This is the authors' submitted version (accepted March 2012). When published, the authoritative version will be available on the JASIST website: http://www.asis.org/jasist.html .

# Can Intermediary-based Science Standards Crosswalking Work? Some Evidence from Mining the Standard Alignment Tool (SAT)

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

Byron Marshall
College of Business
Oregon State University
Corvallis, OR
byron.marshall@bus.oregonstate.edu

Trevor Chart
College of Engineering
Oregon State University
Corvallis, OR
chartt@onid.oregonstate.edu

## Abstract

We explore the feasibility of intermediary-based crosswalking and alignment of K-12 science education standards. With increasing availability of K-12 science, technology, engineering and mathematics (STEM) digital library content, alignment of that content with educational standards is a significant and continuous challenge. Whereas direct, one-to-one alignment of standards is preferable but currently unsustainable in its resource demands, less resource-intensive intermediary-based alignment offers an interesting alternative. But will it work? We present the results from an experiment in which the machine-based Standard Alignment Tool (SAT) —incorporated in the National Science Digital Library (NSDL)— was used to collect over half a million direct alignments between standards from different standard-authoring bodies. These were then used to compute intermediary-based alignments derived from the well-known AAAS Project 2061 Benchmarks and NSES standards. Results show strong variation among authoring bodies in their success to crosswalk with best results for those who modeled their standards on the intermediaries. Results furthermore show a strong inverse relationship between recall and precision when both intermediates where involved in the crosswalking.

## Acknowledgments

This work was supported in part by US National Science Foundation grant #0532709. We also acknowledge the Center for Natural Language Processing at Syracuse University for its support of the Standard Alignment Tool (SAT) and JES&Co. for its support of the Achievement Standards Network (ASN).

## Introduction: The Standard Alignment Challenge

Recently, we have seen a sharp increase in the availability of digital, web-based, K-12 curricular materials or learning resources. Table 1 contains a sample of reputable digital collections of K-12 curricular content and an item count of their holdings. A challenge confronting all these collections concerns the alignment of their curriculum with science education standards. In the USA this challenge stems mainly from US 2001 Public Law 107: "An act to close the achievement ga*p with accountability, flexibility, and choice, so that no child is left behind,"* more commonly known as the 'No Child Left Behind' Act [50]. Compliance with this act challenges content providers to align their collection's curriculum with the many thousands of K-12 educational standards throughout the USA. The standard alignment challenge, however, is not limited to the USA as in many other countries and regions, standards-based teaching and learning—we refer, for instance to Europe's Bologna process [8]—is quickly becoming the norm. The following indicate the magnitude of the alignment challenge:

- The Achievement Standard Network (ASN), a publicly available database of K-12 standards, lists about 60,000 STEM standards in the USA alone [21]. Moreover, USA states, on average, have reformulated their standards about every five years, forcing frequent realignment of the existing curriculum [20, 23, 45].

- Manual (re)alignment by collection maintainers implies a sheer insurmountable task while community-based attempts such as pioneered at curriki.org [30, 51] are in their infancy.
- Inter-rater reliability problems have been reported when comparing human with machine-based alignments [12, 35, 36].
- Whereas private, for-profit enterprises make alignments available, the cost of using their services are significant and the applied alignment process remains hidden.

| Collection | Item count | Estimation method |
|---|---|---|
| Applied Math and Science Education Repository -- AMSER (Richards, 2010) (http://www.amser.org) | 13,416 | Search query on www.nsdl.org |
| Compadre (Mason, 2006) http://www.compadre.org) | 6,226 | Search query on www.nsdl.org |
| Curriki (McAnear, 2007; Wallis and Steptoe, 2006) (http://www.curriki.org) | 23,249 | Search query on Website |
| Engineering is Elementary (Cunningham and Hester, 2007; Lachapelle and Cunningham, 2007) (http://www.mos.org/eie/) | 90 | Count of curricular units multiplied with declared lesson count per unit |
| HotChalk (Kubilinskiene and Dagiene, 2009) (http://hotchalk.com) | 4,000+ | Stated on Website |
| Lesson Planet (http://lessonplanet.com) | 350,000+ | Stated on Website |
| Middle School Portal (Saylor and Minton-Morris, 2006) (http://msteacher.org/) | 2,108 | Search query on www.nsdl.org |
| National Science Digital Library (Zia, 2002; 2005) (http://www.nsdl.org) | 79,000+ | Custom query to NSDL official |
| NetTrekker (Breen, 2008; Felix, 2004) (http://www.nettrekker.com) | 300,000+ | Promotional video on Website |
| Teachers' Domain (Blumenthal, 2003; Johnson, 2003) (http://www.teachersdomain.org) | 3,100 | Search query on www.nsdl.org |
| TeachersTryScience (http://teacherstryscience.org/) | 44 | Stated on Website |
| TeachEngineering (Sullivan *et al*., 2005) (http://www.teachengineering.org) | 1045 | Browse query on Website |

*Table 1: Sample of K-12 STEM digital libraries and their holdings (as per Sep. 2011).*

**Background: Standard Alignment as Information Retrieval**

Inspired by developments in natural language processing techniques on the one hand and the prohibitively high cost of manual (re)alignment of the quickly growing supply of on-line educational resources on the other, several research groups and authors continue to attempt to develop and improve machine-based alignment methods [12, 13, 14, 36]. These alignment approaches fall in two categories. In the 'direct alignment' approach machine classifiers are used to find matches between an educational standard and a resource. The work by Devaul *et al.* [12] and Reitsma and Diekema [36], for instance, assesses the performance of the CAT classifier as a direct alignment tool. Results indicate rather limited success with specific difficulties for the classifier when dealing with so-called 'method' or 'method of inquiry' standards. Reitsma *et al*. [35] make an attempt at unfolding the alignment concept itself. Following the multidimensional notion of 'relevance' advocated by Saracevic [39, 40, 41] and Schamber *et al.* [43], they model alignment as a composite of different aspects or attributes of a resource. They find that the semantic component as represented by thematic content explains up to 70% of alignment variation but they also find significant contributions by attributes such as overall resource quality and the availability of student-ready worksheets.

In the transitive or 'crosswalking' approach to alignment, resources and standards are aligned using other, existing alignments between other resources and a standard. Various types of crosswalking can be considered. In a resource crosswalk, (semantic) similarity between resources can be used to transitively associate a standard with a resource if that resource is similar to another document which was previously associated with that same standard. More common in the K-12 educational standard domain, however, is standard crosswalking. Here a resource and an educational standard are associated with each other, because that standard is semantically similar to another standard that has previously been associated with the resource.

Although many search, retrieval, and matching tasks can be approached with fully automated computational mechanisms such as SVM when these mechanisms are trained with collections of correct; *i.e.*, 'vetted' matches, many important tasks are not well suited to such treatment. The poor inter-rater reliability in educational standards alignment mentioned above [12, 35, 36], for instance, implies that it is difficult to establish such training sets. Instead, we believe that the multi-faceted character of the 'alignment' concept calls for a dynamic and multidimensional retrieval mechanism in which users take part by interactively adding information in real-time, thereby steering and affecting how different alignment aspects are accounted for in the relevance computations. This situation is not unique to the retrieval of educational materials. Combining multiple aspects of relevance in a single computation requires some form of semantic integration. Doan and Halevy [16], for instance, have suggested that semantic integration techniques are appropriate when an application uses structured data representations and when more than one of those representations apply. Melnik *et al.* [31] demonstrated that multifaceted networks of structural and semantic clues can be integrated into a single computational formalism. Of course, a rich literature and tradition on multi-attribute and information gain decision making, covering such diverse areas as consumer behavior and public policy making; *e.g.*, [5, 48, 49], attests to this view of information retrieval. The value of multifaceted networks in facilitating retrieval processes has been demonstrated in law enforcement [27], document summarization [10], and database schema matching [31], to name a few topical domains. However, for such techniques to be employed, relevance clues must be collected and organized for integration.

Multiple, network-driven relevance clues and heuristics can help address challenging retrieval tasks. In the educational standards alignment domain, carefully developed lexicons, *e.g.*, that of the Teacher's Domain digital library [47] designed to support matching to specific standards, have been the basis of important and useful tools. Such resources provide important clues but are often expected to support rather than supplant human efforts [12]. Educational resources being considered 'near' previously vetted items or being positioned interestingly in a network of linked resources can provide these clues [52, 53]. Further, being positioned in or near interesting and transitive network association patterns can provide usable relevance information [26, 27]. An approach which integrates multiple relevance clues as well as real-time user input may well be needed to increase the effectiveness of standards-based document retrieval systems. The importance flooding algorithm discussed in [27] is one such an approach. But for such approaches to be more widely deployed, underlying networks of relevance information need to be collected and validated.

