Indices of Semantic Similarity for Automated Essay Scoring (AES)[1]

Yoav Cohen & Anat Ben-Simon

National Institute for Testing and Evaluation, Jerusalem, Israel

Effi Levi

Hebrew University of Jerusalem

Abstract

Content is one of the main writing dimensions on which essays are judged and rated. Since no automated essay scoring (AES) system is capable (yet) of truly understanding the content of an essay and assessing its breadth, depth and relevance, AES systems use indirect methods and proxy indices for judging its quality. Most such indices are based on measures of semantic similarity between a given essay and some gold standard.

The purpose of this study is to examine the efficiency (validity) of five computer-generated sematic indices used by NiteRater – an AES system for text analysis and essay scoring of Hebrew texts (NiteRater, 2007). These indices can be classified into three categories: (1) indices based on semantic proximity between essays – the similarity of an essay's vocabulary to that of essays in various score-categories; (2) indices based on Principal Component Analysis (PCA) of semantic similarities; and (3) indices based on prompt-related vocabulary – the similarity of the essay's vocabulary to that of the prompt.

Six essay-corpora of various genres were used to study the efficiency of the semantic indices, including essays written by native and non-native Hebrew speakers. The efficiency of these indices was assessed by correlating them with raters' scores. The internal structure of the semantic indices, as well as their differential validity for essays of different genres was also studied.

The results of the study show that indices based on semantic proximity can capture a large proportion of the essay scores (r=.28-.85). These are followed by indices based on PCA of semantic similarities (r=.27-.57) and finally, by indices based on prompt-related vocabulary (r=-.51-.59), which are also the most sensitive to essay genre.

[1] Paper presented at the annual meeting of the International Association of Educational Assessment (IAEA), 2014, Singapore.

Introduction

Automated essay scoring (AES) systems were first presented and applied about 50 years ago (Page, 1966). The systems have matured since then and, used extensively for the past two decades, have been proven to produce reliable and valid measures of writing ability (Shermis & Burstein, 2003; Ben-Simon & Bennett, 2007). Most current systems are used to score essays written in English.

In a typical system, numerous statistical and natural language processing (NLP) features are extracted from large corpora of student essays; the most useful features are identified by correlating the features with human scores and a scoring model is developed. Almost all AES systems attempt to mimic, as closely as possible, the scores produced by human raters.

Most AES systems produce feature scores, as well as scale scores that represent various writing dimensions. Of these, *content* is one of the most commonly used. Since no AES system is capable of truly understanding the content of an essay and assessing its breadth, depth and relevance, AES systems use indirect methods and proxy indices for judging its quality. Most such content indices are designed to measure semantic similarity between a given essay and some gold standard. 'Semantics' in this context has a rather narrow meaning. By 'semantics', we refer here not to the full meaning of a paragraph or a sentence, but to the meaning conveyed by a collection of decontextualized words or lexemes.

The purpose of this study is to examine the efficiency (validity) of five computer-generated semantic indices used by NiteRater – an AES system for scoring essays in Hebrew (NiteRater, 2007). The indices can be classified into three categories: (1) indices based on semantic proximity – the similarity of an essay's vocabulary to that of essays in various score-categories; (2) indices based on Principal Component Analysis (PCA) of semantic similarities and (3) indices based on prompt-related vocabulary – the similarity of the essay's vocabulary to that of the prompt.

NiteRater

NiteRater is a program that was developed by the National Institute for Testing & Evaluation (NITE) in Israel for text analysis and essay scoring of Hebrew and Arabic texts (NiteRater, 2007). The feature-extractor used by NiteRater presently produces about 200 quantified text features (micro-features) from a single text or a set of texts. These features include statistical (surface), grammatical, morphological, lexical and semantic features. The version of NiteRater used in the current study has four writing dimensions, which encompass 15 scoring dimensions and 31 macro-features (see Appendix A for details).

Five of these 15 scoring dimensions are semantic indices: (1) the nearest semantically proximate essay score (NTS); (2) the semantic distance from essays in the highest score category (DHS); (3) the mean score of the K most semantically similar essays (MKN); (4) unsupervised PCA-based semantic rank; and (5) prompt-related vocabulary based on the overlap of essay vocabulary and the essay prompt (letter-string & lexeme) (PRV).

