Paper

René Reitsma
College of Business, Oregon State Univ.
Corvallis, OR
reitsmar@bus.oregonstate.edu

Hsieh, Ping-Hung
College of Business, Oregon State Univ.
Corvallis, OR
Ping-Hung.Hsieh@bus.oregonstate.edu

Robby Robson
Eduworks Corporation
Corvallis, OR
robby.robson@eduworks.com

Abstract

We present a semantic similarity-based recommender service. Our experimental application and validation domain consists of K-12 engineering learning resources. Given a learning resource, we must determine which educational standards it addresses and vice versa, find resources that align with a given standard. One approach to this problem suggests transitively inferring standard alignment from the semantic similarity of other, previously aligned resources. We investigate a bigram-based similarity estimator and a Sammon map-based user interface for visualizing the resulting similarity space. Validation was performed using resources in TeachEngineering.org, a K-12 STEM digital library. Target classifications were derived from author-generated tables of content for these resources. Testing shows good performance of the similarity measure, both in its correspondence to the collection's table of contents and in the form of a two-dimensional Sammon map. The results provide evidence for the feasibility and practicality of using automated similarity measures in standards alignment and similar problems.

Keywords: Natural language processing, Content analysis, Data mining, Interface design, Visualization, Web services, Human-Computer Interaction

Introduction

During the last ten years or so, a variety of on-line repositories of K-12 learning resources have been developed (Zia, 2004; Reitsma *et al*., 2012). K-12 teachers use these resources to develop or improve their teaching with new and interesting lesson materials. It is increasingly important for both teachers and curriculum providers that these lesson materials be 'aligned' with educational standards or learning outcomes. This is a difficult and continuous challenge as there are many different standards, standard sets change regularly, new standard sets such as the US Common Core or Next Generation Science Standards (NGSS) have been and are being developed, and the supply of K-12 learning objects is growing quickly. Manually aligning content with standards requires the engagement of experts which is time consuming, expensive and often lacks consistency.

Several recent papers have explored the ability of machine-based classifiers to support this ongoing alignment task. They show that aligning learning resources to standards is a difficult and somewhat intractable problem (Devaul *et al*., 2011; Reitsma & Diekema, 2011, Marshall & Reitsma, 2011). In an

assessment of the Content Alignment Tool (CAT) machine-based classifier (Diekema & Chen, 2005; Diekema *et al.*, 2007; Yilmazel *et al.*, 2007, 2007A), Devaul *et al.* (2011) found poor inter-rater reliability (IRR) among experts and between CAT results and experts. They report that on average only 32% of the alignments suggested by CAT were shared by two experts, with agreement higher (40%) for standards addressing more general subjects and lower (18%) for more precise ones. Furthermore, Reitsma & Diekema (2011) found that the differences between CAT and human alignments systematically deviate, especially in the area of standards relating to science inquiry and methodology.

One possible reason for these alignment difficulties is that clues to the nature of an educational resource can come from its use, its relationship to a specific text or from other. To study this, Reitsma *et al.* (2010) measured IRR for multiple dimensions of alignment, as suggested by Saracevic (2007, 2007a) and Schamber *et al.* (1994). Two of the dimensions, 'concepts' (the resource includes concepts, keywords, terms and definitions from the standard) and 'background' (the resource provides background related to the standard), addressed semantic alignment. The other dimensions involved issues such as motivation, the availability of examples, ease of use in the classroom, and the availability of references. Wetzler *et al.* (2013) found similar results. Together, the semantic dimensions explained 70% of teacher opinions concerning alignment. Although other dimensions must eventually be considered, these results justify focusing first on the semantic relationship between a resource and an educational standard.

Measuring semantic similarity between a resource and a standard has its own challenges. Systems such as CAT do this by determining and comparing the concepts as expressed in the text of the resource and the standard. This may be called 'direct semantic comparison.' Direct comparison has inherent limitations because education standards are typically highly contextualized short sections of text and because distinct standards can involve different aspects of identical or similar concepts that are signaled by the use of even a single term, *e.g.*, 'communication' *vs.* 'application' or 'calibrate a model' *vs.* 'gravity model.' An alternative approach which we call 'indirect or transitive semantic comparison' is to measure the semantic similarity between yet unaligned resources and resources whose alignments have been vetted and to offer these similarities to users so that they can explore the (semantic) neighborhood of the vetted resources.

To implement this transitive method, we propose a five-step process:

1. Preparation and clean up of learning resources so that non-essential content which can thwart similarity estimation is removed. This is especially relevant for Web-based resources which often display substantial amounts of extraneous information such as advertisements, off-site links, terms of use and similar data.

2. Indexing the learning resources, which involves linguistically parsing the resources' texts.

3. Computation and collection of estimates of similarities between resources.

4. Computation of a spatial configuration of learning resources which optimally represents the similarities computed in the previous step.

5. Visualization of this configuration in an easy-to-navigate, easy-to-acquire and easy-to-deploy display medium.

Viewed more broadly, our problem and suggested solution are an instance of a so-called 'content-based recommender service;' *i.e.*, a recommender service which suggests new and additional resources based on their relationships with an already selected resource. For an overview of these systems we refer to Neumann (2009) and Santos & Boticario (2012).

