ISSN 1744-1986

Technical Report N° 2005/09

A Research Taxonomy for Latent Semantic Analysis-Based Educational Applications

Debra Trusso Haley
Pete Thomas
Anne De Roeck
Marian Petre

31 May 2005

Department of Computing
Faculty of Mathematics and Computing
The Open University
Walton Hall,
Milton Keynes
MK7 6AA
United Kingdom

http://computing.open.ac.uk

A Research Taxonomy for Latent Semantic Analysis-Based Educational Applications

Debra Trusso Haley, Pete Thomas, Anne De Roeck, Marian Petre
The Computing Research Centre
The Open University
Walton Hall, Milton Keynes MK7 6AA UK
[D.T.Haley, P.G.Thomas, A.DeRoeck, M.Petre] at open.ac.uk
OU Tech Report 2005/09

Abstract

The paper presents a taxonomy that summarises and highlights the major research into Latent Semantic Analysis (LSA) based educational applications. The taxonomy identifies five main research themes and emphasises the point that even after more than 15 years of research, much is left to be discovered to bring the LSA theory to maturity. The paper provides a framework for LSA researchers to publish their results in a format that is comprehensive, relatively compact, and useful to other researchers.

1 Introduction

The major contribution of this paper is a taxonomy resulting from an in-depth, systematic review of the literature concerning latent semantic analysis (LSA) research in the domain of educational applications. The taxonomy presents the key points from a representative sample of the literature. Researchers and developers implementing LSA-based educational applications will benefit by studying the taxonomy because it brings to one place the techniques and evidence reported in the vast LSA literature.

We realized the need for a taxonomy while building an LSA-based assessment system for use in computer science courses. Although our original assessment results were encouraging, they were not good enough for the intended task of summative assessment (Thomas, Haley, et al. '04). We conducted a comprehensive, in-depth literature review to find techniques to improve our system. The taxonomy documents our findings and supports the insights gained by studying the literature.

There exists a great deal of literature on LSA. Some of it involves educational uses (Steinhart '01), some concentrates on LSA theory (Landauer & Dumais '97), and some of the newer articles[1] suggest uses of LSA that go beyond analysing prose (Marcus, Sergeyev, et al. '04, Quesada, Kintsch, et al. '01).

The literature demonstrated that others were having difficulty matching the results reported by the original LSA researchers. We found a lot of ambiguity in various critical implementation details (e.g. weighting function used) as well as unreported details. We speculate that the conflicting or unavailable information explains at least some of the inability to match the success of the original researchers.

This paper is not an LSA tutorial. Readers desiring a basic introduction to LSA should consult the references section.

Section 2 explains the taxonomy, section 3 discusses insights gained by studying the taxonomy, and section 4 concludes with a suggestion for other LSA researchers. The taxonomy is in the appendix.

2 About the taxonomy

2.1. Scope of the taxonomy

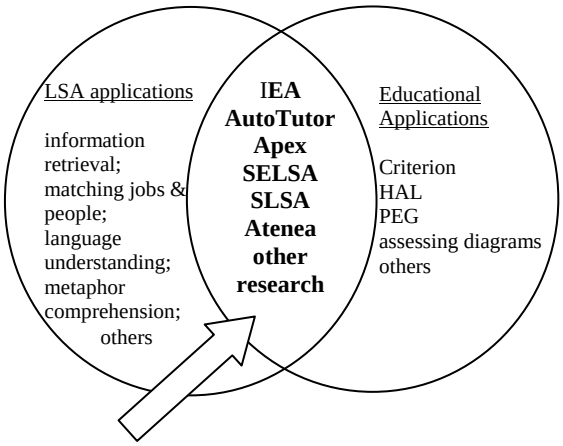

Figure 1. Scope of the Taxonomy – the intersection of LSA and educational applications

The taxonomy summarises and highlights important details from the LSA literature. Because the literature is extensive and our interest is in the

[1] To avoid confusion, we refer to papers in the literature as *articles. Paper* refers to this paper, which includes a taxonomy.

assessment of essays and related artefacts, the taxonomy includes only those LSA research efforts that overlap with educational applications. Therefore, LSA research into such areas as information retrieval (Nakov, Valchanova, et al. '03) and metaphor comprehension (Lemaire & Bianco '03) do not appear in the taxonomy. Similarly, the taxonomy ignores various non-LSA techniques that have been used to assess essays (Burgess, Livesay, et al. '98, Burstein, Chodorow, et al. '03) and diagrams (Anderson & McCartney '03, Thomas, Waugh, et al. in press).

The next subsections discuss the rationale for choosing certain articles over others and the meaning of the headings in the taxonomy.

2.2. Method for choosing articles

The literature review found 150 articles of interest to researchers in the field of LSA-based educational applications. In order to limit this collection to a more reasonable sample, we constructed a citer – citee matrix of articles. That is, each cell entry (i,j) was non blank if article *i* cited article *j*. The articles ranged in date from perhaps the first LSA published article (Furnas, Deerwester, et al. '88), to one published in May 2005 (Perez, Gliozzo, et al. '05). We found the twenty most-cited articles and placed them, along with the remaining 130 articles, in the categories shown in Table 1.

