MIS Quarterly

ISSUES AND OPINIONS

UNCOVERING THE INTELLECTUAL CORE OF THE INFORMATION SYSTEMS DISCIPLINE[1]

By: **Anna Sidorova**
College of Business Administration
University of North Texas
Denton, TX 76203-5249
U.S.A.
sidorova@unt.edu

Joseph S. Valacich
College of Business Administration
Washington State University
Pullman, WA 99164-4743
U.S.A.
jsv@wsu.edu

Nicholas Evangelopoulos
College of Business Administration
University of North Texas
Denton, TX 76203-5249
U.S.A.
evangeln@unt.edu

Thiagarajan Ramakrishnan
College of Business Administration
University of North Texas
Denton, TX 76203-5249
U.S.A.
ramakrit@unt.edu

Appendix A

Latent Semantic Analysis of MIS Research Abstracts

In this appendix we discuss some technical details of our implementation of latent semantic analysis on a set of 1,615 research abstracts published in *MIS Quarterly*, *Information Systems Research*, and *Journal of Management Information Systems*, in the period 1985 through 2006. The reader who is interested in a theoretical introduction to LSA and an illustration example is referred to Appendix C.

Term Reduction

In accordance with well-accepted information retrieval and text mining procedures (Fox 1992; Frakes 1992; Han and Kamber 2006, pp. 614-622; Harman 1992; Porter 1980), we started the analysis by compiling a list of all terms used in the MIS abstracts (*dictionary*). The 1,615 MIS abstracts produced a dictionary of 9,706 MIS research terms. We then examined and eliminated the unique terms (those appearing in only one document). That reduced the dictionary size to 5,776 terms. As a second step, we removed trivial English words (*stopwords*) such as "and," "the," and so on. Our customized stoplist included a few additional words that we felt should be filtered out from our paper abstracts collection. For example, we added the terms "paper" and "author" as stopwords because they were not expected to add information useful to our analysis. This step reduced the dictionary to 5,410 nontrivial terms. As a third step, we removed term suffices, applying what is commonly known as *term stemming*. For example, we replaced "collaborate," "collaborating," "collaboration," and "collaborative" by "collabor–." This resulted in a dictionary of 3,172 stemmed terms. As a fourth step in term reduction we conducted an initial singular value decomposition (SVD) to

identify and retain the terms that explain a large percentage of variability in the first 100 principal components. We chose to focus on the first 100 factors as they are more likely to represent distinct research areas as opposed to a larger number of factors, say 200 or 300, which would accommodate spurious word usage patterns. Using 100 separate factors, about 42 percent of the terms explained 95 percent of the variance (communality). The remaining 58 percent of the terms that explained only 5 percent of the variance were filtered out because they mostly represented "noise," such as *writing style expressions* that do not have any research significance, but are necessary in order to complete the language structure of the abstract. This *communality filtering* process resulted in a final dictionary of 1,318 terms.

Performing SVD on the Term Frequency Matrix

A tabulation of the retained terms and their appearance in the documents (abstracts) produced a term frequency matrix with 1,318 rows (terms) and 1,615 columns (documents). The raw term frequencies were transformed using a weighting and normalization scheme known as *inverse document frequency (IDF) weighting* or *TF-IDF*, a more traditional approach to term-frequency weighting (Han and Kamber 2006, p. 619; Harman 1992 p. 373; Husbands et al. 2001; Salton 1975; Salton and Buckley 1988). Such transformation promotes the occurrence of rare terms and discounts the influence of more common non-stopwords such as "information" or "system." The transformed term frequency matrix was then subjected to a SVD.[1] More information on the mathematics of TF-IDF as well as SVD is provided in Appendix C. This decomposition produced term eigenvectors, document eigenvectors, and square roots of eigenvalues, known as *singular values*, appearing in descending order. Initially, the total number of factors produced this way was equal to 1,318 factors. In order to identify research areas and research themes at different levels of aggregation we chose to explore several solutions with different number of factors. Those involved 2 through 13, and 100 factors respectively. For each solution, multiplying term eigenvectors by the singular values produced a term-by-factor matrix of term loadings. Similarly, multiplying document eigenvectors by the singular values produced a document-by-factor matrix of document loadings.

Factor Rotations and Factor Loading Thresholds

In classical factor analysis, rotations of factor loadings help with factor interpretations by simplifying the factor/variable associations. Similarly, our latent semantic factors were first rotated by performing varimax rotations on the term loadings in order to simplify the list of terms associated with each factor. To preserve the factor space, the same rotation matrix was used to rotate the document factor loadings. A more extensive discussion on the choice of rotation techniques is presented in Appendix C.

In order to discriminate between significant and insignificant term loadings, a related threshold value was selected based on the probability distribution of term loadings.[2] For the case of a k-factor solution, a threshold associated with a tail probability of $1/k$ was sought. The choice of such a tail probability was related to our decision that, for clarity of interpretation, each term and each document should, on average, load high on only one factor. So for a k-factor solution, this would be accomplished by retaining $1/k$ of the loadings. For example, in the case of 100 factors, the threshold was equal to 0.197 and term loadings with absolute value greater than this number were considered significant. This way, for 100 factors and 1,318 terms, an average of one factor per term and an average of 13.2 terms per factor were expected.

In a similar manner, the probability distribution of document loadings was considered and document loading thresholds were established. For the case of 100 factors, using a tail probability of 1 percent, the appropriate threshold loading was determined to be 0.229 and document loadings with absolute value greater than this number were retained. While on average our approach ensured that each term and each document loaded on one factor, this did not exclude the possibility of cross-loading. This is not an unusual occurrence in factor analysis in general, especially when factors may be closely related, and should be expected when extracting factors from a field whose subareas pull from a common language. In fact, for the 100-factor solution, 25.6 percent of the documents failed to load on any of the 100 factors, 51.6 percent loaded on exactly one factor, 19.9 percent of the documents loaded on two factors, and 2.9 percent of the documents (47 papers) loaded on three factors. No document loaded on more than three factors. Other solutions produced different yet similar cross-loading percentages.

Using the retained term and document loadings, tables of ordered high-loading terms and documents were prepared for each factor solution (see Tables A1 and A2). Coexamination of high-loading terms and documents for each factor solution produced factor labels. Table A3 lists factor labels for the 2, 3, 4, 5, 8, 12, and 13-factor solutions. Labels for the 100-factor solution are listed separately in Table A4. Finally, Table A5 presents the relationship between research areas (5-factor solution) and research topics (100-factor solution) based on cross-loading papers. Detailed results related to the interpretation of the 13-factor solution are presented separately in Appendix B.

---

[1]SVD computations were performed using custom-made Java classes, based on matrix algebra code produced by the National Institute of Standards and Technology (http://math.nist.gov/javanumerics/).

[2]In classical factor analysis, loading thresholds of 0.4 or 0.5 are commonly used. In the case of LSA, because the operations are performed on the covariance matrix and not the correlation matrix, using a fixed threshold is not appropriate.