We explore this process using learning resources from the www.teachengineering.org digital library (Sullivan *et al.*, 2005). This growing collection of 1,300+ K-12 STEM curricular resources was developed by 36 US universities as part of their engineering outreach and learning programs. The library served 2.44M unique visitors in the period July, 2014 – August, 2015.

Machine Estimation of Semantic Resource Similarity

No machine technique for estimating curricular document similarity can be expected to be perfect; not in the least because the notion of 'curricular similarity' itself is problematic (Reitsma *et al.* 2010). In empirical studies such as those conducted by Li *et al.* (2008), algorithms for text classification typically achieve at most 85%–90% accuracy, and the 'No Free Lunch Theorem' by Wolpert and Macready (1997)

offers that methods that work well for one class of problem will likely perform poorly on others. Previous work has not yet established how effective readily available semantic similarity tools are in clustering educational resources into standards-relevant topical groups. Providers of digital educational collections would benefit from a semantic similarity tool that provides adequate precision; *i.e.*, retrieved results should be correct, does not require pre-computation or training to be applied to a new domain, is fast and scalable and can be implemented in a Web 2.0 environment.

Our candidate framework for this purpose is a measure developed at Eduworks Corporation which has been used in digital library contexts such as the NSDL (Zia, 2004). The version used for the work presented here measures the coincidence of bigrams; *i.e.*, two-word terms which co-occur in documents (Tan *et al.*, 2002; Bekkerman & Allan, 2003), weighted by the root mean square of their inverse document frequencies (IDF). More precisely, for any document D, define $\beta(D)$ to be the set of bigrams in D. Let $S = \{S_1, \cdots, S_n\}$ be a corpus of documents to which a new 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, *i.e.*, 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 semantic similarity μ between D and S_i is defined as N times the root mean square of the inverse document frequencies of the shared bigrams; *i.e.*, as

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

where $M = |\beta(D)|$ is the number of bigrams in D. We note that a more standard measure counts the number of times that b appears in S_i; *i.e.*, the term frequency (Manning *et al.*, 2008), whereas the measure in (1) just looks at whether or not it appears. 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 completely insensitive to size since a larger document has more terms and therefore a larger chance of a random hit, but 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 computing our measure we do not examine every bigram in the text of the documents as we ignore common words—so-called 'stop words'—and only consider those bigrams in which each term is either a noun, a verb or an adjective.

This measure lends itself to computationally rapid implementations. The instance used was implemented by constructing a B-tree which for each $b \in B$ efficiently finds the set of $S_i \in S$ that contain b. This enables computation in time linear in the length of D. The measure was exposed as a JSON Web service. In tests, variants of this measure and of measures based on more sophisticated semantic analysis such as Latent Semantic Analysis (LSA) did not perform significantly better in the general case and are significantly slower.

The reader might wonder why bigrams are used instead of single terms or n-grams for *n>2*. Heuristically, bigrams can be considered to be the smallest unit of language that does not require sense disambiguation to detect semantic similarity and the largest unit that can consistently detect similarity when it is present. Single terms suffer from too many multiple meanings, leading to false positive matches. For example, using a single (non-stop-word) term example: "A *tree* is a connected graph without loops" matches well with "No trees grew on the loop that connected their homes," even though there is little similarity in the subject matter of the two sentences. Bigrams can have ambiguous meanings and different meanings in different contexts, but not to the extent that single terms do. Trigrams (and *n*-grams), on the other hand, reduce the density of matches, leading to false negatives. For example, the sentences "He had a white cell phone with a black plastic case" and "She had a black cell phone with a clear plastic case" would not match at all using trigrams, whereas bigrams detect "cell phone" and "plastic case."

Learning Resource Preparation and Indexing

Web-based learning resources are likely to contain and/or display substantial amounts of information unrelated to their core (semantic) contents. Examples of these are author names, publication dates and

copyright statements, but also entire segments of text which may only tangentially relate. For instance, many TeachEngineering resources contain instructions for how to conduct an experiment or application meant to illustrate the science and engineering principles covered by the resource. The semantic content of these instructions, however, may have little relationship with those principles, For instance, a lesson on bridge building may contain a detailed description of an exercise involving matches, spaghetti or lollypop sticks, yet the relationships between this description and bridge building principles is entirely implicit and unlikely to be picked up by comparing co-occurrences of terms. To prevent a term-based similarity estimator from associating such a lesson with another entirely unrelated lesson that happens to also refer to matches or spaghetti, we must redact these types of sections and information from the resource prior to computing semantic similarity.

Similarly, to minimize estimation bias, we must remove all items that could explicitly associate a resource with other resources even though there may be no underlying semantic similarity. For instance, each resource in TeachEngineering belongs to a table of content (TOC) group and displays all other members of that group. Leaving these in place could easily bias a similarity estimator in favor of similarity between the resources within a group. Since all TeachEngineering resources are stored in XML, removal of nonessential items such as these was easily automated.