In a recent study of the factorial structure of writing ability using machine-generated linguistic text features, a factor analysis was applied to NiteRater scoring dimensions generated for five essay-corpora (Ben-Simon & Safran, 2012). Results of the study revealed three dimensions: *lexical complexity* (fluency), *topical analysis* (content) and *vocabulary*. Two content features were consistently loaded on the topical analysis dimension: (1) the nearest semantically proximate essay score; and (2) the semantic distance from essays in the highest score category. The unsupervised PCA-based semantic rank did not show a consistent pattern across the five corpora. In all five corpora, the topical analysis dimension produced the highest correlations with both raters' average score on the content dimension (.46-.77) and the raters' total score (.51-.82).

Semantic Features for Essay Scoring

General approach to semantic analysis

One of the most widespread methods used to extract semantic features for a set of texts is to define a vector space with dimension M, where each text is represented by an M-dimension vector. Each coordinate in the space represents a word in the joint vocabulary space of the texts, and the coordinate value for a vector representing a specific text is the number of times that the word corresponding to that coordinate appears in the text.

It is important to note that in this method, a text is treated as a **bag of words**; i.e., the order of words is not considered. Therefore, only individual word semantics is measured regardless of the context in which the word appears.

The next step is to define a metric (or a distance function) on the semantic space. Since the context vectors define a Euclidian space, the natural choice is the Euclidian distance. This choice, however, has a major disadvantage: the Euclidian distance between vectors is highly dependent on their lengths (norms), or the lengths of the texts they represent. Since our goal is to quantify semantic distance (or similarity) between texts, we prefer a distance function that does not factor in text length. For this reason, we used **cosine similarity**, which measures the cosine of the angle between the vectors; hence the distance function in this study is:

$$\text{dist}(x,y) = 1 - \text{cosine\_similarity}(x,y).$$

Once the semantic space and a matching distance function are defined, the features that quantify the semantic relation between different texts can be defined. It is important to note that the method will produce good results only if the set of texts does in fact constitute a semantic space in some sense (e.g., a set of texts relating to a given topic).

Co-occurrence matrix

The first step in semantic analysis is to calculate the co-occurrence matrix (the set of context vectors for all the essays). Based on the assumption that morphology does not affect word semantics, we chose to assign space coordinates to lexemes instead of words. This choice has two main advantages:

a. The dimension of the resulting space is significantly smaller and thus reduces computational complexity.
b. Reducing the redundancy allows for a better representation of the semantic space.

For a set of N essays (divided into a training set – for which the scores are already given – and a test set – for which the features are calculated) with a joint vocabulary of M lexemes, we calculate a MxN matrix, where each column represents a specific essay. The value at the m,n coordinate is the number of times the m^{th} lexeme appears in the n^{th} essay.

Extraction of semantic features

But before calculating the indices, we modified the co-occurrence matrix as follows:

a. All the lexemes that appear in only one of the essays were removed, as each such lexeme constitutes a dimension that is not a part of the *joint* semantic space.

   Different weighting methods were applied to the co-occurrence matrix in order to re-weight the matrix values, for example to assign higher weights to rarer lexemes (mirroring their importance). Various local and global weighting methods were applied:

   i. Local weight: depends only on the value itself. We tried the following transformations: identity, logarithm, binary (0 for zero frequency and 1 otherwise).

   ii. Global weight: depends on the value itself and the values of the entire matrix row.

   The optimal weighting method for each index and corpus was determined empirically for each index.

b. The rank of the co-occurrence matrix was reduced using Singular Value Decomposition (SVD). (First, the matrix is decomposed into a multiplication of three matrices: U*S*V' where S is the matrix of singular values. Then, some of the singular values are removed by deleting them from S. Finally the matrices are re-multiplied).

The result is a matrix with the original co-occurrence matrix dimensions but a lower rank.
The rank reduction procedure produces a matrix with the original matrix dimensions but a lower rank. The rank reduction has three main advantages: some of the computations are faster for lower-ranked matrices, noise cancelation and unification of dimensions that have similar semantic meaning (such as synonyms).

The semantic features

The next step was to calculate semantic features for the essays from the co-occurrence matrix. For this purpose, we defined five semantic indices that rely on the co-occurrence matrix and the cosine similarity distance function. The first three indices are derived from a given training set (for which raters' scores are given), while the fourth and fifth indices are calculated without using a training set.