### A Word on Standard Harmonization Initiatives

Partly driven by the complexity of the educational standard landscape in the USA, the so-called Common Core initiative [4, 32] is an attempt at harmonization of standards across USA states. At the time of this writing all but seven of the 50 USA states have adopted the new Common Cores in English and Mathematics (665 standards) and efforts at developing other common cores are underway. There is little doubt that standard harmonization efforts, be it USA Common Core, European Bologna [8] or others, will reduce the complexity and hence, the effort associated with the alignment process. Nevertheless, we offer that plenty of challenge will remain. To start with, wholesale adoption of future cores for subjects such as science and history possibly face a bigger challenge than cores such as for mathematics and language given the value ladeness of these subjects. But even with significant harmonization within countries such as the USA or between countries such as in the EU, alignment challenges remain. With the global availability of Internet content and the quickly expanding stores of Internet-based learning resources, we expect that standard alignment efforts will become global efforts and that we will see an increase in alignments across national boundaries. Second, not only must, when new harmonized standards are accepted, the existing curriculum be realigned, harmonized standard sets themselves will change again in the future, which again requires realignment. Finally, nongovernmental educational organizations such as AAAS [1, 2, 3], the (US) Academy of Sciences [33, 34] and the International Technology and Engineering Educators Association (ITEEA) [19] are—and for the foreseeable future will continue to be—engaged in the development of new and improved standards.

### Standard Crosswalking Efficiency

Whereas in direct standard alignment learning resources and educational standards are associated with each other, standard crosswalking implies transitive alignment of the following form:

*Premise 1:* *Standard P is similar to standard Q.*
*Premise 2:* *Learning resource X aligns with standard Q.*
------------------------------------------------------------------------
*Conclusion:* *Learning resource X aligns with standard P.*

Collecting similarities between standards, also known as standard 'crosswalks,' however, can be a very time-consuming activity, even when done by machines, because of the sheer number of possible combinations involved. At least three types of standard crosswalking data can be considered:

1. Align each standard of each standard-issuing body against all standards of all other standard bodies. This generates a lot of crosswalks, which, lacking a means to massively parallel process this problem, take a lot of time to collect. The required time—assuming sequential processing—for this method scales exponentially with the number of standard-issuing bodies.
2. Align each standard from each standard-issuing body against the standards of an intermediary standard-issuing body, followed by transitively aligning standards through this intermediary crosswalk. The required time for this method scales linearly with the number of standard-issuing bodies and the number of involved intermediaries.
3. Apply methods 1 or 2, but only to those standards that were directly aligned to the items in a collection. Unlike methods 1 and 2, the time requirements for this method depend on the number of direct alignments in a collection.

Figure 1 estimates the amounts of work required for methods 1 and 2 using the existing USA K-12 STEM standards as example The (approximate) amount of 60,000 STEM standards used for this computation was derived from the Achievement Standard Network (ASN) database [18, 46] of standards. The example further assumes the availability of a machine-based standard alignment tool such as SAT [13, 15] and an average of five seconds per alignment query. The total required times for both methods assume uninterrupted querying; *i.e.*, zero down time. The calculation shows a more than 95% reduction in query time when using an intermediary rather than a standard-to-standard alignment approach.

| |
|---|
| *Method 1: Align each standard from each standard-issuing body against all standards of all other standard-issuing bodies*<br>Number of STEM standards in the ASN: ≈60,000<br>Number of standard bodies: ≈50<br>Mean number of standards per issuing body: ≈1,200<br>Number of standard-issuing body combinations: 50(50 - 1) / 2 ≈ 1,225<br>Total number of required alignment queries: 1,200 * 1,225 ≈ 1,470,000<br>Mean time per query: ≈5 seconds.<br>Total required time: ≈85 days of querying |
| *Method 2: Align each standard from each standard-issuing body against the standards of one intermediary standard-issuing body*<br>Total number of required alignment queries: 50 * 1,200 ≈ 60,000 queries.<br>Total required time: ≈3.47 days |

*Figure 1: Comparison of direct with intermediary-based standard alignment.*

However, although transitive crosswalking—*Method 2*—sharply reduces the computing time compared with direct crosswalking, it introduces an additional transitive element in the alignment process. Whereas Marshall *et al*. [28] offer that even direct crosswalking might not be very successful given the often profound incommensurabilities between the standards of different standard-authoring bodies; *i.e.*, great differences in both wording and semantic content of the various standard sets, transitive, intermediary-based crosswalking introduces a second transitive component yielding the following train of transitive logic:

*Premise 1:* *Standard P aligns with intermediary standard Z.*
*Premise 2:* *Standard Q aligns with intermediary standard Z.*
---------------------------------------------------------------------------
*Conclusion 1 (Premise 3): Standard Q aligns with standard P.*
*Premise 4:* *Learning resource X aligns with standard Q.*
---------------------------------------------------------------------------
*Conclusion 2:* *Learning resource X aligns with standard P.*

Since the inter-rater reliability experiments mentioned earlier render alignments between standards and between standards and resources tentative at best, any additional transitive elements can be expected to further weaken the validity of the final document-standard conclusion (*Conclusion 2*).

Still, the expected computational gain of intermediary-based standard crosswalking is so significant when compared to direct crosswalking, that we decided to test the feasibility of this double transitive approach.

**Intermediaries: Which Ones and How Many?**

As mentioned earlier, the USA K-12 education standard landscape is quite complex. Not only do most states formulate their own standards—education is under the authority of the states rather than the federal government, states also have chosen a wide variety of standard formulations. One aspect of this variety is granularity; *i.e.*, the amount of detail covered by the standards. Whereas states such as Iowa and Alaska have very few science standards (27 and 38 respectively), states such as Illinois, Tennessee and Georgia have several thousand science standards (Figure 2).

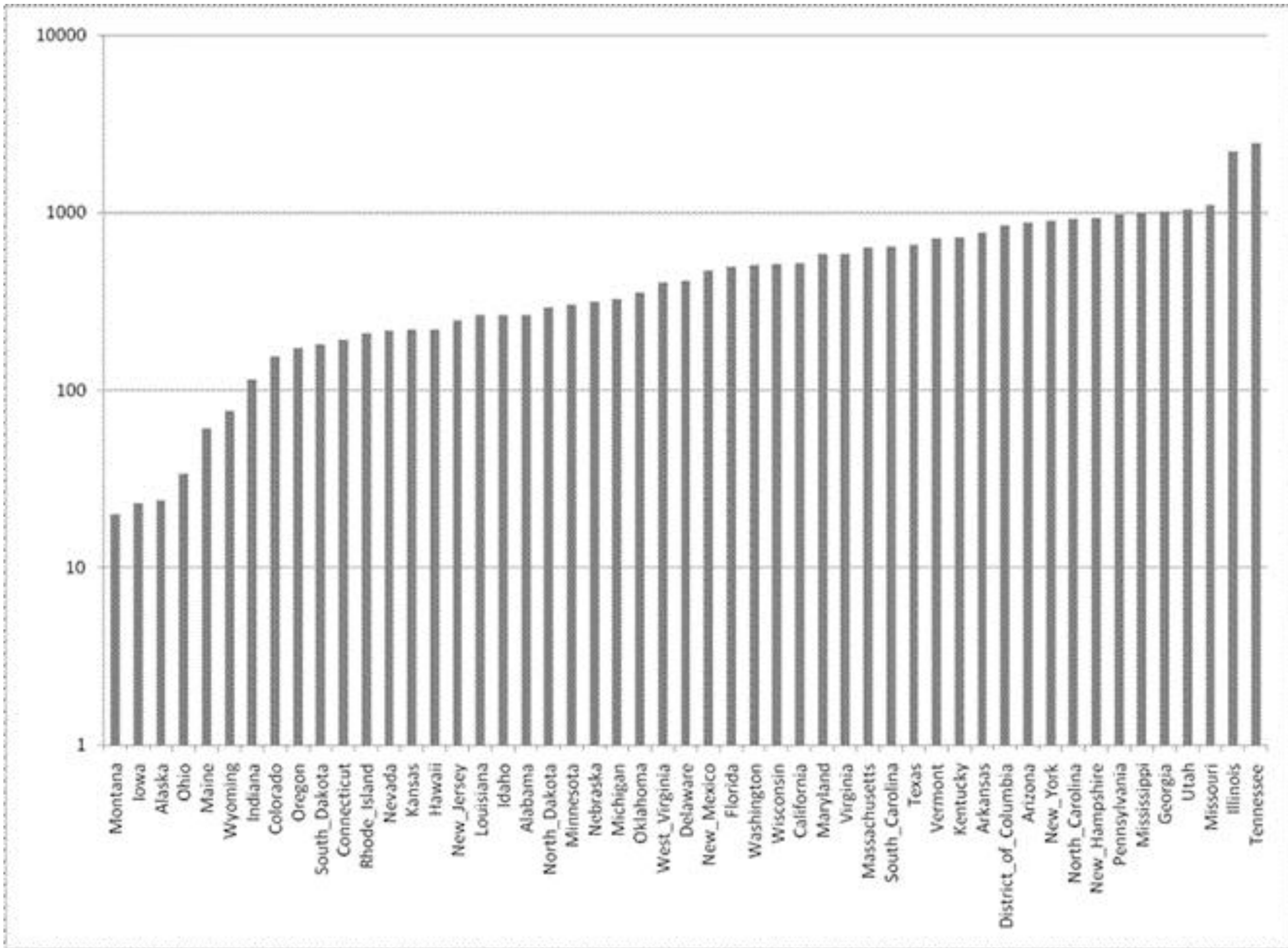


*Figure 2: US states and their number of science standards (logarithmic scale).*

As a consequence of this variability, one can theorize that the success of crosswalking from and to states will vary between states. States with very few, very generally-stated standards; *i.e.*, states with low standard granularity, will be more difficult to crosswalk than high granularity states with lots of very specific standards. In addition, different pedagogical or organizational paradigms have been employed by different standards-issuing bodies. For example, some states tend to segregate empirical and methodological standards into different parts of their standards structure while others integrate empirical and method concepts in individual standards items. Empirical or 'World' standards are those standards that refer to the empirical world, typically containing easy to recognize and specific terms. 'Method' standards, on the other hand, represent methodological principles and say essentially nothing about the empirical world other than how to study and understand it. Findings by Reitsma and Diekema [36] show that recall and precision achieved by machine-based direct alignment is significantly better for 'World' standards than for 'Method' standards. 'Method' standards tend to contain very few specific terms and hence, machine-based classifiers have much more trouble aligning method standards to learning objects.