Assessment of Similarity Estimation Quality

The methods described above can be used to index and compute semantic similarity within any collection of textual learning resources. For this paper, we were interested in assessing how well the similarities compared to similarity derived from the explicit structure of a resource collection. After checking the internal integrity of the results, we compared computed similarity in TeachEngineering with a similarity metric derived from the TeachEngineering TOC groupings.

Internal integrity checking

The algorithm we used computes relative similarities; *i.e.*, similarity of a resource relative to all other resources in the corpus. There is no pre-determined range for the estimates, *i.e.*, the estimates are not normalized, and the estimates will generally be larger when the resource being compared is larger. The estimates are intended to be used ordinally; *i.e.*, we are generally interested in the top *n* estimates. The largest similarity for each document should be self-similarity—the document compared with itself—which was empirically found to be the case with self-similarity being roughly log-linear in the word count. Mean similarity scores for the top 5, 10, 25 and 100 most similar resources where 122.24, 82.43, 48.13 and 18.01 respectively. By way of contrast, we computed the similarities between the CNN home page—a random resource proxy—and all 958 documents in the corpus. The result was only 67 nonzero similarities with a mean value of only .43.

Comparison with TeachEngineering TOC

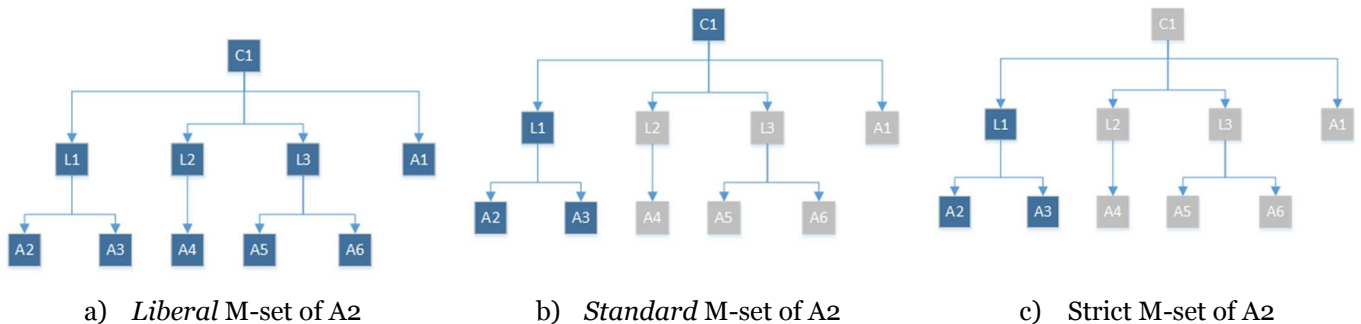a) *Liberal* M-set of A2   b) *Standard* M-set of A2   c) Strict M-set of A2

Figure 1. M-sets and their member resources (C: curricular units; L: lessons; A: activities). Blue: member. Gray: non member

TeachEngineering learning resources come in three types: hands-on activities, lessons and curricular units. The collection is hierarchically organized in that most activities are grouped under lessons and lessons are again grouped under curricular units. (Figure 1: A=activity; L=lesson, C=curricular unit). The lesson level, however, is not mandatory; *i.e.*, activities can 'live' on curricular units; e.g., A1 in Figure 1. We assumed that resources placed closer together in this hierarchy by their authors are more closely related. Since information on these groupings was not part of the similarity computations, the groupings could serve as an independent, expert-derived 'gold standard' against which computed similarities can be assessed. A drawback of this method was that we had to remove all learning resources that were not part of any TOC grouping; *i.e.*, all activities which were not grouped under lessons or curricular units. After removal of these resources, 958 learning resources—532 activities, 364 lessons and 62 curricular units— remained for validation. Given a resource *x*, we defined three match sets (or 'M-sets') of resources related to *x*; each representing a different (and increasingly constrained) type of match. These are most easily described in terms of the three-level TOC structure:

- The *liberal* M-set of resource *x* consists of all resources *y* such that *x* and *y* have a common ancestor, including the ancestor itself (Figure 1a).
- The *standard* M-set of resource *x* consists of all resources *y* such that *x* and *y* have direct lineage to a common ancestor including the ancestor itself (Figure 1b).
- The *strict* M-set of resource *x* consists of all resources *y* such that *x* and *y* have a common parent including the parent (Figure 1c).