1. *Nearest semantically proximate essay score (NTS)*

   This index identifies the score from the training set that is semantically closest to the test essay.

   To calculate this index, all the essays in the training set are clustered into score categories reflecting the discrete scores given on the scoring scale. Each cluster (score category) is then concatenated to a 'super-essay' representing the respective score category. For example, for a six-score scale (1-6), six 'super-essays' will be generated. Next, a co-occurrence matrix is calculated for the 'super-essays' to obtain a matrix in which each column corresponds to a 'super-essay' (specific score).

   Finally, the distance of the text of a test essay from each one of the 'super-essays' is calculated and the final index score assigned to the test essay is the score of the nearest 'super-essay'.

2. *Distance from the highest score category (DHS)*

   DHS is based on the semantic distance of the test essay from the 'super-essay' of the highest score category (as obtained from the training set).

3. *Mean score of the K nearest neighbors (MKN)*

   MKN is arrived at by computing the average score of the K nearest essays (taken from the training set) to the test essay. To calculate this index, we use the defined distance function to rank the essays in the training set according to their distance from the test essay. The MKN index is the average score of the first K essays, where K ranges between 1% and 25%. The exact value of K is determined separately for each essay corpus.

4. *Unsupervised PCA-based semantic rank (PCA)*

The fourth index involves determining the spectral decomposition of the distance matrix between all essay pairs, then calculating the projection of each of the essays on the first principal component.

First, the pair-wise distance matrix is calculated for all essays (according to the semantic space defined by the co-occurrence matrix). Then a PCA procedure is applied to the distance matrix to obtain the decomposition of the semantic distance space to the principal components. The first principal component is the dimension that contains the largest amount of information (explained variance) and hence, is the dimension on which we project each one of the essays. The PCA index for each respective essay is the coefficient of this dimension (or the first coefficient) in the matrix obtained from the PCA procedure. It is important to note that this index, unlike the previous three, does not require a scored training set and is therefore obtained in an unsupervised fashion.

5. *Prompt-related vocabulary*

The fifth index is based on measuring the semantic distance between each of the essays and the prompt that was given with the essay assignment. The prompt was represented by a context vector in the same way as were the essays; then, for each essay, the distance was calculated between the essay's and the prompt's context representations. Unlike the previous indices, this index is fairly sensitive to essay genre and length of the prompt.

Method

Essay-corpora

Six essay-corpora of various student populations were used to study the efficiency of the semantic indices. The essays in the six corpora were written by students of different ages, as well as by native and non-native Hebrew speakers (see Table 1).

Table 1. Essay-corpora used in the study

Corpus	N	Genre*	Grade	Language	Test
NA8-A NA8-B	665 649	Inf. Arg.	8th Grade	Hebrew *native speakers*	National assessment: Hebrew language test (Maytzav)
IA12	659	Arg.	12th Grade	Hebrew *native speakers*	Experimental instructional writing program
HFL-A HFL-B	489 498	Arg.	College Applicants	Hebrew *non-native speakers*	Test of Hebrew as a foreign language (YAEL)
PET-H	1,000	Arg.	College Applicants	Hebrew *native speakers*	Writing task, component of the Israeli admissions test to higher education (PET)

* Arg. - argumentative; Inf. - informative

Analysis

To examine the relationships between the indices, the inter-correlation matrix among the five indices was calculated for each essay-corpus.

To examine the efficiency (validity) of the indices, a simple correlation of each index with the essay scores was calculated. Then, to study the overlap between the various indices, all the indices were incorporated in a single prediction equation generated from a training set and cross-validated with a test set.

The cross-validation procedure was carried out as follows: each corpus was randomly divided into five essay sets. A regression model was generated using four of the five sets (training set). The model was then applied to the remaining set (test set), producing a predicted score for each essay in that test set. The procedure was repeated five times, once for each of the five essay sets serving as the test set, hence producing predicted scores for the complete essay set. The predicted score was then correlated with raters' score.

This cross-validation procedure was repeated 100 times. Finally, the 100 correlation coefficients obtained were averaged. The full procedure was carried out separately for each corpus.

To assess the contribution of the semantic indices to the prediction of final essay scores within the full NiteRater prediction model, the abovementioned cross-validation procedure was applied again, using the five semantic indices and the remaining 10 NiteRater scoring dimensions.