In order to transitively crosswalk standards, we must select an intermediary standard set. Rather than selecting any state-based set of standards, we suggest using a(n) (inter)nationally recognized set such as the AAAS Project 2061 Benchmarks [1, 2, 3] and/or the National Science Education Standards (NSES) [33]. These standard sets were drafted by specialists that were not subject to the policy-making forces of a specific state and, as witnessed by many of the documents accompanying states' publication of their science standards, are widely recognized as being specifically targeted at furthering science education and training.

The availability of both the AAAS and NSES standard bodies introduces some interesting opportunities. Not only can we glean some insight into whether either of these bodies generates more or less successful transitive crosswalks, but we can also join them for two different types of crosswalk; *i.e.*, a crosswalk between state standards through either one or the other intermediary—we denote this approach as 'AAAS ∪ NSES'—or alternatively, to only consider state standards to be aligned if their crosswalk is established through both intermediaries; *i.e.*, 'AAAS ∩ NSES.'

### Alignment Data

Standard alignment data for this study were collected from two sources:

1. The Achievement Standards Network (ASN), published by the nonprofit JES&Co [18, 46] is an XML/JSON repository of all US K-12 education standards. Standards in the ASN are represented hierarchically; *i*.e., standards 'branch' into ever finer detail. Figure 3 shows a sample of Maryland science standards. Every standard in the ASN is identified by a permanent URL or PURL which can be resolved at http://purl.org/ASN/resources/<PURL>; *e.g.*, Maryland standard S102978A in Figure 3 resolves at http://purl.org/ASN/resources/S102978A. Standard updates in the ASN naturally lag the publication of those standards by the standard-issuing bodies. Also, since the various standard-issuing bodies use different schemas and formalisms to issue their standards whereas the ASN uses a single, hierarchical model, differences between the original standard formulation and the ASN representation may exist.
2. For conducting the automated standard alignments we used the Standard Alignment Tool (SAT), developed by the Center for Natural Language Processing (CNLP) at Syracuse University [13, 15] and hosted by the National Science Digital Library (NSDL) [56, 57]. SAT aligns standards based on a comparison between their textual content and those of the standards in the ASN. (Ordinal) alignment scores are the result of a weighted linear combination of fits produced by an information retrieval (IR) scorer, and, optionally, a machine-learning scorer. SAT's IR scorer compares the textual content of a standard with the standards in the ASN using CNLP's in-house natural language processing software TextTagger [54, 55]. TextTagger tokenizes the text, tags tokens with parts-of-speech, and identifies noun phrases, including proper nouns, based on part-of-speech patterns. The vector space model's TF*IDF formula [38] is used to score all of the standards' term vectors against the submitted standard's term vector, where the term vectors are comprised of the single-token and phrasal terms provided by TextTagger. SAT then assigns the highest ranking standard(s). Standards are returned in order of ranking without fit score information.

Alignment data were collected over a nine-month period, during which the SAT service was automatically and daily queried for periods of eight hours at a time. The procedure consisted of submitting each of the ≈60K STEM standards in the ASN to SAT, requesting it to align it with at most five standards from each of the standard–issuing bodies in its database. The limit of five was chosen for two reasons. First, since SAT does not provide a measure of fit other than an ordinal one, it does not essentially indicate quality of fit. Second, in-house testing by the developers of SAT has shown that fit rapidly declines beyond five matches. Hence, the SAT developers recommend setting the limit at five. This resulted in a total of 23M+ standard alignments.

Since our proposed intermediaries were AAAS and NSES, all alignments involving non-science standards were dropped from the data set. All alignments involving so-called ASN 'nonleaves' were also eliminated from the data set. This last statement requires some explanation. Referring to the Maryland Science 2005 example of Figure 3 again, only standards S102990E, S102990F, S102A106 and S102A10B—standards at the deepest level of nesting—are considered ASN 'leaves.'

This is the authors' submitted version (accepted March 2012). When published, the authoritative version will be available on the JASIST website: http://www.asis.org/jasist.html .

Maryland: Science (2005)

- (S102978A) 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
    - (S102989C) Diversity of Life
        - (S102990E) *Observe a variety of* familiar plants and animals to describe how they are alike and how they are different.
        - (S102990F) Observe a variety of familiar animals and plants (perhaps on the school grounds, in the neighborhood, and at home) to discover patterns of similarity and difference among them.
        - *Etc.*
    - (S10298A1) Evolution
        - (S102A106) Explain that individuals of the same kind differ in their characteristics, and sometimes the differences give individuals an advantage in surviving and reproducing.
        - (S102A10B) Describe ways in which organisms in one habitat differ from those in another habitat and consider how these differences help them survive and reproduce.
        - *Etc.*
- *Etc.*

*Figure 3: Single branch of Maryland life science standards and its associated ASN PURLs.*

Since SAT limits the notion of a standard to only the 'leaves' of the ASN standard trees, all alignments involving nonleaves were removed from the data set. This resulted in a data set of 4,790,801 alignments involving 28,795 science standards. These alignments can be subdivided into two groups:

- Bidirectional alignments; *i.e.*, SAT found standard X to align with standard Y and standard Y to align with standard X.
- All other (unidirectional) alignments.

Frequencies of both bidirectional and unidirectional alignments between science standards, including those involving AAAS and NSES standards, are listed in Table 2.

| | *full SAT total (bi- and unidirectional)* | *full SAT bidirectional only* |
|---|---|---|
| *Proportion of NSES standards aligned* | 310/310 (100%) | 310/310 (100%) |
| *NSES alignments* | 188,600 (3.94%) | 24,041 (4.46%) |
| *Proportion of AAAS standards aligned* | 854/855 (99.88%) | 852/855 (99.65%) |
| *AAAS alignments* | 269,062 (5.62)% | 37,152 (6.88%) |
| *Total alignments (science)* | 4,790,801 | 539,615 |

*Table 2: Bidirectional and unidirectional SAT alignments between science standards.*

From Table 2 it becomes clear that although both AAAS and NSES alignments occupy only very minor portions of the total alignment set, coverage of their standards in the SAT crosswalks is comprehensive; *i.e.*, essentially 100%.

Furthermore, the proportions of both NSES and AAAS crosswalks are higher in the bidirectional alignment set than in the unidirectional set. For AAAS crosswalks this proportional difference is statistically significant (chi-squared test); for the NSES portion it is not. This difference in proportion, together with the notion that a bidirectional SAT alignment can be interpreted as SAT's 'confirmed' alignment in both directions, made us limit our analysis of intermediary-based crosswalking to these 539K+ bidirectional SAT alignments.

**Error Sources and Validity Issues**

Before presenting the results of the intermediary-based crosswalk analysis, we care to point out four potential sources of error which may have manifested themselves in the data, mostly as a result of the long duration of data collection and the still experimental nature of both the SAT and the data collection tools:

1. Since US states change their educational standards frequently [20, 23, 45], changes to the standard bodies did take place during the data collection period. This implies that for some states, different standard sets have been used for alignment.
2. Some states have vary few standards or did not have standards until later in the data collection period; *e.g.*, Montana and Iowa. Since these states where not retroactively processed, some or even many of their potential alignments are absent from the data set.
3. Errors 1. and 2. may have been exacerbated because the ASN always lags the publication of standards and because absorption of ASN standards in SAT lags the publication of the ASN.
4. Due to inaccuracies in the data collection process, alignment data for some authors; *e.g.*, Colorado and Michigan are missing.

Although the data set would obviously be both more complete and more accurate if these errors had not occurred and data were not missing, some of these deficiencies must be considered difficult if not impossible to avoid. For instance, US states frequently and independently change their standards and some did only adopt standards late in the data collection process. Retroactively correcting for these problems would have added significant time to the data collection process and might even have been impossible as SAT, at least at the time of data collection, was reinitialized whenever new versions of the ASN were published.

Although some authors are missing from the data set, a sufficiently large data set remained to explore for significant and interesting patterns.