Type of Article	Number in Lit Review	Number in Taxonomy
most cited	20	13
LSA and ed. applications	43	15
LSA but not ed. apps.	13	0
LSI	11	0
theoretical / mathematical	11	0
reviews / summaries	11	0
ed. apps. but not LSA	41	0
Total	**150**	**28**

Table 1. Categories of articles in the literature review and those that were selected for the taxonomy

We chose the twenty most-cited articles for the taxonomy. Some of these most-cited articles were early works explaining the basic theory of Latent Semantic Indexing (LSI).[2] Although not strictly in the scope of the intersection of LSA and educational applications, some of these articles appear in the taxonomy because of their seminal nature. Next, we added articles from the category that combined educational applications with LSA that were of particular interest, either because of a novel domain or technique, or an important result. Finally, we decided to reject certain heavily cited articles because they present no new information pertinent to the taxonomy. This left us with 28 articles in the taxonomy.

2.3. The taxonomy categories

The taxonomy organises the articles involving LSA and educational applications research into three main categories: an *Overview*, *Technical Details*, and *Evaluation*. Figures 2, 3, and 4 show the headings and sub-headings. Most of the headings are self-explanatory; some clarifications are noted in the figures.

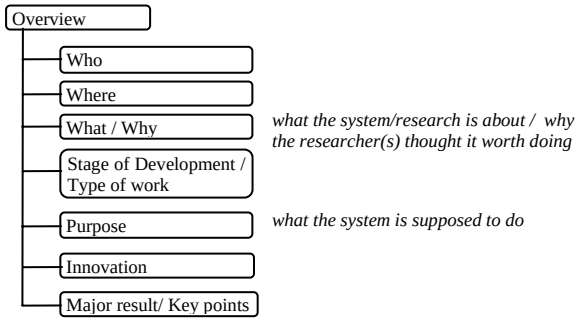

Figure 2. Category A: Overview

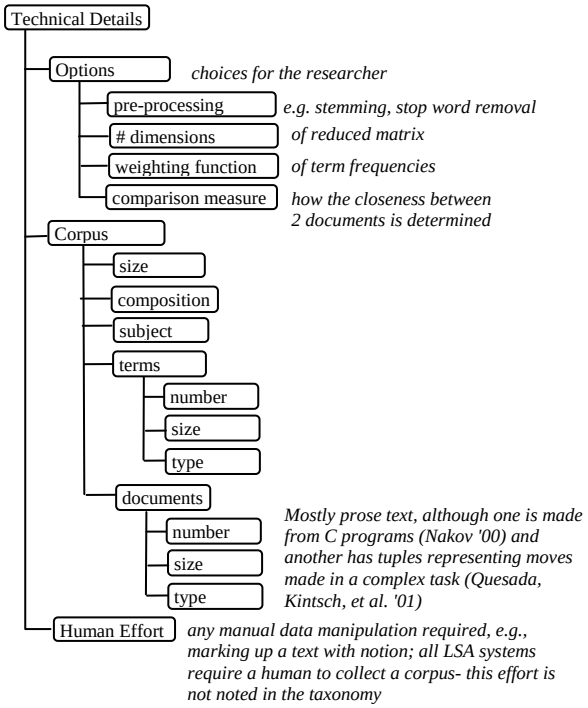

Figure 3. Category B: Technical Details

[2] Researchers trying to improve information retrieval produced the LSI theory. Later, they found that LSI could be useful to analyse text and created the term LSA to describe LSI when used for this additional area.

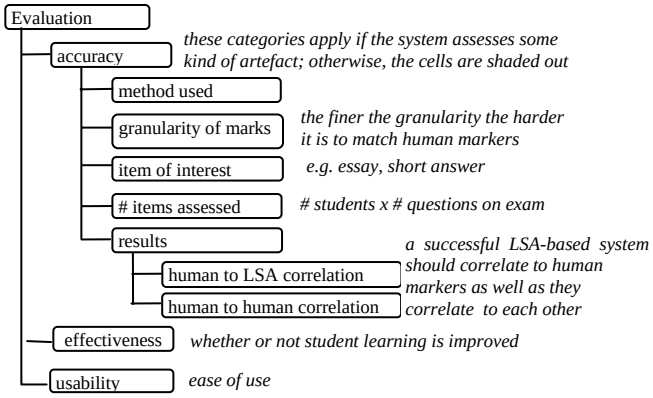

Figure 4. Category C: Evaluation

When looking at the taxonomy, the reader should keep a few points in mind. First, each line presents the data relating to one study. However, one article can report on several studies. In this case, several lines are used for a single article. The cells that would otherwise contain identical information are merged. Second, the shaded cells indicate that the data item is not relevant for the article being categorised. Third, blank cells indicate that we were unable to locate the relevant information in the article.[3] Fourth, the information in the cells was summarised or taken directly from the articles. Thus, the *Reference* column on the far left holds the citation for the information on the entire row. [4]

Organising a huge amount of information in a small space is not easy. The taxonomy in the appendix is based on an elegant solution in (Price, Baecker, et al. '93).