Results

Table 2 shows the means and ranges of the inter-correlations among the semantic indices across the six essay-corpora. Since the PRV index is highly sensitive to essay genre, correlations between the PRV index and the other indices were calculated separately for the informative writing task (NA8-A), which required the students to summarize a 390-word article, and the argumentative essays, which were written in response to much shorter prompts. Moderate to high correlations were observed among the five indices, suggesting a potential overlap and thus redundancy of some of the indices.

Table 3 presents the validities (correlations with the raters' scores) of the various indices for each corpus. The results show that indices based on semantic proximity account for a larger proportion of variance in essay scores (mean r across corpora =.62-.64). Indices based on

PCA of semantic similarities and prompt-related vocabulary account for a smaller proportion of the variance (mean r across corpora =.43 and .40[2], respectively).

The multiple regression coefficients (cross-validation) of the five indices in the prediction of raters' scores ranged from .489 to .892 with a mean r of .733. The mean multiple regression coefficients (cross-validation) obtained for all 15 NiteRater scoring dimensions with raters' scores was only slightly higher (.775). It is interesting to note that the marginal contribution of NiteRater's 10 non-semantic indices to the prediction of raters' scores accounted, on average, for only 6.4% of the explained variance.

Table 2: Means (and ranges) of the inter-correlations among the semantic indices across the six Hebrew essay-corpora

	NTS	DHS	MKN	PCA
DHS	**.61** (.003-.75)			
MKN	**.78** (.57-.87)	**.65** (.07-.81)		
PCA	**.34** (-.08-.39)	**.80** (.52-.94)	**.32** (-.09-.71)	
PRV (argument.)	**-.41** -.35-(-.59)	**-.14** -.51-(-.38)	**-.47** -.69-(-.34)	**.24** (.02-.36)
PRV (informative)	**.68**	**.88**	**.75**	**.76**

NTS – Nearest text score
DHS – Distance from highest score
MKN – Mean score of K nearest neighbors
PCA – Unsupervised PCA-based
PRV – prompt-related vocabulary

[2] The average correlation for the PRV index was calculated using the absolute value of the correlations obtained for each corpus; this practice was applied in light of the fact that the direction of correlation (positive/negative) between this index and raters' scores varies according to the genre.

Table 3: Simple and multiple correlations between semantic indices and raters' scores by essay-corpora

	NA8-A	NA8-B	IA12	HFL-A	HFL-B	PET-H	Mean
Simple correlations*							
NTS	.58	.28	.65	.85	.70	.55	.63
DHS	.62	.34	.52	.79	.74	.58	.62
MKN	.67	.33	.71	.77	.70	.57	.64
PCA	.57	.27	.51	.53	.35	.28	.43
PRV	.59	-.16	-.34	-.36	-.51	-.37	.40**
Multiple regression correlation							
Semantic indices only	.677	.489	.718	.892	.802	.677	.733
15 scoring dimensions	.689	.631	.742	.916	.836	.711	.775
Diff. exp. variance	.017	.159	.035	.043	.056	.047	.064
Inter-rater correlation							
	-	-	.80	.93	.80	.78	-

* Correlations for the NA8 essay-corpus were computed with total score of one rater and correlations for the remaining corpora were computed with the average score of two raters.

**Mean of absolute values

Table 4 gives the regression β weights of the semantic indices in the prediction of raters' scores, and Table 5 gives the regression β weights of NiteRater's 15 scoring dimensions. Also presented in Table 5 are the regression weights obtained for all the scoring dimensions by the stepwise regression procedure, which indicate the marginal contribution of the semantic indices to the predication of raters' scores (entering order and significance).

Marked differences between the indices' β weights were observed both within each corpus and across the corpora. Of the five semantic indices, the three indices based on semantic proximity between essays (MKN, DHS and NTS) were included in four of the six prediction models, while PCA and PRV were included in only one or two of the prediction models. However, though some of the semantic indices may indeed be redundant in some of the prediction models, none of them is redundant in all of the prediction models.