Table A1. High-Loading Terms for the 5-Factor Solution

F5.#	F5 Label	Top 30 Terms
F5.1	IT and Organizations	plan, strateg, busi, firm, organiz, execut, competit, issu, organ, resourc, success, invest, industri, chang, project, system, coordin, role, implement, innov, integr, advantag, technologi, compani, knowledg, inform, corpor, factor, capabl, valu
F5.2	IS Development	dss, decision, design, system, problem, approach, method, requir, databas, techniqu, methodologi, expert, applic, analysi, tool, support, gener, framework, propos, prototyp, base, knowledg, evalu, structur, softwar, object, solv, maker, environ, plan
F5.3	IT and Individuals	instrum, valid, measur, construct, perceiv, satisfac, usag, accept, reliabl, user, factor, eas, influenc, test, job, variabl, survei, comput, behavior, empir, success, individu, inten, attitud, scale, adop, train, relationship, determin, find
F5.4	IT and Markets	price, market, consum, product, seller, custom, buyer, onlin, cost, invest, electron, servic, supplier, firm, trade, network, valu, transac, trust, profit, internet, commerc, econom, optim, strategi, industri, vendor, increas, offer, reduc
F5.5	IT and Groups	gss, team, meet, task, commun, collabor, outcom, gdss, trust, facilit, work, particip, social, experi, support, interac, instrum, electron, learn, virtual, influenc, comput, individu, behavior, idea, perceiv, affect, em, structur, mediat

Table A2. High-Loading Papers for the 5-Factor Solution

F5.#	F5 Label	Selected High-Loading Papers	Factor Loading
F5.1	IT and Organizations	Johnston and Carrico, MISQ, Mar 1988	0.347
		Premkumar and King, ISR, Jun 1994	0.332
		Gold et al., JMIS, Jul 2001	0.326
		Henderson and Sifonis, MISQ, Jun 1988	0.312
		Karimi and Konsynski, JMIS, Apr 1991	0.300
		Wixom and Watson, MISQ, Mar 2001	0.299
		Dansker et al., MISQ, Jun 1987	0.298
		Reich and Benbasat, MISQ, Mar 2000	0.297
		Van de Ven, MISQ, Jun 2005	0.296
		Main and Short, MISQ, Dec 1989	0.290
F5.2	IS Development	Arinzn, JMIS, Jul 1991	0.415
		Liu et al., JMIS, Jul 1990	0.352
		Prietula and March, ISR,Dec 1991	0.313
		Turban and Watkins, MISQ,Jun 1986	0.301
		Konsynski, JMIS, Jan 1985	0.284
		Nanduri and Rugaber, JMIS, Jan 1996	0.271
		Karimi, JMIS, Jan 1987	0.266
		Purao et al., ISR, Sept 2003	0.246
		Ein-Dor and Spiegler, JMIS, Jul 1995	0.242
		Mantha, MISQ, Dec 1987	0.238
F5.3	IT and Individuals	Davis, MISQ, Sept 1989	0.447
		Doll and Torkzadeh, MISQ, Jun 1988	0.439
		Igbaria et al., MISQ, Sept 1997	0.437
		Agarwal and Karahanna, MISQ, Dec 2000	0.390
		Barki and Hartwick, ISR, Dec 1994	0.342
		Magal, JMIS, Jul 1991	0.341
		McKinney et al, ISR, Sept 2002	0.337
		Heijden, MISQ, Dec 2004	0.329
		Torkzadeh, JMIS, Oct 1988	0.323
		Doll et al., JMIS, Jul 2004	0.300
F5.4	IT and Markets	Grover and Ramanlal, MISQ, Dec 1999	0.485
		Bakos, MISQ, Sept 1991	0.442
		Oh and Lucas, MISQ, Sept 2006	0.441
		Dewan et al., JMIS, Oct 2000	0.405
		Choudhury et al, MISQ, Dec 1998	0.399
		Gallaugher and Wang, MISQ, Dec 2002	0.393
		Yoo et al., JMIS, Jan 2003	0.378
		Kocas, JMIS, Jan 2003	0.375
		Chellappa and Kumar, JMIS, Jul 2005	0.366
		Barua et al., JMIS, Apr 1997	0.326
F5.5	IT and Groups	Dennis and Garfield, MISQ, Jun 2003	0.388
		Miranda and Bostrom, JMIS, Apr 1999	0.384
		Jarvenpaa et al., MISQ, Dec 1988	0.366
		Huang and Wei, JMIS, Jun 2000	0.364
		Ellis et al., JMIS, Jan 1990	0.360
		Kwok et al, JMIS, Jan 2003	0.349
		Dennis et al., MISQ, Jun 2001	0.344
		Reinig, JMIS, Mar 2003	0.331
		Jarvenpaa and Shaw, ISR, Sept 2004	0.309
		Reinig and Shin, JMIS, Oct 2002	0.306

Table A3. Factor Labels and Paper Counts for Selected Factor Solutions

F#	Factor Label	Paper Counts				
		85-06	87-91	92-96	97-01	02-06
	2-Factor Solution Label					
F2.1	IT at organizational and societal levels	1013	224	264	122	224
F2.2	IT at individual and group levels	602	101	122	153	205
	3-Factor Solution Label					
F3.1	IS development	665	177	183	130	109
F3.2	IT at individual and group levels	487	80	116	121	144
F3.3	IT at organizational and societal levels	463	67	86	128	170
	4-Factor Solution Label					
F4.1	IT and organizations	491	130	123	106	96
F4.2	IS development	466	115	127	98	83
F4.3	IT at individual and group levels	392	67	93	90	122
F4.4	IT and markets	270	25	39	76	128
	5-Factor Solution Label					
F5.1	IT and organizations	484	100	127	121	107
F5.2	IS development	397	112	102	72	62
F5.3	IT and individuals	288	54	73	53	90
F5.4	IT and markets	229	18	34	60	115
F5.5	IT and groups	217	23	57	65	70
	8-Factor Solution Label					
F8.1	IT and organizations	398	81	107	101	86
F8.2	IS development	331	82	84	72	61
F8.3	IT and markets	206	19	31	56	97
F8.4	HR and project management	185	53	59	34	18
F8.5	IT adoption and use	178	18	35	52	72
F8.6	IT and groups	146	20	41	43	39
F8.7	Research method	115	21	27	22	41
F8.8	Decision support systems	58	25	11	5	4
	12-Factor Solution Label					
F12.1	IS development	274	67	72	60	45
F12.2	IT management	205	81	57	31	12
F12.3	IT adoption and use	182	15	37	52	76
F12.4	Value of IT	195	20	38	63	69
F12.5	IT and markets	143	11	18	35	77
F12.6	IT for group support	121	20	38	41	21
F12.7	Research method	97	16	22	20	36
F12.8	HR issues in IS	126	38	41	20	11
F12.9	Decision support systems	52	21	12	4	2
F12.10	Project and risk management	86	11	22	25	24
F12.11	Virtual collaboration	71	1	11	12	45
F12.12	IT use by individuals	63	13	21	12	13
	13-Factor Solution Label					
F13.1	IS development	246	63	66	47	43
F13.2	IT management	195	81	55	28	6
F13.3	Value of IT	188	18	36	58	72
F13.4	IT adoption and use	167	14	38	47	66
F13.5	IT and markets	134	10	18	34	70
F13.6	IT for group support	119	20	38	39	21
F13.7	Research methodology	95	16	22	19	35
F13.8	IS field development	130	20	27	36	39
F13.9	Decision support systems	49	21	9	4	2
F13.10	HR issues in IS	75	18	26	11	10
F13.11	Virtual collaboration	81	3	13	15	48
F13.12	Project and risk management	82	11	22	22	22
F13.13	IT use by individuals	54	9	20	10	12