3 Discussion

This section discusses the insights revealed by the taxonomy. Sections 3.1 and 3.2 describe what can be found in the literature, and section 3.3 highlights some of the gaps in the literature.

3.1. Main research themes

A great deal of literature exists about LSA and about educational applications. Even the intersection of these two areas contains many articles. However, the taxonomy reveals five main research themes:

- seminal literature describing the new technique named LSI, which was later renamed to LSA
- attempts to reproduce the results reported in the seminal literature, which for the most part failed to achieve the earlier results
- attempts to improve LSA by adding syntax information
- applications that analyse non-prose text.
- attempts to improve LSA by experimenting with corpus size and composition, weighting functions, similarity measures, number of dimensions in the reduced LSA matrix, and various pre-processing techniques – exactly those items in Category B1 of the taxonomy

3.2. Diversity in the research

The taxonomy reveals a great deal of variety in the research. Researchers work in North America, Europe, and Asia on both deployed applications and continuing research. They use a wide variety of options for pre-processing techniques, number of dimensions in the reduced matrix, weighting functions, and composition and size of corpus. They use English, French, Spanish and Bulgarian corpora. The researchers report their evaluation methods with different specificity.

3.3. Gaps in the literature

The great variety of techniques used by researchers mentioned in the previous section leads to difficulty in comparing the results. Other researchers need to know all of the details to fully evaluate and compare reported results.

Much information is missing on page 2 of the taxonomy – *Category B: Technical Details*. These missing data concern the choices researchers must make when they implement their systems. Page 3 of the taxonomy, *Category C: Evaluation*, shows that some researchers have not evaluated the effectiveness or usability of their deployed systems.

The *Method used* subheading under *Accuracy* in *Category C* is a major area for gaps. Although many researchers report correlations between LSA and human graders, they usually do not mention whether they are using the Pearson, Spearman, or Kendall's tau correlation measure.

The existence of the blank cells in the taxonomy is troubling. They imply that researchers often neglect to report critical information, perhaps due to an oversight or page length restrictions. Nevertheless, the ability to reproduce results would be enhanced if more researchers provided more detailed data regarding their LSA implementations.

[3] Please send any corrections to the first author, who will gladly update the taxonomy.

[4] The *Reference* column contains a pointer to the references section at the end of this paper. Each reference contains a code at the end that corresponds to the entry in the *Reference* column. The entries are of the form *xxxnn* where *xxx* are the initials of up to three of the authors. If capitalised, they represent different authors; if the first is capitalised and the second two are lower case, the article has one author. *nn* is the 2-digit year of publication.

4 Conclusions

We hope that future LSA researchers will keep the taxonomy in mind when presenting their work. Using it will serve two main purposes. First, it will be easier to compare various research results. Second, it will ensure that all relevant details are provided in published articles, which will lead to improved understanding and the continued development and refinement of LSA.

The variability in the results documented in the taxonomy shows that LSA is still something of an art. More than 15 years after its invention, the research issues suggested by (Furnas, Deerwester, et al. '88) are still very much open.

Acknowledgements

The work reported in this study was partially supported by the European Community under the Innovation Society Technologies (IST) programme of the 6th Framework Programme for RTD - project ELeGI, contract IST-002205. This document does not represent the opinion of the European Community, and the European Community is not responsible for any use that might be made of data appearing therein.

References

(Anderson & McCartney '03) M. Anderson & R. McCartney, Diagram processing: Computing with Diagrams. Artificial Intelligence, vol. 145, pp. 181-226, 2003 [AM03].

(Bassu & Behrens '03) D. Bassu & C. Behrens. Distributed LSI: Scalable concept-based information retrieval with high semantic resolution. In Proceedings of Text Mining 2003, a workshop held in conjunction with the Third SIAM Int'l Conference on Data Mining. pp, San Francisco, 2003 [BB03].

(Berry, Dumais, et al. '95) M. W. Berry, S. T. Dumais & G. W. O'Brien, Using linear algebra for intelligent information retrieval. SIAM Review 37, vol. 4, pp. 573-595, 1995 [BDO95].

(Burgess, Livesay, et al. '98) C. Burgess, K. Livesay & K. Lund, Explorations in context space: Words, sentences, discourse. Discourse Processes, vol. 25, pp. 211-257, 1998 [BLL98].

(Burstein, Chodorow, et al. '03) J. Burstein, M. Chodorow & C. Leacock. Criterion Online Essay Evaluation: An Application for Automated Evaluation of Student Essays. In Proc. of the Fifteenth Annual Conference on Innovative Applications of Artificial Intelligence. pp, Acapulco, Mexico, 2003 [BCL03].

(Deerwester, Dumais, et al. '90) S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer & R. Harshman, Indexing by Latent Semantic Analysis. Journal of the American Society for Information Science, vol. 41, pp. 391-407, 1990 [DDF90].

(Dumais '91) S. T. Dumais, Improving the retrieval of information from external sources. Behavioral Research Methods, Instruments & Computers, vol. 23, pp. 229-236, 1991 [Dum91].