An obvious issue possibly limiting the validity of these data for evaluating the feasibility of intermediary-based standards crosswalking relates to the exclusive use of SAT. We realize that choosing SAT alignments as ground truth implies somewhat of a leap of faith. SAT is a tool which is still under development, which is still in the process of proving itself and whose results will most certainly be flawed. On the other hand, we know of no other, publicly available, inter-state standard alignment tool with such broad coverage. Moreover, since we have no reason to believe that SAT treats alignments to and from certain states differently than others, we assume that in the aggregate its errors can be considered random. Naturally, any correlation between systematic SAT errors and a corresponding bias of certain states' standards will have an effect on the alignments and crosswalks involving those states. Moreover, as pointed out by several of the previously cited authors, human inter-rater reliability relating to standard alignment have been found to be quite low. Hence, if we were using human alignments instead, we might be confronted with a similar bias problem. Finally, human alignment sets are rare, hard to come by and cannot be used in aggregated form since the conditions under which they were collected—task, score systems, level of training, *etc*.—differ substantially.

Given SAT's experimental status and likely flaws, however, we intuitively prefer the bidirectional alignments over the unidirectional ones because they represent two consistent alignments which SAT independently retrieved in opposite directions.

Overall, we consider the 539K+ dataset of SAT-based science standard alignment a sufficiently good dataset for exploring the feasibility of intermediary-based standard crosswalking.

### Author-specific Alignments

We start our analysis by exploring state-to-state or inter-author SAT alignments; both unidirectional (Figure 4) and bidirectional (Figure 5). Frequencies in both figures are the mean number of alignments per standard—recall that all SAT alignment requests had a constraint set of a maximum of five alignments per standards.

Figure 4 shows that both AAAS and NSES are universally good alignment generators (high row scores). This is a useful result because it implies that they are good candidates for intermediary-based crosswalking. They are followed by states such as Louisiana, Massachusetts, New Hampshire, Kansas and Pennsylvania. The state of Wyoming stands out as a weak 'from' state.

Matters look less promising on the 'receiving' end (columns). No single author stands out as a good alignment target. Alaska, Delaware and Georgia and to a lesser extent Kansas stand out as particularly hard to align to.

Figure 4 shows average bidirectional alignment frequencies. Note that the intra-state alignment frequencies (diagonal) are low except for AAAS and NSES. Whereas this may indicate the absence of significant standard redundancies within states, high intra-author alignments for AAAS and NSES may indicate extended coverage of specific topics or—as, for instance, explicitly mapped in the AAAS Atlas of Scientific Literacy [3]—the linkage between and across topics in different areas or sub categories of science.

Bidirectional alignment between AAAS and NSES is stronger than for any other combination of standard authors. NSES stands out with relatively high frequencies of bidirectional alignments across states. States that stand out as commonly good bidirectional aligners are Rhode Island and Nebraska. On the low end of bidirectional alignment frequencies we see Wyoming and Idaho.

The above results support our choice of using AAAS and NSES as crosswalking intermediaries. Not only are both bodies of standards generated and maintained by nonpolitical, impartial agencies, they also exhibit the best standard alignment frequencies.

This is the authors' submitted version (accepted March 2012). When published, the authoritative version will be available on the JASIST website: http://www.asis.org/jasist.html .

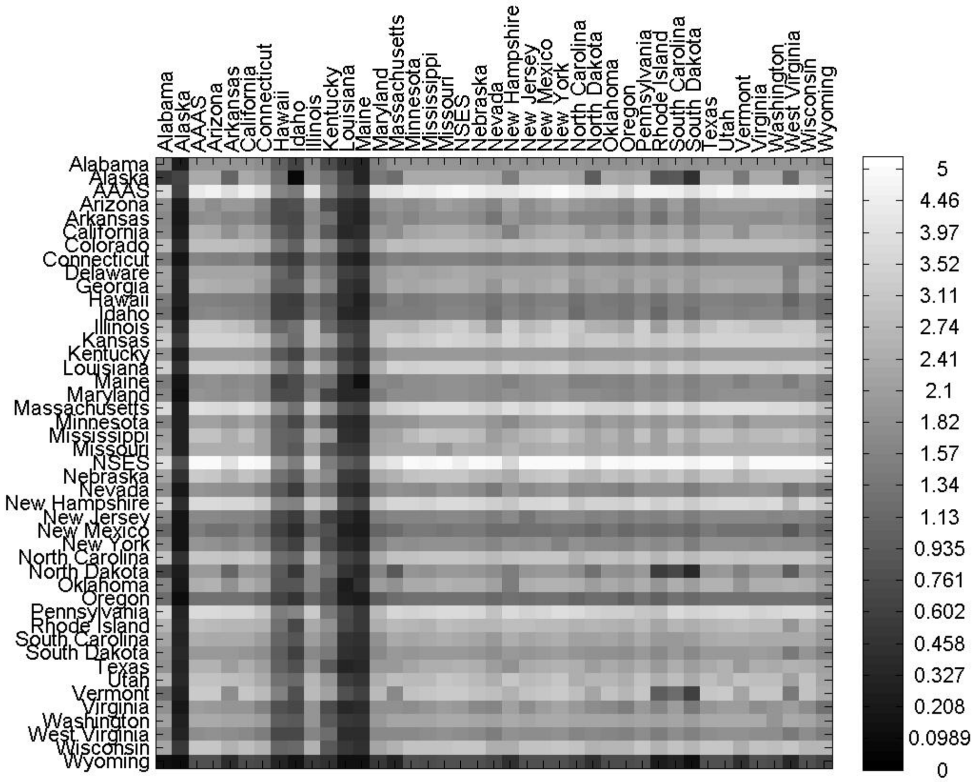

*Figure 4: Average number of alignments per standard by state.*

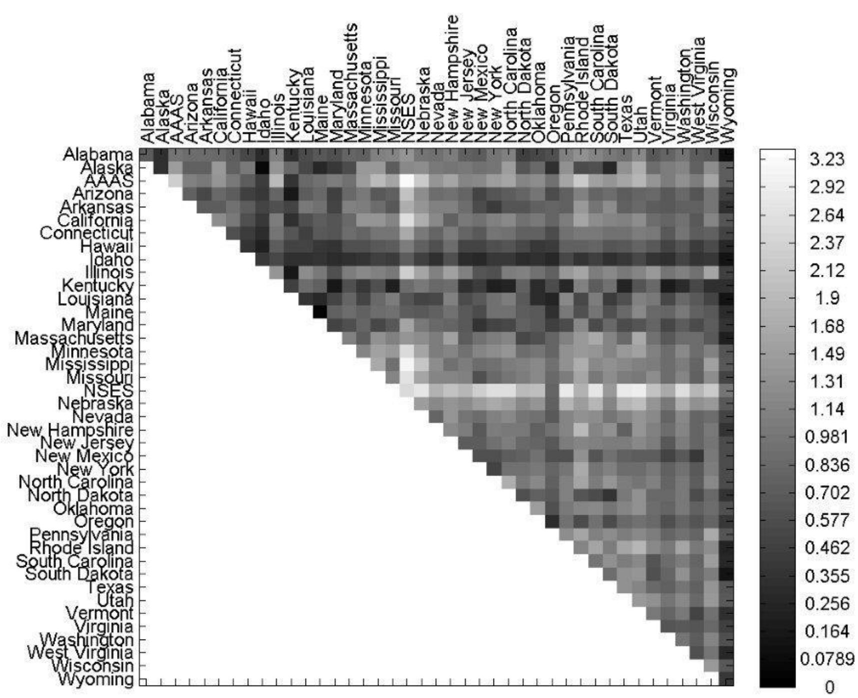

*Figure 5: Average number of bidirectional alignments per standard by state.*

## Standard Crosswalks

For our evaluation of intermediary-based crosswalking feasibility we define a crosswalk as the association of any pair of different standards which can be realized through a two-step, transitive alignment involving an intermediary standard source. Formally:

*Standard RX* ≈ *Standard SY* ≈ *Standard TZ*

where R, S and T represent standard-issuing agencies and X, Y and Z represent science standards written by those agencies.

As crosswalk intermediaries (S) we chose the AAAS and NSES standard sets discussed earlier. With these two intermediaries four sets of crosswalks were derived: crosswalks based on AAAS only, crosswalks based on NSES only, crosswalks supported by both AAAS and NSES (AAAS ∩ NSES) and crosswalks supported by either AAAS or NSES (AAAS ∪ NSES).

## Aggregate Crosswalk Results

Tables 3-6 report the aggregate crosswalking results. Note that the frequencies of the *~Crosswalks*/*~SAT bidirectional* cells were computed based on a SAT-request parameter stating that no more than five results should be returned per standard/author. The loglinear $L^2$ statistic follows a chi-squared distribution and can accordingly be used for significance testing. The $\lambda_{ii}$ parameters provide cell-specific measures of the row-column interaction effects. Positive values indicate a positive relationship. All four tables indicate a statistically significant and positive relationship between SAT bidirectional crosswalking and the four types of intermediary-based crosswalks.