Table A4. Factor Labels and Paper Counts for the 100-Factor Solutions

F100.#	Factor Label	Paper Counts				
		85-06	87-91	92-96	97-01	02-06
F100.1	Decision support systems	38	16	7	3	1
F100.2	Measurement instruments	47	12	13	7	15
F100.3	Individual technology acceptance	28	2	9	7	10
F100.4	Economics of IT	29	1	5	8	15
F100.5	HR issues in IS field	32	7	13	5	3
F100.6	IT for competitive advantage	29	14	4	2	6
F100.7	Virtual teams (leadership in VT)	25	2	4	5	13
F100.8	IT adoption	32	3	6	13	10
F100.9	IS planning	30	13	8	5	0
F100.10	Group support systems	30	0	6	18	6
F100.11	Resource-based view of IT	17	1	0	3	13
F100.12	Communication media	21	4	3	10	4
F100.13	Computer self-efficacy	14	0	5	5	4
F100.14	Database design and data modeling	22	4	7	5	5
F100.15	Group decision support systems	16	8	4	2	1
F100.16	Information systems success	20	2	6	5	5
F100.17	Electronic meeting systems	20	7	8	3	2
F100.18	IS discipline (journals, diversity, etc.)	23	3	7	2	10
F100.19	E-marketplaces and their characteristics	24	1	2	6	15
F100.20	Prototyping (SDLC alternatives)	17	5	1	4	2
F100.21	Knowledge management and knowledge transfer	27	1	3	9	14
F100.22	Role of top management (CEO/CIO)	18	7	6	2	1
F100.23	IT outsourcing	20	0	6	5	9
F100.24	The value of IT investments	24	2	4	7	11
F100.25	IT project failure (management)	19	1	7	7	4
F100.26	EDI and interorganizational systems	14	1	5	7	1
F100.27	Centralized/decentralized IS structure	16	2	6	5	2
F100.28	Critical issues in IS management	16	8	2	3	0
F100.29	Trust in IT-enabled relationships	18	0	1	1	16
F100.30	Software development and maintenance	23	6	5	6	6
F100.31	Power and politics	13	4	2	2	5
F100.32	Customer service	25	2	7	5	11
F100.33	Information centers	13	8	4	1	0
F100.34	Risk management	22	1	7	6	7
F100.35	Web site design	19	0	0	2	17
F100.36	Systems analyst/programmer	24	10	6	2	4
F100.37	Trading systems	14	1	1	11	1
F100.38	Coordination (within and among organizations)	20	3	10	4	2
F100.39	Satisfaction (user and job)	22	7	4	2	7
F100.40	Problem solving	18	5	9	1	1
F100.41	Online consumer (behavior and characteristics)	23	0	1	0	21
F100.42	Electronic brainstorming	19	2	4	8	5
F100.43	Real options and option pricing	14	1	2	5	6
F100.44	Networks (electronic and social)	19	6	4	3	6
F100.45	Executive information systems	18	9	7	2	0
F100.46	Training	19	4	7	3	4
F100.47	Learning and education	20	2	11	3	4
F100.48	Systems development methodologies	17	5	5	2	3
F100.49	Interviews and other knowledge acquisition techniques	14	3	5	2	3
F100.50	End user computing	15	8	5	0	1

Table A4. Factor Labels and Paper Counts for the 100-Factor Solutions (Continued)

F100.#	Factor Label	Paper Counts				
		85-06	87-91	92-96	97-01	02-06
F100.51	Creativity	14	1	4	5	3
F100.52	Languages (programming and query)	21	4	7	5	2
F100.53	Intelligent systems (artificial intelligence)	14	5	4	3	1
F100.54	Supply chain management	13	0	1	2	10
F100.55	Cost-benefit analysis	9	3	1	2	2
F100.56	Industry	11	0	2	6	3
F100.57	Research methodology (qualitative vs. quantitative)	13	4	5	4	0
F100.58	Business process reengineering	18	0	7	10	1
F100.59	Roles (social and organizational)	5	1	1	3	0
F100.60	Neural networks and data-mining	11	0	4	5	2
F100.61	Control	13	2	1	2	5
F100.62	Expert systems	23	8	10	3	1
F100.63	MIS	17	4	4	0	1
F100.64	ERP and IS implementation	14	0	5	1	8
F100.65	Conflict	13	2	7	3	1
F100.66	Task (technology-task fit)	14	1	7	3	3
F100.67	Ethics	14	1	3	6	3
F100.68	Environment (IT-based and organizational)	4	1	0	2	0
F100.69	Object-oriented methodologies	15	4	5	3	3
F100.70	Data and IS Quality	9	0	2	2	5
F100.71	Error detection	12	2	2	5	3
F100.72	Cost and effort estimation	12	3	3	2	3
F100.73	Auctions and other dynamic pricing mechanisms	9	0	1	2	6
F100.74	Graphical information presentation and user interface	11	3	0	3	3
F100.75	IT Innovation	11	1	3	3	4
F100.76	Personalization and privacy	17	4	2	2	7
F100.77	Strategic alignment	7	0	1	3	3
F100.78	Service quality (SERVQUAL instrument)	13	0	6	4	3
F100.79	Attitudes, change and IT adoption	13	2	4	2	4
F100.80	Classification framework	2	0	0	1	1
F100.81	Culture (national and organizational)	13	0	1	6	5
F100.82	Application domain	6	0	2	2	2
F100.83	Negotiations	10	1	2	2	5
F100.84	Collaboration	18	0	3	3	12
F100.85	Communities and digital libraries	4	0	1	0	3
F100.86	Infrastructure	10	0	3	4	3
F100.87	Standards	12	0	0	2	10
F100.88	Security	12	6	2	1	3
F100.89	Public sector (IS in public sector)	9	2	1	4	1
F100.90	Critical Success Factors	6	2	0	0	3
F100.91	Knowledge-based systems and computer-based explanations	8	2	3	2	1
F100.92	User participation in system development	9	1	4	3	1
F100.93	Manufacturing (IT use in manufacturing)	5	2	0	0	3
F100.94	Multimedia (multimedia vs. text environments)	12	1	4	3	4
F100.95	Document management (electronic documents)	6	0	0	2	4
F100.96	Banking (IT in the banking industry)	14	5	2	6	1
F100.97	IT usage	7	1	2	1	3
F100.98	Resource allocation (computer, human and other resources)	5	3	0	1	1
F100.99	Global IT	1	1	0	0	0
F100.100	Internet and social integration of IT	0	0	0	0	0

Table A5. Cross-Loadings between the 5-Factor and the 100-Factor Solutions

F5.#	F5 Label	F100	F100 Label	Paper Count
F5.1	IT and Organization	F100.9	Information system planning	19
		F100.6	IT for competitive advantage	16
		F100.38	Coordination (within and among organizations)	12
		F100.27	Centralized/decentralized IS structure	10
		F100.58	Business process reengineering	10
		F100.22	Role of top management (CEO/CIO)	10
		F100.56	Industry	9
		F100.26	EDI and interorganizational systems	9
		F100.7	Virtual teams (leadership in VT)	9
		F100.24	The value of IT investments	8
		F100.28	Critical issues in IS management	8
		F100.23	IT outsourcing	7
		F100.45	Executive information systems	7
		F100.64	ERP and IS implementation	6
		F100.81	Culture (national and organizational)	6
		F100.43	Real options and option pricing	5
		F100.8	IT adoption	5
		F100.44	Networks (electronic and social)	5
		F100.54	Supply chain management	5
		F100.16	Information systems success	5
		F100.25	IT project failure (management)	5
		F100.21	Knowledge management and knowledge transfer	5
F5.2	IS Development	F100.1	Decision support systems	33
		F100.20	Prototyping (SDLC alternatives)	10
		F100.40	Problem solving	9
		F100.62	Expert systems	8
		F100.52	Languages (programming and query)	8
		F100.14	Database design and data modeling	8
		F100.53	Intelligent systems (artificial intelligence)	6
		F100.9	Information system planning	6
		F100.48	Systems development methodologies	6
		F100.34	Risk management	5
F5.3	IT and Individuals	F100.2	Measurement instruments	30
		F100.3	Individual technology acceptance	18
		F100.5	HR issues in IS field	17
		F100.39	Satisfaction (user and job)	10
		F100.13	Computer self-efficacy	10
		F100.8	IT adoption	9
		F100.50	End user computing	8
		F100.46	Training	7
		F100.33	Information centers	7
		F100.35	Web site design	6
		F100.78	Service quality (SERVQUAL instrument)	5
		F100.29	Trust in IT-enabled relationships	5