(Foltz, Britt, et al. '96) P. W. Foltz, M. A. Britt & C. A. Perfetti. Reasoning from multiple texts: An automatic analysis of readers' situation models. In 18th Annual Cognitive Science Conference. pp 110-115, NJ, 1996 [FBP96].

(Foltz, Kintsch, et al. '98) P. W. Foltz, W. Kintsch & T. K. Landauer, The Measurement of Textural Coherence with Latent Semantic Analysis. Discourse Process, vol. 25, pp. 285-307, 1998 [FKL98].

(Foltz, Laham, et al. '99) P. W. Foltz, D. Laham & T. K. Landauer. Automated Essay Scoring: Applications to Educational Technology. In Proceedings of EdMedia '99. pp, 1999 [FLL99].

(Franceschetti, Karnavat, et al. '01) D. R. Franceschetti, A. Karnavat, J. Marineau, G. L. McCallie, B. A. Olde, B. L. Terry & A. C. Graesser. Development of Physics Text Corpora for Latent Semantic Analysis. In Proc. of the 23rd Annual conference of the Cognitive Science Society. pp, 2001 [FKM01].

(Furnas, Deerwester, et al. '88) G. W. Furnas, S. Deerwester, S. T. Dumais, T. K. Landauer, R. A. Harshman, L. A. Streeter & K. E. Lochbaum. Information retrieval using a singular value decomposition model of latent semantic structure. In Proc. of 11th annual int'l ACM SIGIR conference on Research and development in information retrieval. pp 465-480, 1988 [FDD88].

(Kanejiya, Kumar, et al. '03) D. Kanejiya, A. Kumar & S. Prasad. Automatic Evaluation of Students' Answers using Syntactically Enhanced LSA. In Building Educational Applications Using Natural Language Processing, Proc. of the HLT-NAACL 2003 Workshop. pp 53-60, 2003 [KKP03].

(Kintsch, Steinhart, et al. '00) E. Kintsch, D. Steinhart, G. Stahl, C. Matthews & R. Lamb, Developing summarization skills through the use of LSA-based feedback. Interactive Learning Environments. [Special Issue, J. Psotka, guest editor], vol. 8, pp. 87-109, 2000 [KSS00].

(Landauer '02) T. K. Landauer. On the computational basis of learning and cognition: Arguments from LSA. In The Psychology of Learning and Motivation. edited by B. Ross, 41, pp 43-84, New York, 2002 [Lan02b]

(Landauer & Dumais '97) T. K. Landauer & S. T. Dumais, A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. Psychological Review, vol. 104, pp. 211-240, 1997 [LD97].

(Landauer, Foltz, et al. '98) T. K. Landauer, P. W. Foltz & D. Laham, An introduction to Latent Semantic Analysis. Discourse Processes, vol. 25, pp. 259-284, 1998 [LFL98].

(Landauer, Laham, et al. '97) T. K. Landauer, D. Laham, B. Rehder & M. E. Schreiner. How Well Can Passage Meaning be Derived without Using Word Order? A Comparison of Latent Semantic Analysis and Humans. In Proceedings of the 19th Annual Meeting of the Cognitive Science Society. pp 412-417, 1997 [LLR97].

(Lemaire & Bianco '03) B. Lemaire & M. Bianco. Contextual effects on metaphor comprehension: Experiment and simulation. In Proceedings of the 5th Int'l Conference on Cognitive Modeling (ICCM'2003). pp, Bamberg, Germany, 2003 [LB03].

(Lemaire & Dessus '01) B. Lemaire & P. Dessus, A system to assess the semantic content of student essays. J. of Educational Computing Research, vol. 24, pp. 305-320, 2001 [LD01].

(Marcus, Sergeyev, et al. '04) A. Marcus, A. Sergeyev, V. Rajlich & J. I. Maletic. An Information Retrieval Approach to Concept Location in Source Code. In Proceedings of the 11th IEEE Working Conference on Reverse Engineering. pp 214-223, Delft, The Netherlands, 2004 [MSR04].

(Nakov '00) P. Nakov. Latent Semantic Analysis of Textual Data. In Proceedings of the Int'l Conference on Computer Systems and Technologies. pp, Sofia, Bulgaria, 2000 [Nak00b].

(Nakov, Popova, et al. '01) P. Nakov, A. Popova & P. Mateev. Weight functions impact on LSA performance. In Proc. of the EuroConference Recent Advances in Natural Language Processing (RANLP'01). pp, Tzigov Chark, Bulgaria, 2001 [NPM01].

(Nakov, Valchanova, et al. '03) P. Nakov, E. Valchanova & G. Angelova. Towards Deeper Understanding of the LSA Performance. In Proc. of Recent Advances in Natural Language Processing. pp 311-318, Borovetz, Bulgaria, 2003 [NVA03].

(Olde, Franceschetti, et al. '02) B. A. Olde, D. R. Franceschetti, A. Karnavat & A. C. Graesser. The Right Stuff: Do you need to sanitize your corpus when using Latent Semantic Analysis? In Proceedings of the 24th Annual Meeting of the Cognitive Science Society. pp 708-713, Fairfax, 2002 [OFK02].