Using the classical information retrieval assessment measures of recall and precision [6], we assess the extent to which the transitive alignment process replicated the results of the direct alignment process. Recall is the proportion of direct alignments which was also found in the transitive ones; *i.e.*, the portion of direct alignment that was replicated using transitive alignment. Precision is the proportion of transitive alignments which is also present as direct alignments.

| *AAAS only as intermediary* | | *Crosswalks* | *~Crosswalks* | *Recall* |
|---|---|---|---|---|
| | *SAT bidirectional* | 313,184 | 762,538 | .291 |
| | *~SAT bidirectional* | 144,1266 | 4,357,012 | |
| | *Precision* | .179 | | |
| $L^2$ = 8.45; DF = 1; p < .01; $\lambda_{11}=\lambda_{22}$= .054 | | | | |

*Table 3: Aggregate crosswalking results for AAAS as intermediary.*

| *NSES only as intermediary* | | *Crosswalks* | *~Crosswalks* | *Recall* |
|---|---|---|---|---|
| | *SAT bidirectional* | 291,326 | 784,396 | .271 |
| | *~SAT bidirectional* | 1,462,478 | 5,798,278 | |
| | *Precision* | .166 | | |
| $L^2$ = 1.633; DF = 1; p < .20; $\lambda_{11}=\lambda_{22}$= .024 | | | | |

*Table 4: Aggregate crosswalking results for NSES as intermediary.*

| *(AAAS Ω NSES) as intermediary* | | *Crosswalks (AAAS Ω NSES)* | *~Crosswalks (AAAS Ω NSES)* | *Recall* |
|---|---|---|---|---|
| | *SAT bidirectional* | 145,288 | 930,434 | .135 |
| | *~SAT bidirectional* | 192,234 | 5,606,044 | |
| | *Precision* | .430 | | |
| $L^2 = 201.84$; DF = 1; $p < .01$; $\lambda_{11}=\lambda_{22}= .379$ | | | | |

*Table 5: Aggregate crosswalking results with both AAAS and NSES as intermediary; crosswalks succeed for both AAAS and NSES.*

| *(AAAS U NSES) as intermediary* | | *Crosswalks (AAAS U NSES)* | *~Crosswalks (AAAS U NSES)* | *Recall* |
|---|---|---|---|---|
| | *SAT bidirectional* | 459,222 | 616,500 | .427 |
| | *~SAT bidirectional* | 2,711,510 | 3,086,768 | |
| | *Precision* | .145 | | |
| $L^2 = 6.08$; DF = 1; $p < .013$; $\lambda_{11}=\lambda_{22}= -.041$; $\tau_{11}=\tau_{22}= .960$ | | | | |

*Table 6: Aggregate crosswalking results with both AAAS and NSES as intermediary; crosswalks succeed for either AAAS or NSES.*

The patterns show that when crosswalking using AAAS by itself (Table 3) and AAAS and NSES jointly (Tables 5 and 6), the relationship between crosswalking and SAT bidirectional alignment is statistically significant. Whereas in the 'AAAS by itself' and 'AAAS Ω NSES' cases this relationship is also positive, in the ' AAAS U NSES ' case it is negative.

The tables also display the associated recall and precision for each of the four crosswalks. Both AAAS and NSES by themselves generate recall and precision of about 30% and 17% respectively. If we use AAAS and NSES jointly, we observe a sharp and not unexpected contrast. Whereas recall and precision for the 'AAAS Ω NSES' case are 13.5% and 43% respectively, almost the reverse is true for the 'AAAS UNSES' case. Clearly, when crosswalks are allowed to be constructed through either an NSES or AAAS intermediary, recall will be higher than when having to construct crosswalks through both an AAAS and an NSES standard. The reverse holds for precision.

**State-specific Results**

In addition to studying the aggregate crosswalk data, we can consider specific authors' crosswalking performance. After all, given the strong differences between the standard bodies of different authors, one can expect that some states will be easier to crosswalk than others.

Table 7 lists the minimum, maximum, mean and standard deviations of recall and precision for the various state-specific crosswalk data sets. Although the means track the aggregate recall and precision numbers, they are a little different, because they were computed from state-specific data with each state weighing evenly. In the aggregate tables (Tables 3-6) all crosswalks are pooled, effectively weighing each state for its number of standards.

| | *AAAS* | *NSES* | *AAAS ∩ NSES* | *AAAS ∪ NSES* |
|---|---|---|---|---|
| *min. recall* | .238 | .210 | .103 | .362 |
| *max. recall* | .670 | .489 | .331 | .750 |
| *μ recall* | .355 | .340 | .184 | .510 |
| *σ recall* | .072 | .067 | .055 | .076 |
| *σ/μ recall* | .202 | .195 | .274 | .149 |
| | | | | |
| *min. precision* | .146 | .138 | .340 | .117 |
| *max. precision* | .310 | .219 | .521 | .199 |
| *μ precision* | .185 | .170 | .433 | .147 |
| *σ precision* | .030 | .021 | .036 | .017 |
| *σ/μ precision* | .160 | .123 | .084 | .116 |

*Table 7: Minimum, maximum, mean and standard deviation of state-specific recall and precision.*

Interestingly, whereas mean state-specific precision (Table 7) closely tracks aggregate values (Tables 3-6), mean state-specific recall is higher than in the aggregate (Tables 3-6), indicating that some of the states with smaller numbers of standards must have relatively high (recall) crosswalking performance. Indeed, the correlations between number of standards and state-specific recall for AAAS-only, NSES-only, 'AAAS ∩ NSES' and 'AAAS ∪ NSES' are r= -.43, r= -.66, r= -.59 and r= -.59, respectively. The corresponding correlations for state-specific crosswalking precision are r=.03, r=-.46, r=-.43 and r=-.50, respectively. On consideration, however, this 'size effect' might also represent an artifact of the method for collecting standard alignment rather than a true relationship between size—the number of standards in a state—and the ability to represent SAT alignments through AAAS or NSES crosswalking. After all, the SAT alignments were collected under conditions of a specified fixed maximum of five target standards, regardless of the number of standards in the target state. This fixed limit of five alignments clearly favors recall for states with few numbers of standards.

Alternatively, however, one would expect that states which have modeled their standard bodies on those of AAAS or NSES achieve better crosswalking recall and precision than states that have not done so, simply because it makes a crosswalk between those states' standards and the AAAS and/or NSES intermediary easier to achieve.

To assess and disentangle these multiple effects, we conducted a series of multiple linear regressions with recall and precision as dependent variables and the following independent variables:

A. Number of source (from) standards.
B. Number of target (to) standards.
C. A x B
D. A dummy indicating if the standards in a source state can be considered modeled on the AAAS standards.
E. A dummy indicating if the standards in a target state can be considered modeled on the AAAS standards.
F. A dummy indicating if the standards in a source state can be considered modeled on the AAAS standards.
G. A dummy indicating if the standards in a target state can be considered modeled on the AAAS standards.

| *Recall* | *AAAS* | | *NSES* | | *AAAS ∩ NSES* | | *AAAS ∪ NSES* | |
|---|---|---|---|---|---|---|---|---|
| *$R^2$ with(out) A, B, C* | 0.26 (0.20) | | 0.33 (0.09) | | 0.28 (0.13) | | 0.31 (0.17) | |
| *Intercept* | 42.163 | *** | 44.045 | *** | 25.098 | *** | 61.110 | *** |
| *A. # source standards* | -0.001 | | -0.010 | *** | -0.004 | ** | -0.008 | ** |
| *B. # target_ standards* | -0.009 | *** | -0.014 | *** | -0.009 | *** | -0.014 | *** |
| *C. AxB* | 0.000 | | 0.000 | | 0.000 | | 0.000 | |
| *D. AAAS source state* | 8.110 | *** | 7.233 | *** | 4.718 | *** | 10.625 | *** |
| *E. AAAS target state* | 9.840 | *** | 1.777 | | 3.565 | *** | 8.052 | *** |
| *F. NSES source state* | -5.032 | *** | -1.924 | | -2.645 | ** | -4.311 | ** |
| *G. NSES target state* | -4.065 | ** | 1.317 | | -0.962 | | -1.787 | |
| | | | | | | | | |
| ***Precision*** | | | | | | | | |
| *$R^2$ with(out) A, B, C* | 0.25 (0.13) | | 0.13 (0.02) | | 0.1 (0.03) | | 0.19 (0.08) | |
| *Intercept* | 23.589 | *** | 19.518 | *** | 49.301 | *** | 17.246 | *** |
| *A. # source standards* | -0.004 | ** | -0.004 | *** | -0.006 | ** | -0.003 | *** |
| *B. # target_ standards* | -0.008 | *** | -0.006 | *** | -0.010 | *** | -0.005 | *** |
| *C. AxB* | 0.000 | | 0.000 | *** | 0.000 | | 0.000 | ** |
| *D. AAAS source state* | 3.237 | *** | 0.972 | | 1.508 | | 1.778 | *** |
| *E. AAAS target state* | 3.744 | *** | 0.677 | | 2.675 | ** | 1.898 | *** |
| *F. NSES source state* | -1.961 | ** | 0.443 | | -1.052 | | -0.524 | |
| *G. NSES target state* | -0.669 | | 1.909 | *** | 1.830 | | 0.524 | |

*Table 8: Regression results for recall and precision with size variables and source state type as independent variables. * $p < .10$; ** $p < .05$; ***$p < .01$.*

Determination of the values for the dummies (D-G) was made by consulting the introductory sections of the states' Departments of Education documentation on their standard writing process. If these documents contained clear statements that the writing of the state standards was modeled on or inspired by AAAS or NSES, or if the documents stated that the AAAS and/or NSES standards were consulted as part of the standard writing process, the dummy was set to 1. In the absence of such statements or if statements containing the opposite were found, the dummy was set to 0.