Table 4: Regression β weights of the semantic indices in the prediction of raters' scores by essay-corpora

	NA8-A	NA8-B	IA12	HFL-A	HFL-A	PET-H
NTS	.001	.076	.143	.514	.179	.109
DHS	.038	.600	.081	.128	.143	.201
MKN	.480	.139	.352	.141	.392	.275
PCA	.106	-.184	.148	.226	.261	.199
PRV	.121	-.248	-.192	-.133	-.111	-.204

Table 5: Regression β weights of the semantic indices in the prediction of raters' scores by essay-corpora

	NA8-A			**NA8-B**			**IA12**			**HFL-A**			**HFL-B**			**PET-H**			**Mean**	
	β	SW β	Or.	β	SW β	Or.	β	SW β	Or.	β	SW β	Or.	β	SW β	Or.	β	SW β	Or.	β	SW β
Semantic scoring dimensions (indices)																				
MKN	.383	.406	1	.057	.083	3	.030	.101		.080	.066		.225	.217	3	.238	.197	3	.169	.178
DHS	.045	.038		.226	.238	2	.006	.039		.150	.189	3	.203	.241	1	.177	.250	1	.134	.166
NTS	.022	.030		.041	.038		.097	.139	2	.392	.419	1	.110	.119	5	.060	.074	8	.120	.136
PCA	.151	.179	3	.005	.008		.093	.054		.014	.001		.037	.029		.067	.076		.061	.058
PRV	-.032	-.015		-.134	-.131	5	-.085	-.037		.061	.055	7	-.049	-.047		.011	.036		-.038	-.023
Other NiteRater scoring dimensions																				
Lexical diversity	.140	.173	2	.374	.388	1	.424	.582	1	.370	.358	2	.283	.340	2	.207	.204	2	.300	.341
Spelling errors	-.051	-.053		-.097	-.097	4	-.067	-.066	4	-.065	-.079	4	-.132	-.126	4	-.077	-.081	4	-.081	-.083
Punctuation	.013	.019		.099	.098	7	-.001	-.005		.047	.053	5	.040	.034		.056	.060	7	.042	.043
Vocabulary	.100	.100	4	.005	.007		.052	-.007		.061	.066	6	-.002	.010		.086	.091	6	.050	.044
Syntax complexity	.069	.075	6	.053	.057		.068	.067	5	.007	.002		-.029	-.026		.034	.033		.034	.035
Conjunction diversity	.011	.026		-.005	-.006		.015	.026		.049	.048	8	.059	.069	6	.036	.028		.028	.032
Complement diversity	.027	.033		.013	.020		.031	.039		.002	.004		.019	.013		.060	.081	5	.025	.031
Verb pattern	.004	.012		.093	.095	6	-.040	-.047		.008	.007		.044	.043		-.017	-.010		.015	.017
Style	-.080	-.081	5	.012	.021		.084	.086	3	-.040	-.035		.028	.019		.036	.047		.006	.010
Tense diversity	.005	.001		.017	.011		.007	.004		-.033	-.029		-.019	-.015		.013	.012		-.002	-.003

Notation: β - β regression weights; SW-β - stepwise β regression weights; Or. - order of entry of the scoring dimensions in the stepwise regression

Summary and Discussion

In this study we examined the differential functioning, with regard to predictive validity, of five semantic indices used in the automated scoring of essays. These indices were preliminarily classified into three main categories: (1) indices based on semantic proximity between essays: (i) nearest text score (NTS), (ii) distance from the scores in the highest score category (DHS) and (iii) mean score of the K nearest neighbors (MKN); (2) an index based on Principal Component Analysis (PCA) of semantic similarities; and (3) an index based on prompt-related vocabulary (PRV).

Six heterogeneous essay-corpora were used in the study. Despite the marked heterogeneity among the six essay-corpora with regard to writers' age, mastery of the Hebrew language and essay genre, some consistent results were observed.

The three indices based on semantic proximity between essays (NTS, DHS and MKN) had, on average, the highest correlations with raters' scores (.62-.64) and made a significant contribution to four of the six full prediction models (based on NiteRater's 15 scoring dimensions) generated for the six corpora. The remaining two indices, PCA and PRV, had, on average, somewhat lower correlations with raters' scores (.43 and .40, respectively) and were less often included in the full prediction models (one or two models). However, in spite of their redundancy in some of the prediction models, none of them was redundant in all of the prediction models.

It is interesting to note that using the full set of 15 scoring dimensions had only a marginal contribution to the prediction models in comparison to prediction models based on the semantic indices only (.775 and .733, respectively, on average). As indicated in Table 3, the full set of scoring dimensions accounted on average for about 60% of the score variance, while the prediction model based on the five semantic indices accounts for 54% of the variance.