Table A5. Cross-Loadings between the 5-Factor and the 100-Factor Solutions (Continued)

F5.#	F5 Label	F100	F100 Label	Paper Count
F5.4	IT and Markets	F100.19	E-marketplaces and their characteristics	19
		F100.4	Economics of IT	15
		F100.37	Trading systems	12
		F100.41	Online Consumer	11
		F100.32	Customer service	11
		F100.26	EDI and interorganizational systems	9
		F100.29	Trust in IT-enabled relationships	8
		F100.43	Real options and option pricing	7
		F100.24	The value of IT investments	6
		F100.23	IT outsourcing	5
		F100.8	IT adoption	5
		F100.44	Networks (electronic and social)	5
		F100.73	Auctions and other dynamic pricing mechanisms	5
F5.5	IT and Groups	F100.7	Virtual teams (leadership in VT)	21
		F100.10	Group support systems	19
		F100.17	Electronic meeting systems	16
		F100.15	Group decision support systems	14
		F100.29	Trust in IT-enabled relationships	11
		F100.84	Collaboration	10
		F100.13	Computer self-efficacy	8
		F100.47	Learning and education	6
		F100.51	Creativity	5
		F100.42	Electronic brainstorming	5

Appendix B

Analysis of the 13-Factor Solution

In this appendix we discuss how the body of IS research is represented through 13 factors (see Table A3). This corresponds to a relatively high level of aggregation, yet offers a more detailed insight into IS research than the 5-factor solution, and may be of interest to some scholars. Tables B1 and B2 show high-loading terms and documents for the 13-factor solution respectively. The 13 factors include IS development (F13.1), IT management (F13.2), value of IT (F13.3), IT adoption and use (F13.4), IT and markets (F13.5), IT for group support (F13.6), measurement instruments (F13.7), IS discipline development (F13.8), decision support systems (F13.9), HR issues in IS (F13.10), virtual collaboration (F13.11), project and risk management (F13.12), and IT use by individuals (F13.13).

Examination of the 13 factors suggests that while some of them represent large research areas, others correspond to subareas or prominent research themes. For clarity and consistency, we will refer to these 13 factors as subareas. Analysis of paper counts of the subareas (see Table A3) suggests that while some subareas declined over the past 20 years, others emerged. Subareas that experienced the most significant decline include IT management (F13.2) and decision support systems (F13.9). The subareas exhibiting significant increase in popularity include value of IT (F13.3), IT and markets (F13.5), and virtual collaboration (F13.11).

Comparison of the 13-factor and the 5-factor solutions illustrates the spin-off of research subareas and prominent research themes. For example, subareas IT management (F13.2) and value of IT (F13.3) correspond to the research area of IT and organizations (F5.1). The decline in F13.2 and the rise of F13.3 compensate for each other, resulting in the relative stability of F5.1. Similarly, IS development (F13.1) and DSS (F13.9) are combined, in the 5-factor solution, under the umbrella of IS development (F5.2). The separation of F5.2 into these subareas illuminates the fact that the decline in the IS development (F5.2) research area is largely attributed to the decline in DSS research.

Table B3 shows the correspondence between subareas (13-factor solution) and research themes (100-factor solution) based on the number of cross-loading documents. As evident from Table B3, most subareas span multiple research themes. For example, IT management (F13.2) includes research related to IS planning (F100.9), the role of top management (F100.22), the structure of the IS function (F100.27), IT for

competitive advantage (F100.6), and so on. Yet, some other subareas, such as decision support systems (F13.9) and HR issues in IS (F13.10), are represented by only one or two research themes.

Table B4 shows how the focus within each subarea evolved over time. For example, research on IS development evolved significantly from DSS and expert systems in the late 1980s, to database design and languages in the early 1990s, and to document management and Web-site design in the 2000s. On the other hand, some subareas maintained constant focus on one or two research themes, and rose or declined together with those themes. For example, the dynamics of IT management (F13.2) subarea mirrors the dynamics of IS planning (F100.9) research theme.

Table B1. High-Loading Terms for the 13-Factor Solution

F13.#	F13 Label	Top 30 Terms
F13.1	IS development	databas, method, design, requir, system, approach, languag, techniqu, problem, network, applic, queri, structur, knowledg, represent, prototyp, expert, integr, tool, object, form, data, propos, describ, environ, base, analysi, gener, methodologi
F13.2	IS management	plan, execut, strateg, success, issu, implement, top, system, corpor, busi, organ, function, competit, resourc, comput, interview, ic, factor, oper, integr, respons, critic, senior, mi, meet, organiz, center, ei, identifi
F13.3	Value of IT	invest, firm, valu, industri, capabl, busi, competit, option, perform, strateg, cost, outsourc, infrastructur, benefit, custom, econom, impact, advantag, resourc, financi, organiz, product, market, relat, edi, supplier, relationship, servic, innov, increa
F13.4	IT adoption and use	adop, perceiv, usag, influenc, behavior, accept, factor, inten, trust, eas, outsourc, theori, social, individu, attitud, belief, innov, test, context, adopt, edi, relationship, construct, empir, theoret, find, percep, determin, variabl, success
F13.5	IT and markets	price, market, seller, consum, onlin, product, buyer, custom, servic, cost, trust, trade, electron, internet, profit, web, offer, supplier, strategi, optim, transac, vendor, commerc, search, softwar, qualiti, network, reduc, marketplac, increas
F13.6	IT for group support	gss, meet, task, gdss, facilit, particip, em, commun, support, idea, electron, tool, outcom, experi, team, interac, structur, satisfac, decision, brainstorm, collabor, gener, creativ, conflict, qualiti, report, social, work, effect, consensu
F13.7	Measurement instruments	instrum, valid, measur, construct, reliabl, satisfac, scale, item, qualiti, dimen, accept, web, servic, eas, perceiv, empir, assess, servqual, test, metric, user, evid, euc, analysi, us, commerc, trust, site, factor, survei
F13.8	IS discipline development	mi, chang, knowledg, issu, field, theori, methodologi, social, framework, perspect, innov, organiz, practic, understand, journal, transform, outsourc, action, scienc, gss, theoret, interpret, approach, analysi, role, literatur, articl, organ, discuss, eme
F13.9	Decision support systems	dss, decision, maker, support, problem, design, compon, system, cognit, es, expert, strategi, effort, solv, featur, aid, strateg, network, creativ, experi, literatur, restrict, assump, activ, subject, behavior, guidanc, involv, theori, improv
F13.10	HR issues in IS	job, satisfac, career, work, profession, analyst, employe, skill, personnel, orient, role, variabl, user, survei, comput, found, turnov, ic, requir, characterist, mi, plan, differ, risk, qualiti, percep, indic, motiv, task, programm
F13.11	Virtual collaboration	team, trust, virtual, collabor, project, commun, knowledg, coordin, integr, web, custom, enabl, mechan, capabl, work, mi, learn, leader, electron, commerc, organiz, busi, role, relationship, invest, structur, compet, perspect, challeng, build
F13.12	Project and risk management	project, risk, softwar, team, control, cost, electron, outsourc, option, estim, invest, approach, failur, coordin, network, methodologi, edi, escal, real, method, success, qualit, practic, conting, problem, custom, schedul, capabl, goal, factor
F13.13	IT use by individuals	train, learn, comput, efficaci, self, euc, educ, collabor, person, program, coordin, edi, web, cognit, skill, student, knowledg, meet, electron, plan, method, behavior, individu, experi, supplier, outcom, higher, buyer, interfac, em