(Perez, Gliozzo, et al. '05) D. Perez, A. Gliozzo, C. Strapparava, E. Alfonseca, P. Rodriquez & B. Magnini. Automatic Assessment of Students' free-text Answers underpinned by the combination of a Bleu-inspired algorithm and LSA. In Proceedings of the 18th Int'l FLAIRS Conference. pp, Clearwater Beach, Florida, 2005 [PGS05].

(Price, Baecker, et al. '93) B. A. Price, R. M. Baecker & I. S. Small, A Principled Taxonomy of Software Visualization. Journal of Visual Languages and Computing, vol. 4, pp. 211-266, 1993 [PBS93].

(Quesada, Kintsch, et al. '01) J. Quesada, W. Kintsch & E. Gomez, A computational theory of complex problem solving using the vector space model (part 1): Latent Semantic Analysis, through the path of thousands of ants. Cognitive Research with Microworlds, vol. 43-84, pp. 117-131, 2001 [QKG01a].

(Rehder, Schreiner, et al. '98) B. Rehder, M. E. Schreiner, M. B. W. Wolfe, D. Laham, T. K. Landauer & W. Kintsch, Using Latent Semantic Analysis to assess knowledge: some technical considerations. Discourse Process, vol. 25, pp. 337-354, 1998 [RSW98].

(Steinhart '01) D. J. Steinhart, Summary Street: An intelligent tutoring system for improving student writing through the use of Latent Semantic Analysis. Unpublished PhD Thesis, Department of Psychology, University of Colorado, Boulder, 2001 [Ste01].

(Thomas, Haley, et al. '04) P. Thomas, D. Haley, A. De Roeck & M. Petre. E-Assessment using Latent Semantic Analysis in the Computer Science Domain: A Pilot Study. In Proc. of the eLearning for Computational Linguistics and Computational Linguistics for eLearning Workshop at COLING 2004. pp 38-44, Geneva, 2004 [THD04].

(Thomas, Waugh, et al. in press) P. Thomas, K. Waugh & N. Smith. Experiments in the automatic marking of ER-Diagrams. In Proc. of ITiCSE 05. pp, Lisbon, Portugal, in press [TWS05].

(Wiemer-Hastings '00) P. Wiemer-Hastings. Adding syntactic information to LSA. In 22nd Annual Conference of the Cognitive Science Society. pp 989-993, 2000 [Wie00].

(Wiemer-Hastings & Graesser '00) P. Wiemer-Hastings & A. C. Graesser, Select-a-Kibitzer: A computer tool that gives meaningful feedback on student compositions. Interactive Learning Environments, vol. 8, pp. 149-169, 2000 [WG00].

(Wiemer-Hastings, Wiemer-Hastings, et al. '99) P. Wiemer-Hastings, K. Wiemer-Hastings & A. C. Graesser. Improving an intelligent tutor's comprehension of students with Latent Semantic Analysis. In Artificial Intelligence in Education. pp, Amsterdam, 1999 [WWG99].

(Wiemer-Hastings & Zipitria '01) P. Wiemer-Hastings & I. Zipitria. Rules for Syntax, Vectors for Semantics. In Proc. of the 23rd Cognitive Science Conference. pp, 2001 [WZ01].

(Wolfe, Schreiner, et al. '98) M. B. W. Wolfe, M. E. Schreiner, B. Rehder, D. Laham, P. W. Foltz, W. Kintsch & T. K. Landauer, Learning from text: Matching readers and texts by Latent Semantic Analysis. Discourse Processes, vol. 25, pp. 309-336, 1998 [WSR98].