Histograms showing the size distribution of these M-sets are shown in Figure 2. *Liberal* M-sets contained as many as 71 resources but *standard* and *strict* M-sets, because of their constrained membership, were small, often containing fewer than five resources. Thus, the chance that a resource chosen at random is part of another resource's M-set is very small, especially using the *standard* or *strict* definition of relatedness.
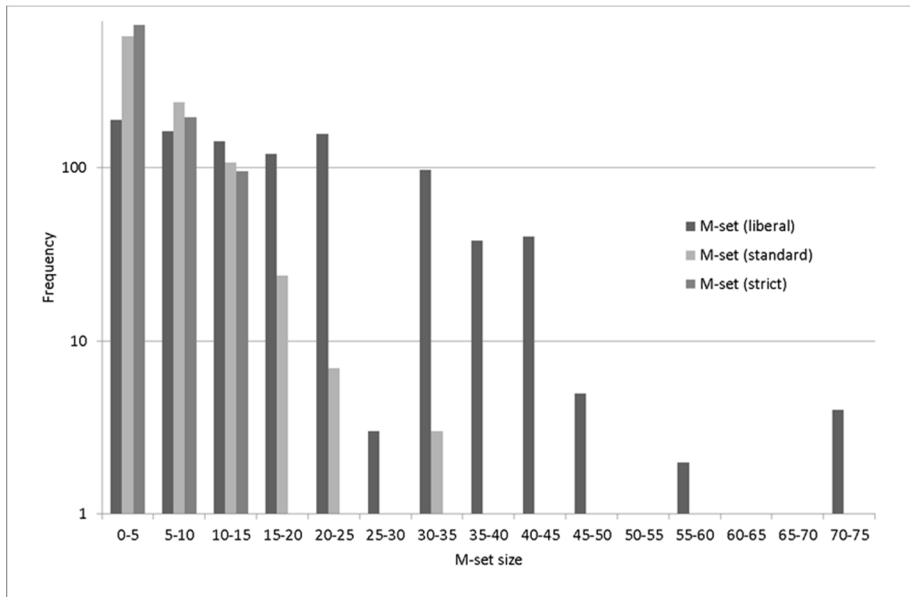

Figure 2. Histogram of TeachEngineering M-set sizes.

To compare the fit between the similarity estimates and M-set membership we computed the R-precision at all ranks of the estimator; *i.e.*, for each resource we computed how many of its top-ranked similarities were present in its M-sets for all its ranks. We then computed the mean precision at any given rank *k* by averaging the precision at that rank over all resources (precision is the fraction of correct retrievals over total number of retrievals). Mean precisions at all ranks are shown in Figure 3. Estimated precision as would be achieved by random association and its 1σ, 2σ and 3σ confidence intervals are provided for reference.

As is apparent from Figure 3, precision is very high at ranks 1 to 3 and significantly outperforms random association at all ranks. We thus conclude that if we use the Eduworks similarity estimator to find TeachEngineering resources semantically related to a given one, the top-ranked results are consistently in the given resource's M-set; *i.e.*, are among the small number of resources that according to the library's contributors are, in fact, related to the resource.
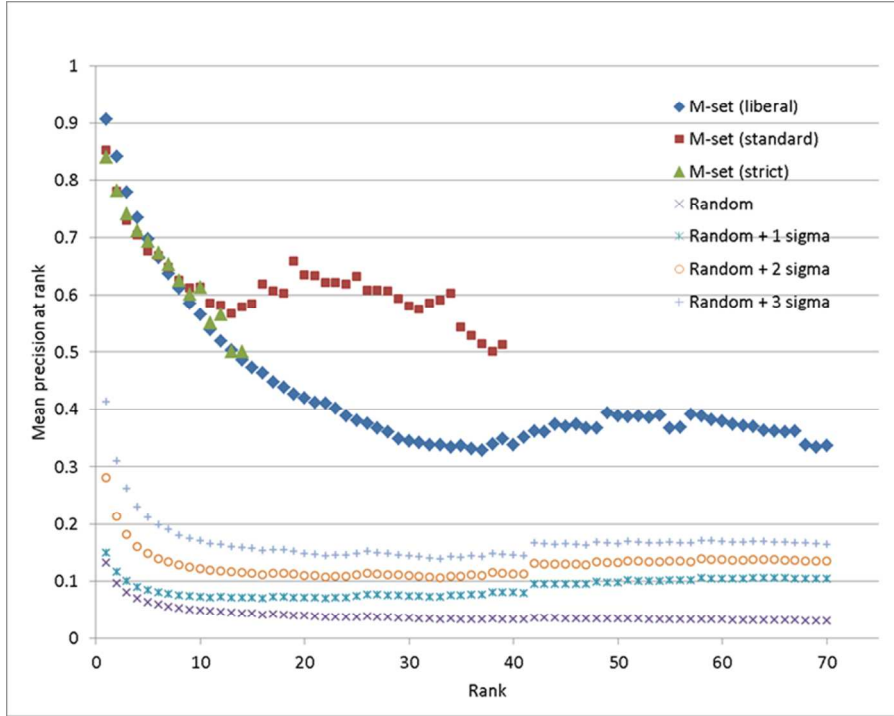

Figure 3. R-precision of the similarity estimator relative to the various M-sets.

It is important to note that these results likely underestimate the precision of the machine estimates; *i.e.*, the estimator is almost certainly better than the test results indicate. The reason for this is that in practice, authors contributing new resources to TeachEngineering often create a new curricular unit rather than distributing their materials over or adding them to already existing units. An example are two units on the topic of *simple machines*, developed independently by two different institutions. These *simple machine* resources are not part of each other's M-set even though they would be considered very much related if examined by an expert and, for that matter, by our similarity estimator. This falsely reduces our precision measures and hence, underestimates the performance of the estimator.