The results of the regressions are displayed in Table 8. Separate models were run for AAAS-only, NSES-only, 'AAAS ∩ NSES' and 'AAAS ∪ NSES.' The explained variance ($R^2$) is given for each model including and excluding the size variables (A-C).

We observe that being categorized as an AAAS source or target state has by far the greatest and uniformly positive contribution to recall and precision. Being an NSES state, on the other hand, either contributes negatively to both recall and precision in almost all crosswalk types or does not achieve statistical significance. Interestingly, this is particularly true for the NSES-only crosswalks.

The influence of size (A-C), is present yet quite moderate. The effect is statistically significant in most cases and in the direction previously hypothesized; *i.e.*, smaller states—fewer standards—leads to higher recall and precision. However, the size of the effect is quite small with parameter values in the order of -.001 to -.01.

Returning to our discussion of the differences between states (Figures 4-7), we observe several patterns:

- There are some great differences between minimum and maximum recall and precision crosswalking performance between states. For instance, whereas 'AAAS only' crosswalking accomplished an impressive 67% recall and 31% precision for Rhode Island—the means are 35.5% and 18.5% respectively—the corresponding minima were 23.8% (recall) and 14.6% (precision) for Kentucky and Mississippi, respectively.

  One explanation for Rhode Island's high performance on AAAS-based crosswalking which is supported by the regressions presented above, would be that the Rhode Island science standards were closely modeled on the AAAS standards. Since this would greatly increase the accuracy of the Rhode Island–AAAS side of the crosswalk, it would improve the performance of the crosswalk overall.

  We furthermore care to observe that Rhode Island's (relatively) high recall and precision provide some supporting evidence of the performance and accuracy of the SAT tool. Indeed, if we would have known of the close relationship between the Rhode Island and AAAS standards prior to conducting this analysis, we would have been looking for such high recall and precision, merely because SAT should have picked up this connection on the RI-AAAS side of the crosswalk. That it did in fact do so, speaks in its favor.

- The state-specific recall and precision show some interesting relationships. For all data sets except 'AAAS ∩NSES,' all recall measures—minimum, maximum, mean, sigma and sigma/mean—are significantly higher than the corresponding precision numbers. This indicates that AAAS/NSES-based crosswalking is 'optimistic,' *i.e.*, it is biased in favor of false positives relative to false negatives.

  This relationship is strongest in the 'AAAS ∪ NSES' crosswalks. Reminding the reader that in this data set a crosswalk could be established through either the AAAS or NSES as intermediary, maximum recall is 75% whereas maximum precision is only 19.7%—both Rhode Island's. This positive jump in recall and corresponding negative one in precision is to be expected. After all, by increasing the set of intermediary alignments the bias in favor of false positives is increased.

  Going in the other direction, however; *i.e.*, considering a crosswalk only if it is supported by both AAAS and NSES (AAAS ∪ NSES), shows the opposite effect. In this case, recall ($\mu$ = 18.4%) is sharply reduced, but precision ($\mu$ = 43.3%) greatly improves.