Table B2. High-Loading Documents for the 13-Factor Solution

Factor 13.#	F13 Label	Selected High-Loading Papers	Factor Loading
F13.1	IS development	Leitheiser and March, JMIS, Apr 1996	0.3864
		Ein-Dor and Spiegler, JMIS, Jul 1995	0.3488
		Shibata et al, JMIS, Jan 1997	0.3411
		Storey and Goldstein, MISQ, Mar 1993	0.3403
		Choobineh and Lo, JMIS, Jan 2005	0.3098
		Adam et al, JMIS, Oct 1994	0.3021
		Nanduri and Rugaber, JMIS, Jan 1996	0.2997
		Orman, JMIS, Jan 1989	0.2684
		Konsynski, JMIS, Jan 1985	0.2604
		Janson and Smith, MISQ, Dec 1985	0.2557
F13.2	IS management	Premkumar and King, ISR, Jun 1994	0.4622
		Lederer and Mendelow, JMIS, Oct 1989	0.3754
		Brancheau and Wetherbe, MISQ, Mar 1987	0.3513
		Saunders and Jones, JMIS, Apr 1992	0.3349
		Raghunathan and Raghunathan, JMIS, Jul 1989	0.3142
		Applegate and Elam, MISQ, Dec 1992	0.3087
		Miller and Doyle, MISQ, Mar 1987	0.299
		Wixom and Watson, MISQ, Mar 2001	0.2944
		Reich and Benbasat, MISQ, Mar 2000	0.2905
		Jarvenpaa and Ives, MISQ, Jun 1991	0.2855
F13.3	Value of IT	Dos Santos et al, ISR, Mar 1993	0.4582
		Chatterjee et al, JMIS, Oct 2002	0.412
		Santhanam and Hartono, MISQ, Mar 2003	0.3778
		Kumar, JMIS, Oct 2004	0.3366
		Santos, JMIS, Apr 1991	0.3296
		Sambamurthy et al, MISQ, Jun 2003	0.3276
		Thatcher and Oliver, JMIS, Oct 2001	0.3233
		Subramani, MISQ, Mar 2004	0.3198
		Davern and Kauffman, JMIS, Apr 2000	0.3128
		Ray et al, MISQ, Dec 2005	0.3125
F13.4	IT adoption and use	Igbaria et al, MISQ, Sep 1997	0.4651
		Venkatesh and Morris, MISQ, Mar 2000	0.3814
		Karahanna et al, MISQ, Jun 1999	0.3786
		Davis, MISQ, Sep 1989	0.2834
		Taylor and Todd, ISR, Jun 1995	0.2558
		Burton-Jones and Straub, ISR, Sep 2006	0.2537
		Kaufman et al, ISR, Mar 2000	0.2496
		Thong, JMIS, Apr 1999	0.2297
		Iacovou et al, MISQ, Dec 1995	0.2281
		Moore and Benbasat, ISR, Sep 1991	0.224
F13.5	IT and markets	Oh and Lucas, MISQ, Sep 2006	0.5205
		Grover and Ramanlal, MISQ, Dec 1999	0.4653
		Bakos, MISQ, Sep 1991	0.4644
		Dewan et al, JMIS, Oct 2000	0.4375
		Sen et al, JMIS, Jul 2006	0.4347
		Kauffman and Wang, JMIS, Oct 2001	0.4
		Choudhury et al, MISQ, Dec 1998	0.3985
		Bakos et al, ISR, Dec 2005	0.3703
		Gupta et al, JMIS, Jul 2000	0.3495
		Yoo et al, JMIS, Jan 2003	0.3491

Table B2. High-Loading Documents for the 13-Factor Solution (Continued)

Factor 13.#	F13 Label	Selected High-Loading Papers	Factor Loading
F13.6	IT for group support	Miranda and Bostrom, JMIS, Apr 1999	0.5458
		Dennis et al, MISQ, Jun 2001	0.5029
		Dennis et al, JMIS, Jul 1997	0.4496
		Huang and Wei, JMIS, Oct 2000	0.4427
		Nunamaker et al, JMIS, Jan 1997	0.3849
		Dennis, MISQ, Dec 1996	0.3819
		Zigurs and Buckland, MISQ, Sep 1998	0.3752
		George et al, ISR, Dec 1990	0.3739
		Jarvenpaa et al, MISQ, Dec 1988	0.3122
		Sambamurthy and Poole, ISR, Sep 1992	0.3114
F13.7	Measurement instruments	Straub, MISQ, Jun 1989	0.5665
		Doll and Torkzadeh, MISQ, Jun 1988	0.5418
		Chang and King, JMIS, Jul 2005	0.5096
		Szajna, MISQ, Sep 1994	0.5025
		Pitt et al, MISQ, Jun 1995	0.4819
		Jiang et al, MISQ, Jun 2002	0.4683
		Torkzadeh and Dhillon, ISR, Jun 2002	0.4566
		Byrd and Turner, JMIS, Jul 2000	0.456
		Doll et al, MISQ, Dec 1994	0.4505
		Torkzadeh, JMIS, Oct 1988	0.4451
F13.8	IS discipline development	Culnan and Swanson, MISQ, Sep 1986	0.3126
		Orlikowski and Barley, MISQ, Jun 2001	0.2618
		Alavi and Carlson, JMIS, Apr 1992	0.2494
		Robey and Boudreau, ISR, Jun 1999	0.2429
		Nunamaker et al, JMIS, Jan 1991	0.2337
		Orlikowski, ISR, Mar 1996	0.231
		Culnan, MISQ, Sep 1987	0.229
		Vessey et al, JMIS, Oct 2002	0.2263
		Agarwal and Lucas, MISQ, Sep 2005	0.2161
		Gregor, MISQ, Sep 2006	0.2101
F13.9	Decision support systems	Goul et al, JMIS, Apr 1986	0.6329
		Kasper, ISR, Jun 1996	0.6284
		Silver, ISR, Mar 1990	0.6106
		Todd and Benbasat, MISQ, Sep 1992	0.6046
		Arinzn, JMIS, Jul 1991	0.5845
		Hogue, JMIS, Jul 1987	0.5365
		Todd and Benbasat, ISR, Dec 1999	0.5339
		Todd and Benbasat, ISR, Jun 1991	0.5213
		Remus and Kottemann, MISQ, Dec 1986	0.521
		Goslar and Mann, JMIS, Jul 1986	0.5206
F13.10	HR issues in IS	McMurtrey et al, JMIS, Oct 2002	0.5009
		Igbaria and Guimaraes, JMIS, Apr 1993	0.4997
		Igbaria et al, MISQ, Jun 1991	0.4761
		Igbaria et al, MISQ, Jun 1994	0.4581
		Igbaria and Baroudi, MISQ, Mar 1995	0.4327
		Yoon and Guimaraes, JMIS, Jul 1995	0.4223
		Guimaraes and Igbaria, ISR, Sep 1992	0.4038
		Green, MISQ, Jun 1989	0.3855
		Li and Sham, JMIS, Apr 1991	0.385
		Millman and Hartwick, MISQ, Dec 1987	0.3811

Table B2. High-Loading Documents for the 13-Factor Solution (Continued)