To be clear, and possibly ironically, we expect that the greatest value of the transitive approach as outlined here is not in being able to recreate a collection's table of contents from the similarities between its documents. On the contrary, as will be shown later, we would consider it a significant contribution being able to recommend to users resources which, although very similar to the ones they have already picked and vetted, have not been cataloged under the same TOC groupings. We believe that it is those cross-TOC similarities which can bring whole new sets of useful documents into users' views. Documents which, if they would only be guided by the collection's TOC or by simple team searches, might forever remain out of view.

From the above results we may conclude that the similarity estimator and its underlying measure produce scores with good correspondence, at least for educational resources in TeachEngineering. And since the 36 universities which contributed curriculum to TeachEngineering can be considered to broadly represent the US K-12 engineering education community, we may tentatively conclude that our estimator comprises a definite candidate for machine-based learning resource similarity estimation in this domain. High precision at top ranks is important for workflows in which teachers search for resources related to ones

already considered suitable or aligned with a standard, but where they can only invest in exploring a few of these resources because of the high (time) cost of such exploration.

We note that the similarity measure was effective even though it is not very complex. One reason for this is that the resources in TeachEngineering were both reasonably sized with a mean document size of 1,848 terms (after cleanup and removal of stop words) and reasonably homogeneous in their thematic content. These are favorable conditions for measures of semantic similarity matching, and we believe they are typical of many educational digital libraries. The other, perhaps more important reason, we hypothesize, is the relative high power of bigrams in domain-specific areas. As is true for any specialized field, STEM fields abound with two-word terms which, when considered separately would take on entirely different meanings.

Similarity Map of Learning Resources

Once we have reliable similarity estimates between all pairs of resources in the corpus, we want to use those similarities to help library patrons find and explore those related resources. Although we could simply offer patrons lists of the most similar resources—more or less in the Netflix/Amazon tradition of 'if you like this you might also like this'—we might also consider visualizing these resources in the context of their TOC groupings. This 'focus & context' principle (Mukherjea, & Hara, 1997) is used widely in information visualization—a comprehensive overview of 100+ applications is provided by Lima (2011)—and we believe that it applicable here as well.

Inspired by the Web-based nature of readily available tools such as the Google Maps API and available methods for data dimensionality reduction such as factor analysis and multi-dimensional scaling, we used the paired similarity estimates to spatially project the TeachEngineering resources so that they could be offered for interactive spatial browsing in tools such as Google Maps. As we did with our similarity estimates, we wanted to ensure that this derived spatial projection also properly represented the TOC groupings of resources as created by the resources' authors. This required satisfying two demands: 1) the spatial projection must properly reflect the similarity data; *i.e.*, highly similar resources must be located close to each other, and 2) the resource groupings as specified by the resources' authors must be spatially retained; *i.e.*, resources which are bundled by authors in discrete groups must be clustered in space as well. For the *simple machines* example mentioned earlier this implies that the resources present in both curricular units must be located close to each other on the map, yet the curricular units must still be spatially distinguishable from each other.

As our mapping method we chose—after testing a variety of alternative techniques—Sammon mapping. Sammon maps are a variant of multi-dimensional scaling in which we interpret the similarity between objects as the inverse of their distance and we try to place the objects (our documents) in an n-dimensional space so that their paired spatial distances maximally represent the inverse similarities (Sammon, 1969; Sun *et al.*, 2011).