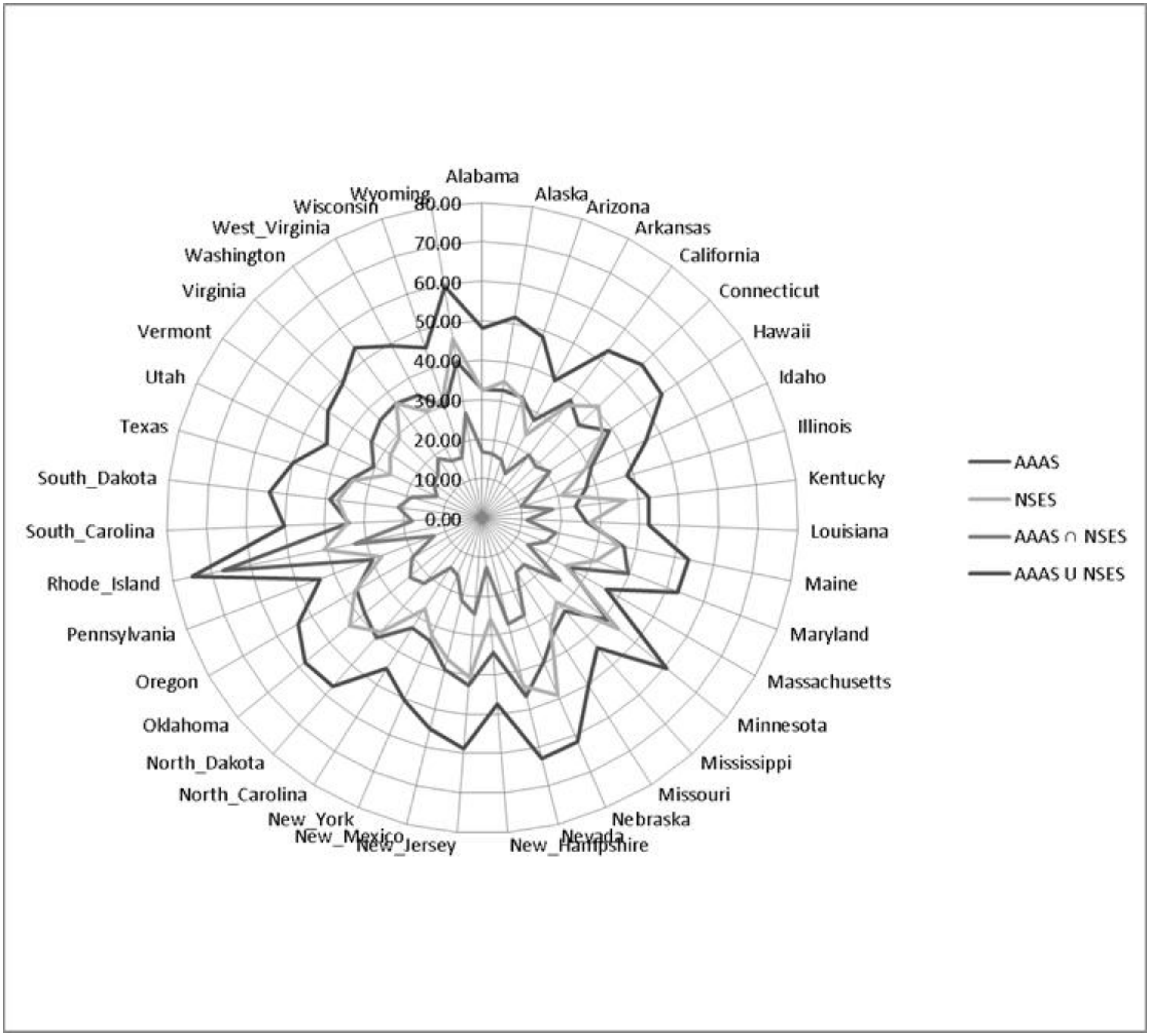


*Figure 6: Recall by state for the four intermediary-based crosswalk types.*

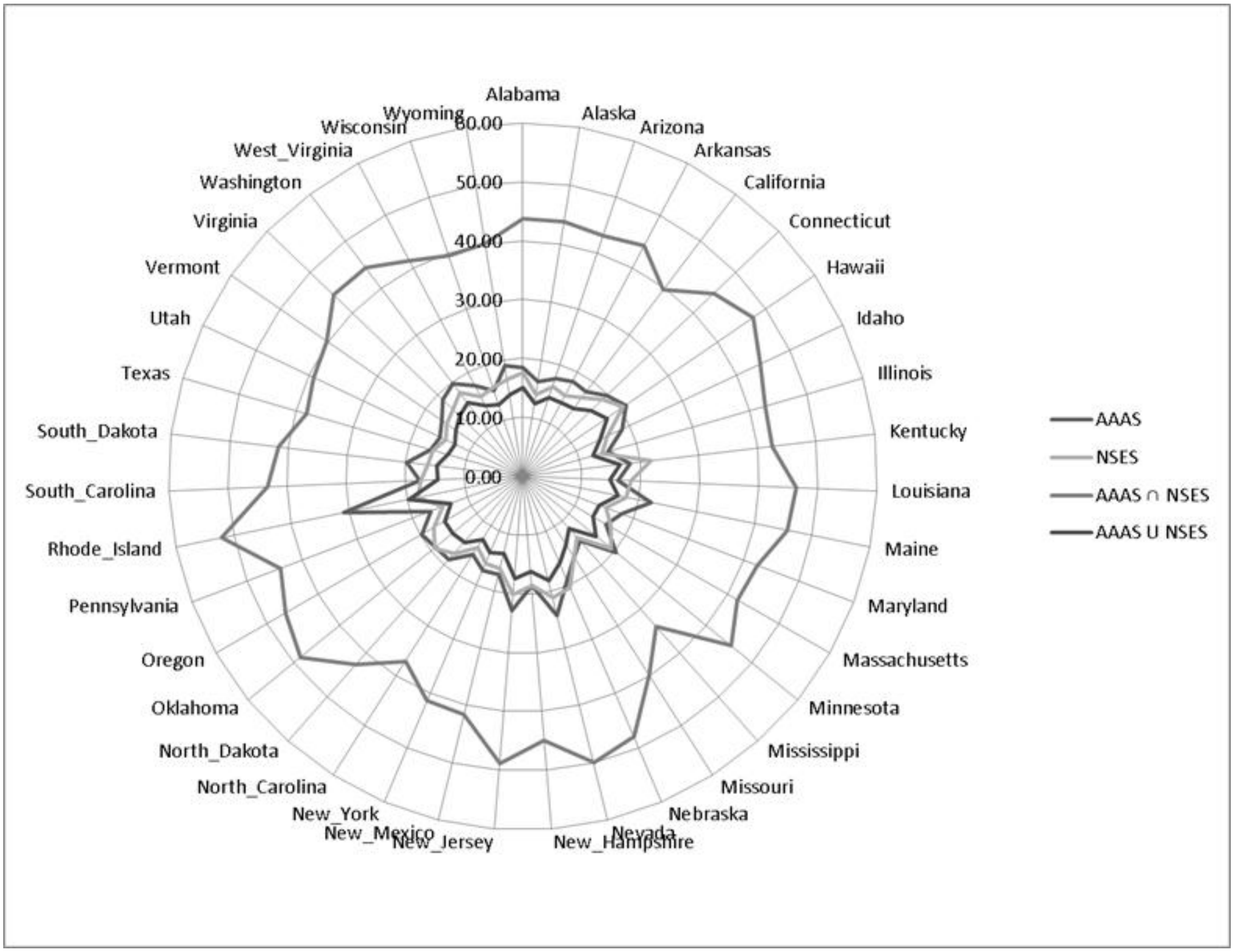


*Figure 7: Precision by state for the four intermediary-based crosswalk types.*

**Conclusion**

We explored the feasibility of automated, intermediary-based science standard crosswalking. We are interested in this type of crosswalking because it may contribute to a solution to the daunting yet real-world challenge of aligning an ever-changing and growing supply of learning resources with a large and changing body of educational standards. In addition, this formulation—transitive matches involving multi-sourced intermediaries—may be applied to other, similar retrieval environments such as work flow or web service matching, forensic financial transaction analysis, law enforcement, or distributed image collections to name a few. A better understanding of how to assess the accuracy of transitive relations may be of use in developing robust, multi-faceted network-driven search mechanisms.

From a standards crosswalking value perspective, the results of this feasibility study are mixed. The overall situation, averaged across states, is not very promising as precision and recall are quite low in most cases. Although precision levels of around 40% can be reached when requiring alignment with both intermediaries, the corresponding recall is quite low. We attribute this lack of crosswalking success to the great variety in standard formulations across states. Alternatively, one might want to attribute the lack of success to SAT. Perhaps SAT is just not (yet) good enough to find the crosswalks? Speaking in SAT's favor, however, is that we did find some convincing evidence that as states model their standards on those contained in the intermediary body, promising recall and precision was achieved. We see this reflected in the results for individual states such as Rhode Island, but also in the aggregated statistical results of the regressions with recall and precision as dependent variables. Whereas both recall and precision for some states reach levels which can be considered practical and ready for real-world use, aggregate levels of recall and precision are perhaps not high enough to consider this type of crosswalking suitable across just any pair of standard authoring bodies. In addition, we observed that crosswalking precision and recall have a (very) strong inverse relationship when two intermediaries are used in either a union (AAAS ∪ NSES) or intersection (AAAS ∩ NSES) mode. These results provide some support for the assessment that the lack of overall success in crosswalking should not be mainly attributed to SAT. Although all analysis was based on data derived from SAT, and hence, any potential SAT-bias was inherent, the high recall and precision of a typical AAAS-modeled state such as Rhode Island support SAT as a valid standard crosswalking tool.

Despite these mixed results, however, we should not lose sight of the real advantage that automated and intermediary-based crosswalking promises, namely the ability to provide alignments in linear rather than exponential time. Neither should we lose sight of the fact that tools such as SAT are quickly improving which may lead to better recall and precision.

As mentioned earlier, standard harmonization efforts such as the Common Core initiative should alleviate an important part of the burden of standards crosswalking. Reducing a standard set from ≈60K to a much smaller, broadly accepted set of standards greatly simplifies the curriculum-standard alignment challenge. However, whereas harmonization efforts reduce the need for standard crosswalking, globalization of learning resources through media such as the Internet and the continued efforts of nongovernmental organizations in redefining and redeveloping standards, add to the crosswalking challenge.

Current and future standard changes imply the need for realignment of previously aligned resources. Whereas both direct—resource-to-standard—and standard crosswalking approaches might be used for such realignment, changes in how standards are structured and formulated may have significant consequences for either of these approaches. One such change, namely the integration of both content and methods of inquiry into single standards is apparent in the latest releases of the science standards of some of the USA states. The following two examples from the 2009 Colorado Science standards illustrate this:

- S11424E3-CO- Physical Science-Sixth Grade: *Identify evidence that suggests there is a fundamental building block of matter.*
- S11424F1- CO- Physical Science-Fifth Grade: *Develop, communicate, and justify a procedure to separate simple mixtures based on physical properties.*

In both of these examples—many more can be found in the latest releases of various USA states' science standards as well as in Common Core Standards for Literacy in History/Social Studies, Science, and Technical Subjects and the latest proposals by the US National Research Council's Board on Science Education—content and method are tightly integrated reflecting the standard authors' views on the nature of knowledge and learning. Useful and appropriate as this content/method integration may be from a pedagogical and learning perspective, it poses a serious challenge to automated alignment mechanisms because as discussed earlier, these mechanisms seem to have special difficulty in recognizing methodological content. Hence, we foresee situations in which the mechanism triggers an alignment based on content, yet misses or worse, misaligns, the methods aspects of a standard.

Regardless of these difficulties, however, our analysis of intermediary-based standard crosswalking indicates the inherent difficulty of that approach on the one hand and at the same time provides support for standard harmonization efforts on the other. Adoption or adaption of a shared, third party standard, be it NSES, AAAS, ITEEA or indeed a Common Core, either whole or in part, greatly increases the likelihood of being able to connect one's learning resources with those of others developed elsewhere.

We have only just begun to understand the notion of standard alignment and it is proving to be a much more complex and challenging phenomenon than perhaps originally considered. We offer that network-driven, multi-faceted retrieval mechanisms might support automated alignment efforts. Our results, however, suggest that these mechanisms must be used with appropriate care. Developers of educational resource portals using automatically generated crosswalk data should likely consider such evidence more or less compelling for different pairs of jurisdictions. Evaluation such as the one presented here may be useful in preparing weight adjustment schemes noting that such adjustments should depend not only on the type of specific sources—all analyzed here are either regional or national organizations—but also on measurable cross-walking effectiveness scores. Retrieval systems which seek to identify learning objects for standards might, for example increase the weighting of transitive associations between Rhode Island and Nevada, where high precision values were noted in the matches, while considering the ones between Kentucky and Mississippi, which had a relatively small number of transitive relationships, less compelling. The process we use here to identify where an automatic process is more or less consistent in establishing the likely correctness of transitive matches may be applied in other network-driven retrieval applications.

## References


1. AAAS - American Association for the Advancement of Science (1989) Science for all Americans: Summary: Project 2061. AAAS. Washington, DC.
2. AAAS - American Association for the Advancement of Science (1989) Science for all Americans: a Project 2061 Report on Literacy Goals in Science, Mathematics, and Technology. AAAS. Washington, DC.
3. AAAS - American Association for the Advancement of Science (2001) Atlas of Science Literacy. AAAS. Washington, DC.
4. Anderson, N. (2010) Governors, State School Superintendents Propose Common Academic Standards. Education Brief. The Washington Post. March 10, 2010. http://www.socialstudies.org/governors_state_school_superintendents_propose_common_academic_standards
5. Arentze, T.A., Timmermans, H.J.P (2005) Information Gain, Novelty Seeking and Travel: a Model of Dynamic Activity-Travel Behavior Under Conditions of Uncertainty. Transportation Research Part A: Policy and Practice. 39. 125–145.
6. Baeza-Yates, R., Ribiero-Neto, B. (1999) Modern Information Retrieval. Addison Wesley–ACM Press. Harlow, UK.
7. Blumenthal, D. (2003) Evaluation Approaches for a K-12 Digital Library Collection. WGBH Educational Foundation. Boston. MA. http://eduimpact.comm.nsdl.org/evalworkshop/blumenthal.doc.
8. Bologna Secretariat (2010). Welcome to the website of the European Higher Education Area: The Official Bologna Process Website 2007–2010. Benelux Bologna Secretariat, Brussels. http://www.ond.vlaanderen.be/hogeronderwijs/bologna/.
9. Breen, C. (2008) Review: NetTrekker. MacWorld.com. http://www.macworld.com/article/134787/2008/08/nettrekker.html.
10. Cai, X., Li, W. (2011) Enhancing Sentence-level Clustering with Integrated and Interactive Frameworks for Theme-based Summarization. Journal of the American Society for Information Science and Technology. 62. 2067–2082.
11. Cunningham, C.M., Hester, K. (2007) Engineering is Elementary: An Engineering and Technology Curriculum for Children. Proceedings of the ASEE Annual Conference, Honolulu, HI. ASEE, Washington D.C. http://www.starwars.mos.org/eie/pdf/research/asee_2007_dev_paper.pdf.
12. Devaul, H., Diekema, A.R., Ostwald, J. (2011) Computer-Assisted Assignment of Educational Standards Using Natural Language Processing. Journal of the American Society for Information Science and Technology. 62. 395–405.
13. Diekema, A.R., Chen, J. (2005) Experimenting with the Automatic Assignment of Educational Standards to Digital Library Content. In: Proceedings of the 5th ASM/IEEE Joint Conference on Digital Libraries. ACM. New York, N.Y.
14. Diekema, A.R., Yilmazel, O., Bailey, J., Harwell, S.C., Liddy E.D. (2007) Standards Alignment for Metadata Assignment. Proceedings of the Joint Conference of Digital Libraries (JCDL'07). Vancouver, BC. IEEE-ACM. New York, NY.
15. Diekema, A.R. (2009). Implications and Challenges of Educational Standards Metadata. Journal of Library Metadata. 9.