Category A: Overview

System Name	Refer-ence	A1 Who	A2 Where	A3 What / Why	A4 Stage of Developmen t/ Type of work	A5 Purpose	A6 Innovation	A7 Major Result / Key points
indexing not assessing essays	DDF90	Deerwester, Dumais, Furnas, Landauer, Harshman	U of Chicago, Bellcore, U of W. Ontario	explain new theory that overcomes the deficiencies of term-matching	LSI research	information retrieval	LSI: explains SVD and dimension reduction steps	for Med: for all but the two lowest levels of recall, precision of the LSI method lies well above that obtained with straight-forward term matching; no difference for CISI
	Dum91	Dumais	Bellcore	attempt better LSI results	LSI research	information retrieval	compared different weighting functions	log entropy best weighting function; stemming and phrases showed only 1-5% improvement; 40% better than raw frequency weighting
	BD095	Berry, Dumais, O'Brien	U of Tenn, Bellcore	explain new theory	LSI research	information retrieval	LSI	LSI - completely automatic indexing method using SVD, shows how to do SVD updating of new terms
	FBP94	Foltz, Britt, Perfetti	New Mexico State University, Slippery Rock U, U of Pittsburgh	matching summaries to text read, determine if LSA can work as well as coding propositions	LSA research	text comprehension to evaluate a reader's situation model	matching summaries to text read, analyses knowledge structures of subjects and compares them to those generated by LSA	representation generated by LSA is sufficiently simillar to the readers' situation model to be able to characterize the quality of their essays
	FKL93	Foltz, Kintsch, Landauer		measure text coherence	LSA research		using LSA to measure text coherence	LSA needs a corpus of at least 200 documents; online encyclopedia articles can be added
	LD97	Landauer, Dumais	U of Colorado, BellCore	explain new theory	LSA research			LSA could be a model of human knowledge acquisition
	LLR97	Landauer, Laham, Rehder, Schreiner	U of Colorado	compared essays scores given by readers and LSA, to determine importance of word order	LSA theory	grading essays	investigating the importance of word order; combined quality (cosine) and quantity (vector length)	LSA predicted scores as well as human graders; separating tech and non-technical words made no improvement
	RSW98	Rehder, Shreiner, Wolfe, Laham, Landauer, Kintsch	U of Colorado	explore certain technical issues	LSA research	grading essays	investigated technical vocabulary, essay length, optimal measure of semantic relatedness, and directionality of knowledge in the high dimensional space	nothing to be gained by separating essay into tech and non tech terms cosine and length of essay vector are best predictors of mark
	WSR98	Wolfe,Shreiner, Render, Laham, Foltz, Kintsch, Landauer	U of Colorado, New Mexico State Univ	compared essay scores after reading one of 4 texts	LSA research	select appropriate text	using LSA to select appropriate text	LSA can measure prior knowledge to select appropriate texts
Intelligent Essay Assessor (IEA) http://psych.nmsu.edu.essay	FLL99	Foltz, Landauer, Laham	New Mexico State University, Knowledge Analysis Technologies, U of Colorado	reports on various studies using LSA for automated essay scoring	deployed application for formative assessment	practice essay writing		Over many diverse topics, the IEA scores agreed with human experts as accurately as expert scores agreed with each other.
Summary Street http://www.k-a-t.com/cu.shtml	KSS00	Kintsch, Steinhart, Stahl, LSA Research Group, Matthews, Lamb	U of Colorado, Platt Middle School	helps students summarize essays to improve reading comprehension skills	deployed application for formative assessment	provide feedback on length, topics covered, redundancy, relevance	graphical interface, optimal sequencing of feedback	students produced better summaries and spent more time on task with Summary Street
Summary Street http://www.k-a-t.com/cu.shtml	Ste01	Steinhart	U of Colorado	helps students summarize essays to improve reading comprehension skills	deployed application for formative assessment	provide feedback on length, topics covered, redundancy, relevance	graphical interface, optimal sequencing of feedback	the more difficult the text, the better was the result of using Summary Street, feedback doubled time on task
	Lan02	Landauer	U of Colorado	explaining LSA		LSA general research		LSA works by solving a system of simultaneous equations
AutoTutor	WWG99	Wiemer-Hastings, P., Wiemer-Hastings, K, Graesser, A.	U of Memphis	test theory that LSA can facilitate more natural tutorial dialogue in an intelligent tutoring system (ITS)	deployed application for formative assessment	assess short answers given to Intelligent Tutoring System	tested size and composition of corpus for best LSA results	LSA works best when specific texts comprise at least 1/2 of the corpus and the rest is subject related; works best on essays > 200 words
	Wie00	Wiemer-Hastings	U of Memphis	determine effectiveness of adding syntactic info to LSA	LSA research	assess short answers given to ITS	added syntactic info to LSA	adding syntax decreased the effectiveness of LSA - as compared to Wie99 study
Select-a-Kibitzer	WG00	Wiemer-Hastings, Graesser	U of Memphis	give meaningful feed back on essays using agents	deployed application for formative assessment	assess short answers given to ITS	investigated types of corpora for best results	best corpus is specific enough to allow subtle semantic distinctions within the domain, but general enough that moderate variations in terminology won't be lost