Figure 4 shows a sample of 107 resources consisting of 11 curricular units (C) and their constituent lessons (L) and activities (A) projected into a 2D Sammon space computed from the similarities between them . Lines connecting resources indicate their membership of a TOC group. We note that the locations of the curricular unit resources (C) were not computed using the Sammon procedure. Because these resources tend to be small (few words), their associated similarity estimates are less reliable than those for lessons and activities. Hence, we computed their positions as the mean coordinates of the lessons and activities comprising them.
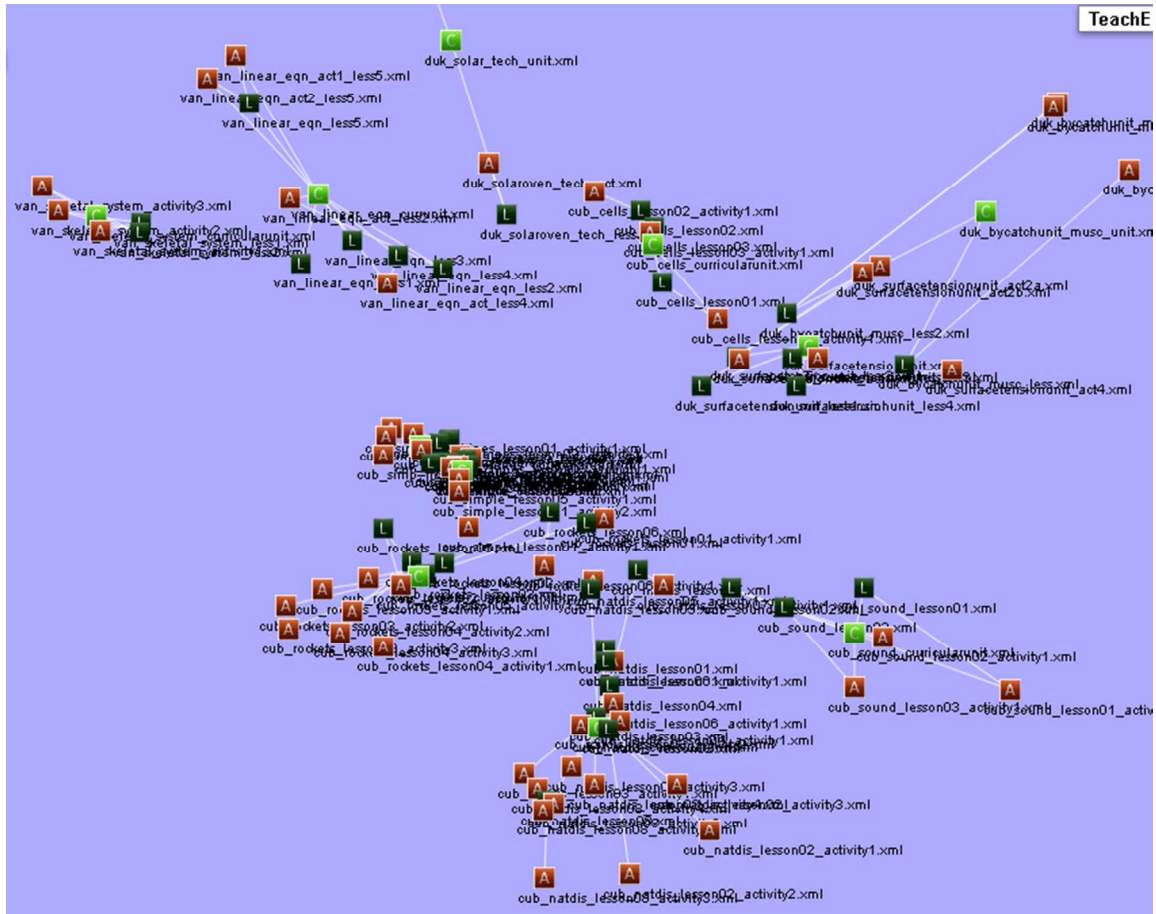
Figure 4. 2D Sammon map of 11 TeachEngineering curricular units (107 resources) rendered in Google Maps API.

We observe the following:

- The solution adequately separates resource groupings. TOC groups of resources render as clusters thereby spatially expressing that they are 'different' from each other.

- The two *simple machines* units are located close to each other (the tight bundle of resources near the center of Figure 4.) Yet when zooming in on them in the Google Maps API, they separate into two distinct groups (not shown here).

- Despite the spatial separation between resources of different groups, the similarities between some resources from those groups render them located close to each other. It is these proximities of resources from different TOC groups which we consider especially promising for users to explore.

  Figure 5 provides a good example of this. In this case the map contains the similarity neighborhood of a specifically selected document—the activity *Sugar Spill*—and the documents belonging to their TOC groups. We draw attention to the close correspondence between the spatial clustering of the documents and TOC groupings (connecting lines). In addition, and this is very good news, the combination of the similarity estimates and the Sammon map results in close proximity of the *Sugar Spill* document with other documents such as *Oil Spill* (the document immediately 'north' of *Sugar Spill*) which are not related by means of the TOC but which contain related terms such as *environmental pollution*, *cleanup* and *contamination*, *absorption* or *containment*.
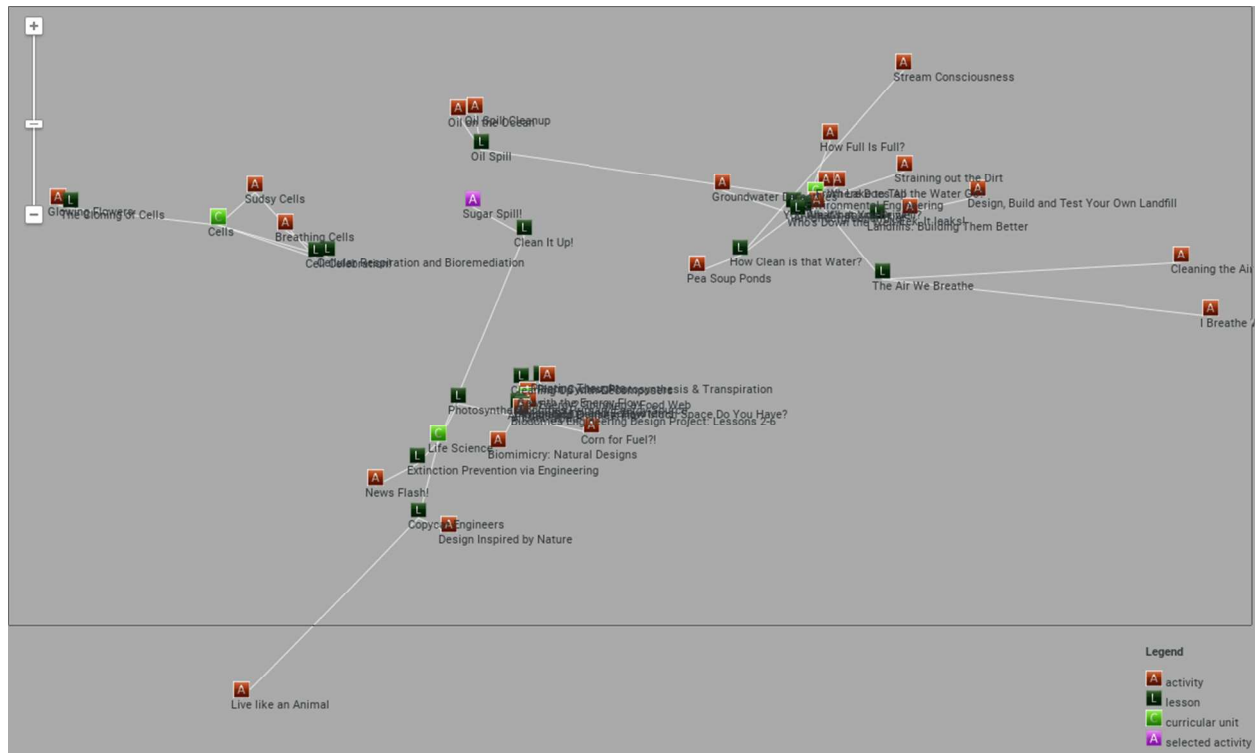
Figure 5. 2D Sammon map of the *Sugar Spill* activity and its semantic neighborhood.