Factor 13.#	F13 Label	Selected High-Loading Papers	Factor Loading
F13.11	Virtual collaboration	Malhotra et al, MISQ, Jun 2001	0.5234
		Jarvenpaa et al, JMIS, Apr 1998	0.5023
		Piccoli and Ives, MISQ, Sep 2003	0.4869
		Pauleen, JMIS, Jan 2004	0.4318
		Kayworth and Leidner, JMIS, Jan 2002	0.3888
		Brown et al, JMIS, Apr 2004	0.3648
		Guinan et al, ISR, Jun 1998	0.3467
		Griffith et al, MISQ, Jun 2003	0.3347
		Leimeister et al, JMIS, Apr 2005	0.3326
		Paul, JMIS, Apr 2006	0.3166
F13.12	Project and risk management	Barki et al, JMIS, Apr 2001	0.5208
		Barki et al, JMIS, Oct 1993	0.4123
		Keil and Robey, JMIS, Apr 1999	0.3618
		Benaroch et al, MISQ, Dec 2006	0.3517
		Keil et al, MISQ, Jun 2000	0.3405
		Hu et al, JMIS, Jul 1998	0.3328
		Schmidt et al, JMIS, Apr 2001	0.3324
		Nidumolu, ISR, Sep 1995	0.3236
		Choudhury and Sabherwal, ISR, Sep 2003	0.2847
		Deephouse et al, JMIS, Jan 1996	0.2836
F13.13	IT use by individuals	Compeau and Higgins, ISR, Jun 1995	0.5083
		Yi and Davis, ISR, Jun 2003	0.4496
		Davis and Davis, JMIS, Oct 1990	0.438
		Webster and Martocchio, MISQ, Jun 1992	0.4067
		Simon et al, ISR, Dec 1996	0.3867
		Piccoli et al, MISQ, Dec 2001	0.3508
		Kang and Santhanam, JMIS, Jan 2004	0.3502
		Alavi and Leidner, ISR, Mar 2001	0.337
		Compeau and Higgins, MISQ, Jun 1995	0.3214
		Compeau et al, MISQ, Jun 1999	0.3111

Table B3. Cross-Loadings between the 13-Factor and the 100-Factor Solutions

F13.#	F13 Label	F100	F100 Label	Paper Count
F13.1	IS development	F100.52	Languages (programming and query)	11
		F100.20	Prototyping (SDLC alternatives)	10
		F100.14	Database design and data modeling	9
		F100.40	Problem solving	8
		F100.44	Networks (electronic and social)	6
		F100.1	Decision support systems	5
		F100.53	Intelligent systems (artificial intelligence)	5
		F100.94	Multimedia (multimedia vs. text environments)	4
		F100.95	Document management (electronic documents)	4
		F100.62	Expert systems	4
		F100.61	Control	4
F13.2	IS management	F100.9	Information system planning	19
		F100.45	Executive information systems	13
		F100.6	IT for competitive advantage	11
		F100.22	Role of top management (CEO/CIO)	11
		F100.33	Information centers	10
		F100.28	Critical issues in IS management	8
		F100.27	Centalized / Decentralized IS structure	6
		F100.63	MIS	5
		F100.64	ERP and IS implementation	4
		F100.16	Information systems success	4
		F100.44	Networks (electronic and social)	4
		F100.38	Coordination (within and among organizations)	4
F13.3	Value of IT	F100.43	Real options and option pricing	11
		F100.6	IT for competitive advantage	11
		F100.24	The value of IT investments	11
		F100.23	IT outsourcing	10
		F100.26	EDI and interorganizational systems	8
		F100.4	Economics of IT	7
		F100.56	Industry	6
		F100.32	Customer service	6
		F100.38	Coordination (within and among organizations)	6
		F100.54	Supply chain management	5
		F100.19	Electronic marketplaces and their characteristics	4
F13.4	IT adoption and use	F100.3	Individual technology acceptance	18
		F100.8	IT adoption	15
		F100.23	IT outsourcing	12
		F100.29	Trust in IT-enabled relationships	10
		F100.13	Computer self-efficacy	9
		F100.26	EDI and interorganizational systems	6
		F100.92	User participation in system development	4
		F100.31	Power and politics	4
		F100.79	Attitudes, change and IT adoption	4
		F100.2	Measurement instruments	4
		F100.75	IT innovation	4
		F100.41	Online consumer (behavior and characteristics)	4

Table B3. Cross-Loadings between the 13-Factor and the 100-Factor Solutions (Continued)

F13.#	F13 Label	F100	F100 Label	Paper Count
F13.5	IT and markets	F100.19	Electronic marketplaces and their characteristics	19
		F100.4	Economics of IT	15
		F100.41	Online consumer (behavior and characteristics)	11
		F100.37	Trading systems	11
		F100.32	Customer service	7
		F100.29	Trust in IT-enabled relationships	6
		F100.73	Auctions and other dynamic pricing mechanisms	4
		F100.35	Web site design	4
		F100.96	Banking (IT in the banking industry)	4
F13.6	IT for group support	F100.10	Group support systems	19
		F100.17	Electronic meeting systems	18
		F100.15	Group decision support systems	15
		F100.42	Electronic brainstorming	11
		F100.51	Creativity	5
F13.7	Measurement instruments	F100.2	Measurement instruments	30
		F100.3	Individual technology acceptance	9
		F100.78	Service quality (SERVQUAL instrument)	7
		F100.35	Web site design	5
		F100.50	End user computing	4
F13.8	IS discipline development	F100.18	IS Discipline (journals, diversity, etc)	9
		F100.63	MIS	8
		F100.28	Critical issues in IS management	5
		F100.31	Power and politics	5
		F100.81	Culture (national and organizational)	5
		F100.6	IT for competitive advantage	5
F13.9	Decision support systems	F100.1	Decision support systems	33
F13.10	HR issues in IS	F100.5	HR issues in IS field	22
		F100.39	Satisfaction (user and job)	9
F13.11	Virtual collaboration	F100.7	Virtual teams (leadership in VT)	21
		F100.29	Trust in IT-enabled relationships	16
		F100.84	Collaboration	8
		F100.38	Coordination (within and among organizations)	5
		F100.27	Centalized/decentralized IS structure	4
F13.12	Project and risk management	F100.34	Risk management	15
		F100.25	IT project failure (management)	13
		F100.7	Virtual teams (leadership in VT)	9
		F100.43	Real options and option pricing	7
		F100.38	Coordination (within and among organizations)	5
		F100.23	IT outsourcing	4
		F100.61	Control	4
		F100.72	Cost and effort estimation	4
F13.13	IT use by individuals	F100.46	Training	15
		F100.13	Computer self-efficacy	10
		F100.47	Learning and education	9