239–251.

16. Doan, A. H., Halevy, A.Y. (2005). Semantic Integration Research in the Database Community: A Brief Survey. AI Magazine. 26. 83-94.
17. Felix, K. (2004) In the Spotlight. NetTrekker. http://www.infotoday.com/MMSchools/jan04/felix.shtml.
18. Gateway (2007) NSDL:ASN Achievement Standards Network. http://www.thegateway.org/asn.
19. ITEEA (International Technology and Engineering Educators Association) (2011) Opportunities for Technology and Engineering in the Sciences Framework. http://www.iteaconnect.org/Resources/PressRoom/2011/Jul/ScienceStdsRespPR.pdf
20. Jay, M., Longdon, D. (2003) Death, Taxes and Correlations: A Primer on the State of Correlation in the K-12 Education. Upgrade, SIIA. 20-21.
21. Gateway (2007) NSDL:ASN Achievement Standards Network. http://www.thegateway.org/asn.
22. Johnson, A. (2003) Summative Evaluation Teachers' Domain. Report submitted to WGBH Educational Productions. WGBH Boston, MA.
23. Kendall, J.S., Marzano, R.J. (1997) Content Knowledge. A Compendium of Standards and Benchmarks for K-12 Education. Second Edition. McREL. Aurora, CO
24. Kubilinskiene, S., Dagiene, V. (2009) Technology-based Lesson plans: Preparation and Description. Informatics in Education. 9. 217-228.
25. Lachapelle, C.P., Cunningham, C.M. (2007) Engineering is Elementary: Children's Changing Understandings of Science and Engineering. Proceedings of the ASEE Annual Conference, Honolulu, HI.
26. Lin, S.D., Chalupsky, H. (2003) Using Unsupervised Link Discovery Methods to Find Interesting Facts and Connections in a Bibliography Dataset. SIGKDD Explor. Newsl. 5. 173-178.
27. Marshall, B., Chen, H., Kaza, S. (2008 ) Using Importance Flooding to Identify Interesting Networks of Criminal Activity. Journal of the American Society for Information Science and Technology. 59. 2099-2114.
28. Marshall, B., Reitsma, R.F., Cyr, M.N. (2007) Semantics or Standards for Curriculum Search? Proceedings of the Joint Conference on Digital Libraries (JCDL'07), Vancouver, BC. IEEE-ACM.
29. Mason, B. (2006) Digital Libraries in Support of Science Education. Computing in Science and Engineering. 8. 62-65.
30. McAnear, A. (2007) What Does Globalization Mean for Education? Learning and Leading with Technology. 5. Available: http://www.iste.org/Content/NavigationMenu/Publications/LL/LLIssues/Volume_34_2006_2007_/February_5_/34505m.pdf.
31. Melnik, S., Garcia-Molina, H., Rahm, E. (2002) In Similarity Flooding: a Versatile Graph Matching Algorithm and its Application to Schema Matching. Proceedings of the 18th International Conference on Data Engineering (ICDE '02), IEEE.
32. National Governors Association Center for Best Practices (2011) Common Core State Standards Initiatives. http://www.corestandards.org.
33. NRC - National Research Council (1996) National Science Education Standards. National Academies Press, Washington, DC.
34. NRC - National Research Council (2011) A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas. National Academies Press. Washington, DC.
35. Reitsma, R., Marshall, B. Zarske, M. (2010) Aspects of 'Relevance' in the Alignment of Curriculum with Educational Standards. Information Processing and Management. 46. 362-376.
36. Reitsma, R., Diekema A. (2011) Comparison of Human and Machine-based Educational Standard Assignment Networks. International Journal on Digital Libraries. 11. 209-222.
37. Richards, L.M. (2010) AMSER: Applied Math and Science Education Repository. Reference Reviews. 24. 38-39.
38. Salton, G., Yang, C.S., Wong, A. (1975) A Vector Space Model for Automatic Indexing. Communications of the ACM. 18. 613-620.
39. Saracevic, T. (1975) Relevance: A Review of the Literature and a Framework for Thinking on the Notion in Information Science. Journal of the American Society of Information Science. 26(6). 321-343.
40. 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. Journal of the American Society of Information Science and Technology. 58. 1915-1933.

41. 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. Journal of the American Society of Information Science and Technology. 58. 2126-2144.
42. Saylor, J.M., Minton-Morris, C. (2006) The National Science Digital Library: An Update on Systems, Services and Collection Development. Science & Technology Libraries. 26. 61-78.
43. Schamber, L., Eisenberg, M.B., Nilan, M.S. (1994) A Re-examination of Relevance: Toward a Dynamic, Situational Definition. Information Processing and Management. 26. 3-48.
44. Sullivan, J.F., Cyr, M.N., Mooney, M.A., Reitsma, R.F., Shaw, N.C., Zarske, M.S., Klenk, P.A. (2005) The TeachEngineering Digital Library: Engineering Comes Alive for K-12 Youth. Proc. ASEE Annual Conference; Portland, OR. ASEE, Washington, D.C. http://www.teachengineering.org/documents/ASEE%202005-851_TE_Final.pdf.
45. Sumner, T. Ahmad, F., Bhushan, S., Gu, Q., Molina, F., Stedman, W., Wright, M., Davis, L., Janée, G. (2005) Linking Learning Goals and Educational Resources Through Interactive Concept Map Visualizations. International Journal on Digital Libraries. 5. 18-24.
46. Sutton, S.A., Golder, D. (2008) Achievement Standards Network (ASN): An Application Profile for Mapping K-12 Educational Resources to Achievement Standards. Proceedings of the 2008 International Conference on Dublin Core and Metadata Applications. Humboldt University, Berlin, Germany. 69-79.
47. Teacher's Domain (2011) Standards Correlation and Lexicon Terms; available: http://www.teachersdomain.org/tdwiki/doku.php/partner:standards_correlation_and_lexicon_terms
48. Thill, J.C. (Ed.) (1999) Spatial Multicriteria Decision Making and Analysis. Ashgate Publishing, Ltd., Aldershot, UK.
49. Timmermans, H.J.P., van der Heijden, R., Westerveld, H. (1984) Decision Making Between Multi-attribute Choice Alternatives: a Model of Spatial Shopping Behavior Using Conjoint Measurements. Environment & Planning A. 16. 377-387.
50. USGPO (United States Printing Office) (2001) Public Law 107. An Act to Close the Achievement Gap with Accountability, Flexibility, and Choice, so that No Child Is Left Behind. http://www.gpo.gov/fdsys/pkg/PLAW-107publ110/content-detail.html.
51. Wallis, C., Steptoe, S. (2006) How to Bring Our Schools Out of the 20th Century. Time, Dec. 10. http://lutheransouth-hs.org/ourpages/publications/Article%20Time%20-%20How%20to%20Bring%20Schools%20out%20of%2020th%20century.pdf.
52. White, H.D. (2003) Pathfinder Networks and Author Cocitation Analysis: A Remapping of Paradigmatic Information Scientists. Journal of the American Society for Information Science and Technology. 43. 423-434.
53. Xu, J., Chen, H. (2004) Fighting Organized Crime: Using Shortest-Path Algorithms to Identify Associations in Criminal Networks. Decision Support Systems. 38. 473-487.
54. Yilmazel, O., Balasubramanian, N., Harwell, S.C., Bailey, J., Diekema, A.R., Liddy, E.D. (2007) Text Categorization for Aligning Educational Standards. Proceedings of the 40th Hawaii International Conference of Systems Sciences. IEEE. New York, NY.
55. Yilmazel, O., Ingersoll, G., Liddy, E.D. (2007) Finding Questions to Your Answers. IEEE 23rd International Conference on Data Engineering (ICDE). Istanbul, Turkey.
56. Zia, L. (2002) The NSF National Science, Technology, Engineering and Mathematics Education Digital Library (NSDL) Program. New Projects in Fiscal Year 2002. D-Lib Magazine. 8. http://www.dlib.org/dlib/november02/zia/11zia.html.
57. Zia, L. (2005) Digital Libraries for Education: A Progress Report on the National Science Foundation's (NSF) National Science, Technology, Engineering, and Mathematics Education Digital Library (NSDL) Program. Lecture Notes in Computer Science. 3334. 45-50.