Despite the visually good results of the Sammon mapping, we wanted to more formally test the correspondence of the resultant maps. Following the same approach as used in testing the similarity estimations, we defined a good result as the spatial distribution (clustering) of the documents corresponding with the TOC groupings of the resources. To quantitatively test for this we followed a two-step procedure:

1. Cluster the Sammon results in as many clusters as there are TOC groups, with each cluster containing the number of resources equal to the number of resources in the corresponding TOC group and each Sammon cluster centered 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 *vs.* TOC clusters. If all frequencies are on the diagonal of this table, indicating perfect Sammon-TOC cluster matching, Cramer's V will be 1.0 and χ^2 will be highly significant (p ≈ 0.0). As off-diagonals cells contain values, indicating cluster mismatches, V and χ^2 decline but can still be statistically significant.

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 (MILP) problem of the following kind:

Constants:

- d_{ij} = Sammon space distance between resource *i* and the center (mean x-y coordinates) of TOC group *j*.

- c_j = number of resources in TOC group *j*.

Decision variables:

- x_{ij} = 1 if resource *i* is part of TOC group *j*; 0 otherwise.

Objective function:

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

Constraints:

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

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

Solving this model for our 107 documents and 11 curricular unit sample problem, giving 107 × 11 = 1177 variables, and comparing the resultant clusters with the TOC group membership of the resources yields an almost perfect fit (χ^2 = 449.80, df = 100; p ≈ 0; V = .943).

To check for the overall dimensionality of the data we also computed a 3D Sammon solution. Solution stress; *i.e.*, a measure of how well the original similarities correspond to the distances on the map, remained essentially the same (2D: .787; 3D: .784). However, the 3D MILP cluster-by-cluster solution was not quite as good (nine off-diagonal cells *vs.* five off-diagonal cells in the 2D solution and a correspondingly lower χ^2). From these results we conclude that the 2D solution suffices.

To test the sensitivity of the MILP solution to the constraint that clusters must have the same number of resources as their corresponding TOC groups, we computed the MILP excluding that constraint; a problem equivalent to computing the space's ordinary Voronoi sets (a Voronoi set is a grouping of points in nD space where the group membership of a point is based on the minimum distance of that point to the spatial centers or 'generators' of all the available groups (Okabe *et al.*, 2000)). This resulted in nine off-diagonals (χ^2 = 431.60, df = 100; p ≈ 0; V = .914) indicating that the constraints do play a role, but that computed either with or without the constraint the Sammon clusters map the TOC groupings well.

Conclusion and Next Steps

From the above results we conclude that our methods of computing digital library resource similarity based on bigrams and visualizing these similarities by means of 2D Sammon mapping seem to work well. At least in our application domain of K-12 STEM and K-12 engineering they both generate results which closely correspond with the resource groupings as determined by the resource authors. This is a welcome result as the methods are quite practical, not very computationally intensive and fit well in the modern world of exposing Web-services and interactive mapping.

Now that we have validated our methods for generating similarity-based resource maps, we have started the human subjects phase of our project. In this phase we ask actual TeachEngineering users; *i.e.*, K-12 science and engineering teachers to search for resources which support several hypothetical teaching tasks. Some users are given a system implementation containing the Sammon maps whereas others are given a simple one-dimensional listing of the most similar resources. Some of the dependent variables we will measure are the speed with which searches progress, usage of the similarity data, the amount of resources found and users' confidence in having found the appropriate resources and all available useful resources.

Acknowledgments

References

Bekkerman, R., Allan, J. 2003. "Using Bigrams in Text Categorization," *Technical Report IR-408*, Center of Intelligent Information Retrieval, UMass Amherst. MA.

Devaul, H., Diekema, A., Ostwald, J. 2011. "Computer-Assisted Assignment of Educational Standards Using Natural Language Processing," *J. Am. Soc. of Info. Sc. and Techn* (62), pp. 395–405.