Table B4 Cross-Loadings between the 13-Factor and 100-Factor Solutions

	1987–1991		1991–1996		1997–2001		2002–2006	
	Theme	*Ct.*	*Theme*	*Ct.*	*Theme*	*Ct.*	*Theme*	*Ct.*
F13.1 IS Development	Problem solving	5	Database design	5	Languages	3	Document management	2
	Networks	5	Languages	4	Document management	2	Web site design	2
	Decision support systems	3	Problem solving	3	Decision support systems	2	Collaboration	2
	Prototyping	2	Intelligent systems (AI)	2	Prototyping	2		
	Expert systems	2	Multimedia	2	BPR	2		
	Intelligent systems (AI)	2						
F13.2 IT Management	IS planning	9	IS planning	6	IS planning	2		
	Information centers	7	Executive IS	6	IS adoption	2		
	IT for compet. advantage	7	Role of top mgmt. (CEO/CIO)	3				
	Executive IS	6	ERP implementation	3				
	Critical issues in IS mgmt.	5	Centr./decentr. IS struct.	3				
	Role of top mgmt. (CEO/CIO)	5						
F13.3 Value of IT	IT for competitive advantage	5	Value of IT investments	3	Real options	4	IT outsourcing	4
			Coordination	3	EDI and interorg. systems	4	Value of IT investments	4
			EDP and interorg. systems	3	Value of IT investments	3	Economics of IT	3
			Customer service	2	Industry	3	IT for compet. advantage	3
			E-marketplaces	2			Real options	3
			IT outsourcing	2				
			IT for compet. advantage	2				
F13.4 IT Adoption and Use	Measurement instruments	2	Indiv. tech. acceptance	6	IT adoption	10	Trust	8
			Computer self-efficacy	3	Indiv. tech. acceptance	7	IT outsourcing	5
			User participation	3	Computer self-efficacy	3	Online consumer	3
					IT outsourcing	3		
					EDI and inteorg. systs.	3		
F13.5 IT and Markets			Customer service	3	Trading systems	8	Economics of IT	8
			E-marketplaces	2	E-marketplaces	4	E-marketplaces	8
					Economics of IT	3	Online consumer	7
					Banking	2	Trust	5
					IT adoption	2	Web site design	4
F13.6 IT for Group Support	Electronic meeting systems	7	Electronic meeting systems	8	Group support systems	13	Group support systems	4
	GDSS	7	GDSS	4	Electronic brainstorming	6	Collaboration	2
			Creativity	2	Electronic meeting systems	3	Electronic brainstorming	2
			Group support systems	2	GDSS	2	Creativity	2
			User participation	2			Virtual teams	2
			Learning & education	2				
			Electronic brainstorming	2				
F13.7 Measurement Instruments	Measurement instruments	6	Measurement instruments	11	Service qual. (SERVQUAL)	4	Measurement instruments	10
	End-user computing	2	Indiv. tech. acceptance	3	Measurement instruments	3	Web site design	4
			Service qual. (SERVQUAL)	2	Indiv. tech. acceptance	3	Online consumer	3
					Data & IS quality	2	Trust	3
F13.8 IS Discipline Development	Critical issues in IS mgmt.	3	MIS	3	Knowledge management	2	IT for compet. advantage	2
	IS discipline	3	Learning & education	2	Culture	2		
	Systems dev. methodologies	2			Power and politics	2		
	MIS	2						
F13.9 Decision Support Systems	Decision support systems	13	Decision support systems	5	Decision support systems	3		
F13.10 HR Issues in IS	HR issues in IS	3	HR issues in IS field	10	HR issues in IS field	4	HR issues in IS field	3
	Information centers	2	Satisfaction (user and job)	3				
	Satisfaction	2						
F13.11 Virtual Collaboration			Virtual teams	3	Virtual teams	5	Trust	11
			Centr./decentr. IS struct.	2	Coordination	2	Virtual teams	10
			Coordination	2	Centr./decentr. IS struct.	2	Collaboration	5
							IT project management	2
F13.12 Project and Risk Management	Cost and effort estimation	2	Coordination	5	IT project management	6	Control	3
			IT project management	4	Risk management	5	Risk management	3
			Risk management	3	Real options	2	Virtual teams	3
			Virtual teams	3	Virtual teams	2		
F13.13 IT Use by Individuals	Training	4	Training	5	Computer self-efficacy	3	Training	3
	End-user computing	2	Learning & education	5	Training	2	Computer self-efficacy	2
			Computer self-efficacy	4	Learning & education	2	Learning & education	2

Appendix C

Introduction to Latent Semantic Analysis

This appendix serves as a brief introduction to latent semantic analysis (LSA) as it applies to exploratory summarization of document collections. LSA allows for computerized extraction of concepts hidden in text data and holds great promise for free text analysis, as it allows for identification of key common themes in a collection of documents without an *a priori* theoretical model, based solely on word usage within the documents. Because researchers usually develop discipline-specific vocabularies and rely on common word patterns to address specific research topics, latent semantic factors are likely to reveal such topics. Some mathematical details are presented in the next section, followed by a small but insightful illustration example.

The Mathematics of LSA

Singular Value Decomposition. The mathematics of LSA are based on a matrix operation called singular value decomposition (SVD), applied to a term-by-document matrix holding the frequency of use of all terms in all documents in a given collection. Given a $t \times d$ matrix $\mathbf{X}$ of terms by documents containing raw or weighted term frequencies, with $\text{rank}(\mathbf{X}) = r \leq \min(t,d)$, the SVD of $\mathbf{X}$ is given by $\mathbf{X} = \mathbf{TSD}^T$, where $\mathbf{T}$ is the $t \times r$ matrix of eigenvectors of the square symmetric matrix of term covariances $\mathbf{XX}^T$, $\mathbf{D}$ is the $d \times r$ matrix of eigenvectors of the square symmetric matrix of document covariances $\mathbf{X}^T\mathbf{X}$, and $\mathbf{S}$ is an $r \times r$ diagonal matrix containing the square roots of eigenvalues (called singular values) of both $\mathbf{XX}^T$ and $\mathbf{X}^T\mathbf{X}$. Then, **TS** are the factor loadings for terms and **DS** are the factor loadings for documents. Retaining a small number of significant factors k, $\mathbf{X}$ can be represented by its least squares approximation $\hat{\mathbf{X}} = \mathbf{T}_k\mathbf{S}_k\mathbf{D}_k^T$. See Berry et al. (1995) and Park et al. (2001) for more detailed discussions of matrix rank reduction and SVD.

Inverse Document Frequency (TF-IDF) Transformation. *Inverse document frequency* transformation, commonly referred to as TF-IDF, is a traditional approach to term-frequency weighting (Han and Kamber 2006, p. 619; Harman 1992, p. 373; Husbands et al. 2001; Salton 1975; Salton and Buckley 1988). As a part of the *TF-IDF* transformation, the raw term frequencies are replaced by the product $w_{ij} = tf_{ij} * idf_i$, where $idf_i = \log_2(N/n_i) + 1$, N is the number of documents in the collection, tf_{ij} is the raw term frequency of term i in document j, n_i is the term frequency of term i in the entire collection of documents, and the inverse document frequency (IDF) idf_i serves as a metric of rarity of term i in the entire collection of documents. Such transformation promotes the occurrence of rare terms and discounts the influence of more common non-stopwords such as "information" or "system." After weighting, the term frequencies are typically also *normalized* so that the sum of squared transformed frequencies of all term occurrences within each document is equal to one (Harman 1992, p. 375; Salton and Buckley 1988). A number of alternative term frequency transformations have been proposed in the literature. Some of them, notably the *log-entropy* transformation (Chew et al. 2007; Dumais 1991), have been found to outperform TF-IDF for purposes of information retrieval and document classification. For purposes of document summarization, however, one may want to try more than one transformation to ensure interpretative consistency.

Factor Rotations. Rotations of loadings can be performed in a number of ways. One way would be to first rotate the term loadings $\mathbf{L}_T = \mathbf{T}_k\mathbf{S}_k$ into $\mathbf{L}_T\mathbf{M}$, by multiplying them by a rotation matrix $\mathbf{M}$ according to some term structure simplification criterion and then reciprocate with the rotation of the document loadings matrix $\mathbf{L}_D = \mathbf{D}_k\mathbf{S}_k$ into $\mathbf{L}_D\mathbf{M}$. A second way to perform loading rotations would be to first rotate the document loadings $\mathbf{L}_D$ and then reciprocate with the rotation of $\mathbf{L}_T$. A third way would be to implement a matching rotation technique (Cheng and Dunkerton 1995; Kiers 1997; Peay 1988), that combines $\mathbf{L}_T$ and $\mathbf{L}_D$, for example, by rotating $\begin{pmatrix} \mathbf{T}_k\mathbf{S}_k^{1/2} \\ \mathbf{D}_k\mathbf{S}_k^{1/2} \end{pmatrix}$. In our paper we apply *varimax* rotations (Crawford and Ferguson 1970) on the term factor loadings alone. The rationale behind this choice is that a simpler term structure will facilitate factor interpretation in a more straightforward manner than a simpler document structure. The same rotations are subsequently applied to the document structure so that both terms and documents maintain the same factor space representation.