As can be seen in Table 5, only one scoring dimension – lexical diversity (Cohen & Safran, 2012) – has a larger predictive power than any of the semantic indices. This is also the only index that is included in all six full prediction regression models generated by the stepwise regression procedure. While the standardized beta weight of the semantic indices is between -.038 (for the PRV index) and 0.169 (for the MKN index), the standardized beta weight of the lexical diversity scoring dimension is 0.300.

Thus, all the single-word based indices – the semantic indices and lexical diversity – are the most predictive measures of essay scores. This observation stresses the importance and centrality of the choice of word sets in writing a good essay. As Jacques Barzun has noted (Barzun, 2001, p. 9):

> *"It is proper for the ordinary reader to absorb the meaning of a story or description as if the words were a transparent sheet of glass. But he can do so only because the writer has taken pains to choose and adjust them with care. They were not glass to him, but mere lumps of potential meaning."*

But this finding also raises an old concern regarding the way in which scoring dimensions are put together in predicting essay scores. Using scoring dimensions in an additive way, like a simple regression equation where each variable appears as a single addend in the equation, results in a compensatory, or disjunctive, system, in which high scores on a small set of scoring dimensions can offset a low score on other dimensions. This may lead to a situation in which a "bag", or collection, of well-selected words, with minimal use of syntactical and discourse-related means, can automatically be scored high. This observation draws attention to the need to devise prediction equations that are more conjunctive, i.e., where a high score on one dimension is not a sufficient condition for getting a high total score.

Another concern when developing an AES system is the generality or specificity of the system. Ideally, we would like to have an AES system that needs no supervision, but which is given a set of essays as an input and produces a corresponding vector of scores. In practice, we are not there yet. Actually, even when using human readers, we start by training them on a set of pre-scored essays, or at least reading together a set of essays in order to reach agreement on scoring principles and set the "leniency level".

In a typical AES session, we start by having the system examine a set of pre-scored essays. But then it is always a question of how far can we go in devising scoring dimensions that are general and do not depend too much on a particular set of essays.

References

Barzun, J. (2001). *Simple & Direct*. 4th ed. New York: Harper.

Ben-Simon, A. & Bennett, R.E. (2007). Toward More Substantively Meaningful Automated Essay Scoring. *Journal of Technology, Learning, and Assessment, 6*(1). Retrieved [date] from http://www.jtla.org.

Ben-Simon, A. & Safran, Y. (2012). *The factorial structure of written Hebrew and its application to AES.* Paper presented at the annual meeting of the National Council of Measurement in Education, Vancouver, Canada.

Cohen, Y. & Safran, Y. (2012). *Indices of Lexical Diversity for Automated Essay Scoring*. Paper presented at the annual meeting of the National Council of Measurement in Education, Vancouver, BC.

NiteRater [Computer software]. (2007). Jerusalem: National Institute for Testing & Evaluation.

Page, E.B. (1966). The imminence of grading essays by computer. *Phi Delta Kappan, 48,* 238–243.

Shermis, M. D., & Burstein, J. (2003). *Automated essay scoring: A cross-disciplinary perspective*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Appendix A: NiteRater's writing and scoring dimensions

Writing Dimension	Scoring dimensions	Description
Grammar	Mechanics	Spelling errors (letter-string & lexemes)
Word complexity	Vocabulary	Average frequency of lexemes based on a large corpus of texts
	Lexical diversity	Letter-string & lexeme diversity
	Conjunction diversity	Conjunction diversity
	Complement diversity	Subordinate & preposition diversity
	Tense diversity	Tense diversity
Organization & Development	Verb pattern	Usage of verb patterns
	Style	Possessive/patient suffix
	Punctuation	Based on proportion of very long sentences, and punctuation types and diversity
	Syntax complexity	Preposition- & adjective-to-noun ratios
Topical analysis	Nearest text score	
	Semantic proximity to top essays	Based on similarity (values of cosine correlations) of essay vocabulary to prompt-specific vocabulary of essays in the highest score category
	Mean score of K nearest neighbors	Average score of the K most semantically similar essays. The similarity is computed using LSA, based on the vocabulary of prompt-specific essays
	PCA semantic rank	Based on Principal Component Analysis of the semantic similarities (values of cosine correlations) based on the vocabulary of the essay-corpus
	Prompt-related vocabulary	Based on measuring the semantic distance between each of the essays and the prompt that was given with the essay assignment.