Diekema, A., Chen, J. 2005. "Experimenting with the Automatic Assignment of Educational Standards to Digital Library Content," *Proc. JCDL'05*. IEEE-ACM, New York.

Diekema, A., Yilmazel, O., Bailey, J., Harwell, S., Liddy E. 2007. "Standards Alignment for Metadata Assignment," *Proc. JCDL'07*. Vancouver, BC. IEEE-ACM. New York.

Li, T., Zhu, S. Ogihara, M. 2008. "Text Categorization via Generalized Discriminant Analysis," *Inf. Proc. and Mgt.* (44), pp. 1684-1697.

Lima, M. (2011) "*Visual Complexity. Mapping Patterns of Information.*" Princeton Architectural Press.

Manning, C.D., Raghavan, P., Schütze, H. 2008. "*Introduction to Information Retrieval.*" Cambridge Univ. Press, New York, NY.

Marshall, B., Reitsma. R. 2011. "World *vs.* Method: Educational Standard Formulation Impacts Document Retrieval," *Proc. JCDL'11*. IEEE-ACM. New York, N.Y.

Mukherjea, S., Hara, Y. 1997. "Focus + Context Views of World-Wide Web Nodes," *Proc. Hypertext 97*. ACM Press. pp. 187-196.

Neumann, A. W. 2009. "*Recommender Systems for Information Providers*." Physica Verlag. Heidelberg, Germany.

Okabe, A., Boots, B., Sugihara, K., Chiu, S.N. 2000. "*Spatial Tesselations. Concepts and Applications of Voronoi Diagrams.*" Second Edition. Wiley Series in Probability and Statistics. John Wiley & Sons, Chichester, England.

Reitsma, R., Marshall, B., Zarske, M. 2010. "Aspects of 'Relevance' in the Alignment of Curriculum with Educational Standards," *Inf. Proc. and Mgt.* (46), pp. 362-376.

Reitsma, R., Diekema, A. 2011. "Comparison of Human and Machine-based Educational Standard Assignment Networks," *Intern. J. on Dig. Libs.* (11), pp. 209-222.

Reitsma, R., Marshall, B., Chart, T. 2012. "Can Intermediary-based Science Standards Crosswalking Work? Some Evidence from Mining the Standard Alignment Tool (SAT)," *J. Am. Soc. of Info. Sc. and Techn.* (63), pp. 1843-1858.

Sammon, J.W. 1969. "A Nonlinear Mapping for Data Structure Analysis," *IEEE Transactions on Computers* (18), pp. 401–409.

Santos, O.C., Boticario, J.G (Eds.) 2012. "*Educational Recommender Systems and Technologies.*" Information Science Reference. Hersey, PA.

Saracevic, T. 2007. "Relevance: A Review of the Literature and a Framework for Thinking on the Notion in Information Science. Part II: Nature and Manifestations of Relevance," *J. Am. Soc. of Info. Sc. and Techn.* (58), pp. 1915-1933.

Saracevic, T. 2007a. "Relevance: A Review of the Literature and a Framework for Thinking on the Notion in Information Science. Part III: Behavior and Effects of Relevance," *J. Am. Soc. of Info. Sc. and Techn.* (58), pp. 2126-2144.

Schamber, L, Eisenberg, M.B., Nilan, M.S. 1994. "A Re-examination of Relevance: Toward a Dynamic, Situational Definition," *Inf. Proc. and Mgt.* (26), pp. 3-48.

Sullivan, J., Cyr, M., Mooney, M., Reitsma, R., Shaw, N., Zarske, M., Klenk, P. 2005. "The TeachEngineering Digital Library: Engineering Comes Alive for K-12 Youth," *Proc. ASEE Annual Conf.* ASEE, Washington D.C.

Sun, J, Fyfe, C., Crowe, M. 2011. "Extending Sammon Mapping with Bregman Divergences," *Inf. Sciences* (187), pp. 72-92.

Tan, C., Wang , Y., Lee, C. 2002. "The Use of Bigrams to Enhance Text Categorization," *Inf. Proc. and Mgt.* (38), pp. 529-546.

Wetzler, P., Bethard, S., Leary, H., Butcher, K., Bahreini, S.D., Zhao, J., Martin, J., Sumner, T. 2013. "Characterizing and Predicting the Multifaceted Nature of Quality in Educational Web Resources," *ACM Trans. on Interactive Intelligent Systems* (3).3.

Wolpert, D.H., Macready, W.G. 1997. "No Free Lunch Theorems for Optimization," *IEEE Trans. on Evolutionary Computation* (1), pp. 67-82.

Yilmazel, O., Balasubramanian, N., Harwell, S., Bailey, J., Diekema, A., Liddy, E. 2007. "Text Categorization for Aligning Educational Standards," *Proc. 40th HICSS*. IEEE. New York, NY.

Yilmazel, O., Ingersoll, G., Liddy, E. 2007A. "Finding Questions to Your Answers," *ICDE'07*. Istanbul, Turkey. Syracuse, NY: CNLP.

Zia, L. 2004. "The NSF National Science, Technology, Engineering and Mathematics Education Digital Library (NSDL) Program." *D-Lib Magazine*. (10), 3.