The Taxonomy

SLSA - Structured LSA	WZ01	Wiemer-Hastings, Zipitria	U of Edinburgh	evaluate student answers for use in ITS	LSA research	assess short answers given to ITS	combines rule-based syntactic processing with LSA - adds part of speech	adding structure-derived information improves performance of LSA; LSA does worse on texts < 200 words
	Nak00b	Nakov	Sofia University	explore uses of LSA in textual research	LSA research		uses correlation matrix to display results; analysis of C programs	
	NPM01	Nakov, Popova, Mateev	Sofia University	evaluate weighting function for text categorisation	LSA research	analyse English literature texts	compared 2 local weighting times 6 global weighting methods	log entropy works better than classical entropy
	FKM01	Franceschetti, Karnavat, Marineau, et al	U of Memphis	constructing different types of physics corpora to evaluate best type for an ITS	LSA research for formative assessment	intelligent tutoring	used 5 different corpora to compare vector lengths of words	carefully constructed smaller corpus may provide more accurate representation of fundamental physical concepts than much larger one
	OFK02	Olde, Franceschetti, Karnavat, et al	U of Memphis, CHI Systems	evaluate corpora with different specificities for use in ITS	LSA research for formative assessment	intelligent tutoring	used 5 different corpora to compare essay grades	sanitizing the corpus provides no advantage
Apex	LD01	Lemaire, Dessus	U of Grenoble-II	web-based learning system, automatic marking with feedback	deployed application for formative assessment	provide feedback on topic, outline and coherence		LSA is a promising method to grade essays
	QKG01a	Quesada, Kintsch, Gomez	U of Colorado, U of Grenada	investigate complex problem solving using LSA	CPS and LSA research		represent actions taken in a Microworld as tuples for LSA	LSA is a promising tool for representing actions in Microworlds.
Distributed LSI	BB03	Bassu, Behrens	Telcordia	improve LSI by addressing scalability problem	LSI research	information retrieval	subdivide corpus into several homogeneous subcollections	a divide-and-conquer approach to IR not only tackles its scalability problems but actually increases the quality of returned documents
SELSA	KKP03	Kanejiya, Kumar, Prasad	Indian Institute of Technology	evaluate student answers in an ITS	LSA research	intelligent tutoring	augment each word with POS tag of preceding word, used 2 unusual measures for evaluation: MAD and Correct vs False evaluation	SELSA has limited improvement over LSA
indexing not assessing essays	NVA03	Nakov, Valchanova, Angelova	U of Cal, Berkeley, Bulgarian Academy of Sciences	investigating the most effective meaning of "word"	LSA research	text categorisation	compared various methods of term weighting with NLP pre-processing	linguistic pre-processing (stemming, POS annotation, etc) does not substantially improve LSA; proper term weighting makes more difference
	THD04	Thomas, Haley, DeRoeck, Petre	The Open University	assess computer science essays	LSA research for summative assessment	assess essays	used a very small, very specifc corpus necessitating a small # of dimensions	LSA works ok when the granularity is coarse; need to try a larger corpus
Atenea	PGS05	Perez, Gliozzo, Strapparava, Alfonseca, Rodriquez, Magnini	U de Madrid; Istituto per la Ricerca Scientifica e Technologica	web-based system to assess free-text answers	LSA + ERB research		combine LSA with a BLEU-inspired algorithm; ie combines syntax and semantics	achieves state-of-the-art correlations to the teachers' scores while keeping the language-independence and without requiring any domain specific knowledge

Category B: Technical Details

	B1: Options				B2: Corpus									B3
	B1.1	B1.2	B1.3	B1.4	B2.1	B2.2	B2.3	B2.4: Terms			B2.5: Documents			
								B2.4.1	B2.4.2	B2.4.3	B2.5.1	B2.5.2	B2.5.3	
Refer-ence	pre-processing	# dimensions	weighting function	comparison measure	size	composition	Subject	Number	Size	Type	Number	Size	Type	Human Effort
DDF90	remove 439 stop words (from SMART)	100		cosine		MED	medical abstracts	5,823	words		1,033	average 50 words	title and abstract	
		100		cosine		CISI	information science abstracts	5,135	words		1,460	avg 45 words		
Dum91	remove 439 stop words (from SMART)	60, 100	log entropy			MED, CISI, CRAN, TIME, ADI	various (described in paper)	374 - 5831	words		82 - 1460		title and abstract	
BD095	none	70-100	log entropy	cosine										none
FBP94		100		cosine	27.8 K	21 articles about the Panama Canal; 8 encyclopedia articles, excerpts from 2 books	Panama Canal	4829	word	prose text	607			
		100		cosine		21 articles on the the heart	heart	2,781	words	prose text				
LD97		300	ln(1+freq)/ entropy over all contexts	cosine	4.6M	Grolier's Academic American Encyclopedia		60.7k	word	prose	30.4k	average 151	words	
LLR97	remove 439 stop words	94		cosine vector length		27 articles from Grolier's Academic Amer. Encyclopedia	heart anatomy	3034	word	prose	830	sentence	words	
	no stop words	1500				textbook	psychology -	19,153	words	prose	4,904	paragraphs	words	
RSW98				cosine		27 articles fromGrolier's Academic Amer. Encyclopedia	heart anatomy							separated essays into technical and non technical words
														created subsections of essays
WSR98		100		cosine	17,880	36 encyclopedia articles	heart	3,034	word	prose				
FLL99						a portion of the textbook	psycholinguistics							
							standardised test - opinion essays							
							standardised test - argument essays							
							diverse							
							heart							
KSS00	correct spelling			cosine		specialized texts	heart and lung	17,688	1 word	prose text	830		prose text	no pregraded summaries but mark up text into topics to appear in summaries
							Meso-American history	46,951	1 word	prose text	530		prose text	
Ste01	correct spelling			cosine		general knowledge space	sources of energy		1 word	prose text				
				cosine		specialized texts	heart and lung							
				cosine		specialized texts	Meso-American history							
Lan02		300		cosine										
WWG99		200	log entropy	cosine	2.3 MB	2 complete computer literacy textbooks, ten articles on each of the tutoring topics, entire curriculum script including expected good answers	computer literacy							collect good and bad answers
Wie00	yes , see human effort			cosine			computer literacy		1 tuple	subject - verb - object		1 tuple	subject - verb - object	segmented sentences into subject, verb, object tuples; resolved anaphora; resolved ambiguities with "and" and "or"
WG00														researcher's task to find or create appropriate texts to serve as the corpus and comparison texts