An Illustration Example

In order to illustrate LSA, we consider the small collection of documents shown on Table C1. These are six selected titles of papers published in *MIS Quarterly (MISQ)* between 1998 (Volume 22) and 2007 (Volume 31). The first task is to create a dictionary of relevant terms for our document indexing purposes. Following generally accepted information retrieval practices, trivial words of the English language, such as "and" or "the," are ignored. Terms that appear only once in the collection are also ignored, since they cannot contribute to the formation of patterns. The dictionary now consists of only six terms: {*acceptance*, *information*, *media*, *model*, *selection*, *technology*}. Table C1 marks the occurrence

of these six terms within the documents by boldfacing. Table C2 shows the raw term frequencies for each of the six documents, organized in a 6 × 6 term-by-document matrix. Table C3 shows the term frequency matrix after a transformation based on *inverse document frequencies* (*TF-IDF* transformation), which penalizes frequent terms and promotes rare terms. This matrix is subjected to an SVD. Figure C1 shows a *scree* plot of the six eigenvalues produced by this analysis.[3] Based on this plot, keeping the first two principal components seems appropriate.

Interpretation of these first two factors is the next step in our analysis. Table C4 shows the term loadings before and after a varimax rotation. Factor F1 appears to be mostly related to terms {*acceptance*, *information*, *technology*}, and somewhat related to {*model*}, whereas factor F2 appears to be primarily related to terms {*media*, *model*, *selection*}. Table C5 shows the document loadings before and after the same varimax rotation that was applied to the term loadings (i.e., using the same rotation matrix). Factor F1 loads high on documents D2, D4, and D5. Factor F2 loads high on documents D1, D3, and D6. Reading again the corresponding titles from Table C1, it is plausible to infer that factor F1 is about information technology acceptance and factor F2 about media selection.

Table C1. Titles of Selected Articles Published in *MIS Quarterly*

ID	Title	MISQ Reference
D1	An Investigation of **Media Selection** Among Directors and Managers: From "Self" to "Other" Orientation	22:3, pp. 335-362
D2	User **Acceptance** of **Information Technology**: Toward a Unified View	27:3, pp. 425-478
D3	Unraveling the Temporal Fabric of Knowledge Conversion: A **Model** of **Media Selection** and Use	30:1, pp. 99-114
D4	Influence Processes for **Information Technology Acceptance**: An Elaboration Likelihood **Model**	30:4, pp. 805-825
D5	Reconceptualizing Compatibility Beliefs in **Technology Acceptance** Research	30:4, pp. 781-804
D6	Communcation **Media** Repertories: Dealing with the Multiplicity of **Media** Choices	31:2, pp. 267-293

Table C2. Raw Term Frequencies for the Titles in Table C1, Organized as a 6 × 6 Matrix

	Document					
Term	D1	D2	D3	D4	D5	D6
acceptance	0	1	0	1	1	0
information	0	1	0	1	0	0
media	1	0	1	0	0	2
model	0	0	1	1	0	0
selection	1	0	1	0	0	0
technology	0	1	0	1	1	0

Table C3. Transformed Term Frequencies After TF-IDF Weighting

	Document					
Term	D1	D2	D3	D4	D5	D6
acceptance	0	0.471	0	0.377	0.707	0
information	0	0.746	0	0.598	0	0
media	0.346	0	0.253	0	0	1
model	0	0	0.684	0.598	0	0
selection	0.938	0	0.684	0	0	0
technology	0	0.471	0	0.377	0.707	0

[3]The fact that the sixth eigenvalue shown in Figure C1 is equal to zero is not surprising, since terms *acceptance* and *technology* always appear together in this small collection of documents. This causes rows 1 and 6 of the term frequency matrix to be identical, resulting in its rank reduction.

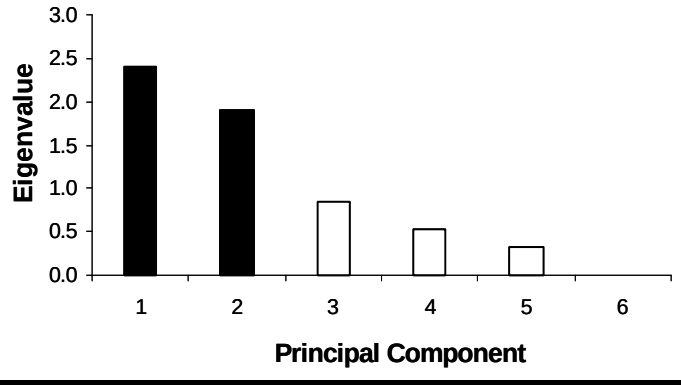

Figure C1. Scree Plot

Table C4. Term Loadings Before and After Varimax Rotation

Term	Unrotated		Rotated	
	F1	F2	F1	F2
acceptance	0.830	0.302	**0.883**	–0.006
information	0.769	0.227	**0.880**	–0.055
media	0.211	–0.764	–0.068	**–0.790**
model	0.533	–0.348	0.378	**–0.512**
selection	0.335	–0.981	–0.028	**–1.037**
technology	0.830	0.302	**0.883**	–0.006

Table C5. Document Loadings Before and After Varimax Rotation

Term	Unrotated		Rotated	
	F1	F2	F1	F2
acceptance	0.250	–0.860	–0.006	**–0.893**
information	0.873	0.329	**0.933**	0.004
media	0.417	–0.799	0.112	**–0.895**
model	0.905	0.113	**0.888**	–0.210
selection	0.756	0.310	**0.817**	0.027
technology	0.136	–0.554	–0.065	**–0.567**

In order to better understand how terms and documents are represented in the latent semantic factor space, let us now examine how term frequencies are approximated by reconstructing the term frequency matrix after retaining the first two principal components (see Table C6). Using this two-factor space, the term frequencies appear modified from their original values in Table C3. For example, even though the term *selection* did not appear at all in document D6 (see Table C3, column 6), it now does, and its frequency is quite high (see Table C6, column 6, highlighted cell). After examining this term-document structure and considering the statistical patterns that are represented by the first two latent semantic factors, our LSA model suggests that when document D6 mentions *media*, it actually refers to *media selection*.[4]

[4]Interestingly, the word *choice* appears in document D6. In our small example this word is ignored, since it appears only once in the entire collection. In a larger set of documents, however, *choice* would be a participating term and LSA would then treat *selection* and *choice* as synonyms by listing both of them as high-loading terms in the media selection/choice factor.

Table C6 Approximated Term Frequencies, Produced Using the First Two Principal Factors

	Document					
Term	D1	D2	D3	D4	D5	D6
acceptance	–0.055	0.539	0.048	0.509	0.472	–0.049
information	–0.018	0.487	0.075	0.467	0.426	–0.024
media	0.510	–0.063	0.500	0.061	–0.069	0.326
model	0.302	0.217	0.345	0.282	0.182	0.187
selection	0.666	–0.045	0.659	0.115	–0.057	0.424
technology	–0.055	0.539	0.048	0.509	0.472	–0.409

It has been argued in the psychology literature (Landauer 2002; Landauer et al. 1998) that this approximation imitates the way our human brain learns and draws conclusions. To illustrate this point, let us suppose that a human student tried to understand Information Systems by studying only this extremely minimalist collection of six small documents, made even smaller by considering only the six terms used in our example. By reading the sixth *MISQ* paper title (document D6), the student would learn that IS research is concerned with *media*. By reading document D1, he/she would learn that IS research is particularly interested in *media selection*. Finally, by reading document D3, he/she would learn that IS researchers have proposed a *media selection model*. After reading all the documents and pausing for some reflection, the student would realize that IS research (as represented by this document collection) is dominated by the themes of (1) information technology acceptance and (2) media selection. Therefore, when document D6 mentions *media*, it is essentially discussing media selection, even though the word *selection* is missing from that document.