WZ01	removed 440 stop words			cosine	2.3 MB	same as WWG99	computer literacy							segmented sentences into subject, verb, object tuples; resolved anaphora and ambiguities with "and" and "or"
Nak00b	removed 938 stop words	30	log and or entropy			religious texts	religion	20,433			196			
						C programs				C code				
NPM01	removed stop words and those occuring only once	15	6 different	dot product / cosine	974 K	Huckleberry Finn and Adventures of Sherlock Holmes		5534	words	prose	487	2 KB	prose	
FKM01		300		cosine		physics text book and other science text books	physics					paragraph	paragraph	prepare specialised corpora
OFK02		300		cosine		physics text book + related to curriculum script	physics	from 1,564 to 6,536	word	prose	from 416 to 3,778	paragraph	prose	sanitize corpora; write "expectations" for each answer
LD01					290K + size of course text	3 French novels plus course text	sociology of education							no pre-graded essays; mark up text into topics and notions
QKG01a						tuples representing actions in a Microworld	complex problem solving	75565	1	tuple	3441	1 trial		
BB03							various							create a classification scheme for LSI vector spaces
KKP03	removed stop words		log entropy	cosine	2.3M	used Auto tutor corpus	computer literacy	9,194	word	word - part of speech tags	5,596	paragraph	prose	part of speech tagging
NVA03	removed 442 stop words, stemming; POS	0, 10, 220, 40	various			Bulgarian	various - see paper for details							
THD04	none	10	log (no global weighting)	cosine	< 2,000	human marked answers to the essays	computer literacy				17	1 paragraph	prose	none
PGS05			tf-idf			10 different corpora: student answers + text from popular computer magazines								

Category C: Evaluation

	C1: Accuracy						C2	C3
	C1.1	C1.2	C1.3	C1.4	C1.5: results			
					C1.5.1	C1.5.2		
Refer-ence	Method used	Granularity of marks	Item of Interest	Number items assessed	Human to LSA correlation	Human to Human correlation	Effectiveness	Usability
DDF90	evaluate using recall and precision		queries	30				
			queries	35				
Dum91	evaluate using recall and precision							
BD095	evaluate using recall and precision							
FBP94	compare against human graders	100	essay	24	0.68	.367 to .768		
	compared sentences with cosine measure							
LD97			TOEFL - multiple choice test	80	LSA: 64.4%; students: 64.5%			
LLR97	compare against human graders	5	short essay - 250 words	94	0.77	0.77		
	gold standard - a short text written by an expert			94	0.72			
	compare against human graders			273	0.64	0.65		
RSW98	compare with 1 or more target texts		short essay - 250 words	106				
WSR98	compared with 4 texts of increasing difficulty and specificity	5 point scale	essay of about 250 words	106	0.63	0.77		
FLL99	holistic - compare with graded essays		essays		0.8	0.73	average grade 85; after revisions, average grade 92	survey showed 98% of students would definitely or probably use system
				695	0.86	0.86		
				668	0.86	0.87		
				1,205	0.701	0.707		
	holistic			188	0.8	0.83		
KSS00	compare with teacher - provided topic list	10	summary of essay				no sig difference	in classroom 1997-1999; students like immediate feed back
		10	summary of essay	50	0.64	0.69	scores of those using SS for difficult texts significantly higher than those not using SS	in classroom 1997-1999; students like immediate feed back
Ste01		5	summary of essay	108			scores of those using SS for difficult texts significantly higher than those not using SS	
		5		52				
		10		52				
Lan02	holistic, Pearson product-moment correlation coefficient	5 or 10 points	essay	3,500	0.81	0.83		
WWG99	compare against pre-graded answers for completeness and compatibility	2: threshold of .55	short answers average length is 16 words	192	0.49	0.51		
Wie00	compared tuples in student answer with tuples in expected answer				.18, .24, and .4			
WG00								

WZ01	evaluate two texts using cosine							
Nak00b	created correlation matrices							
NPM01	defined precision as ration of chunks from same text to num of chunks at a level							
FKM01	compared vector lengths of words for 5 different corpora							
OFK02	compared experts' marks against LSA marks using a gold standard	5	short answer	1,000	best result about .45	0.72		
LD01	compare with teacher - provided topic list	0-20	essay	31	0.59	0.68	no sig difference between 3 groups - 1 - control - no help 2 - human help provided; 3 - Apex help	
QKG01a	compare LSA with human assessment		moves in Microworld		0.57			
BB03	uses 2 similarity measures							
KKP03	used 20 good answers to each of 8 questions; correlation coefficient, MAD, correct vs false evaluations	2		192	0.47	0.59		
NVA03								
THD04	use Spearman's rho correlation to compare average human grade with LSA grade	8,2,7	essay	18	only 1 set was correlated statistically			
PGS05	Pearson correlation coefficient between humans' scores and Atenea's scores		short essays		0.5	not clear		