LSA: From Information Retrieval to Team Performance

Dr. Melanie J. Martin Mathematics and Computer Science Speaker Series CSU Stanislaus September 28, 2007

Outline of Talk

- Some Context
- A Tool for Information Retrieval - Latent Semantic Analysis
- More Than a Tool for Information Retrieval - Latent Semantic Analysis applied to Team Discourse
- Conclusion

First Some Context

- I am an Assistant Professor of Computer Science at CSU Stanislaus, I also teach Mathematics
- My dissertation title is "Reliability and Verification of Natural Language Text on the World Wide Web"
- This talk is about joint work with Peter W. Foltz at the Computing Research Laboratory at New Mexico State University

Areas I currently work in:

- Artificial Intelligence
 - Data Mining
 - Machine Learning
 - Natural Language Processing
- Information Retrieval
- Health Informatics
- Digital Media
- Computer Science Education

Information Retrieval

- Documents
- User has a need
- Find relevant content
- Indexing
- How to represent DocumentsRequests (queries)
- Searching
 - Query to document
 Based on indexing



One way to do this:

- Vector Space Model (Salton et al. 1968)
- Represent a collection of documents
 - Term (rows) by document (columns) matrix, based on occurrence
 - Translate into vectors in a vector space
 - One dimension for each term (!)
 - · Documents and queries are vectors of terms

Term	by I	Context	Μ	latrix
	-			

Doc/	Doc	Doc	Doc	Doc	Doc	Doc
Term	1	2	3	4	5	6
Information	0	0	1	1	1	0
Retrieval	0	0	1	1	1	1
System	1	1	1	0	0	0



One way to do this:

- Rank documents by how "close" they are to the query
 - "Close" is usually the computed using the cosine to measure distance between vectors (documents)
 - small angle = large cosine = similar
 - large angle = small cosine = dissimilar



Some Minor Issues

- What exactly are terms?
 - Stemming
 - Stop words
- Are all terms of equal importance?
 - Weighting schemes

Two Major Issues

- Words have multiple meanings (polysemy) - For example: model, python, chip, bank
- Multiple ways to refer to the same object or concept (synonymy)
 - For example: car-automobile



Outline of Talk

• Some Context

- A Tool for Information Retrieval – Latent Semantic Analysis
- More Than a Tool for Information Retrieval
 Latent Semantic Analysis applied to Team
 Discourse
- Conclusion

Latent Semantic Analysis (LSA)

- Latent Semantic Analysis was developed at Bellcore (now Telcordia) in the late 1980s (1988). It was patented in 1989.
- Goal
 - Address synonymy and polysemy issues
 - Model underlying/hidden/latent semantic relationship between terms and documents
 - Things with same meaning should be close together in the space

Latent Semantic Analysis

• Strategy

- Reduce dimension of space
- Semantically related terms end up in same dimension
- Exploit term co-occurrence
 - Two or more terms occur in same documents more often than chance
 - If words co-occur in similar contexts => evidence of semantic relatedness
 - D1: user interface HCI interaction
 - D2: HCI interaction
 - Query: user interface

LSA and Co-occurrence

- LSA operates on the deep level (latent) meaning of words rather than the surface characteristics (exact matches). Example:
 - The doctor operates on the patient.
- The physician is in surgery.
- No term overlap

	Term	LSA
 Doctor — Doctor 	1.0	1.0
 Doctor—Physician 	0.0	0.8
- Doctor-Surgeon	0.0	0.7

Constraint Constraint Constraint 0.0017 0.00184 0.00184 0.00184 0.00184 0.0017 0.00187 0.00184 0.00184 0.00184 0.00184 0.0017 0.00187 0.00184 0.00184 0.00184 0.00184 0.00187 0.00187 0.00184 0.00184 0.00184 0.00184 0.00187 0.00187 0.00187 0.00184 0.0118 0.0118 0.00187 0.00187 0.00187 0.00184 0.0118 0.0118 0.00187 0.00187 0.00187 0.00187 0.00187 0.00187 0.00187 0.00181 0.00187 0.00187 0.00187 0.00187 0.00181 0.00181 0.00187 0.00187 0.00187 0.00187 0.00181 0.00181 0.00187 0.00187 0.00187 0.00187 0.00181 0.00181 0.00187 0.00187 0.00187 0.00187 0.00181 0.01121 0.01121 0.01121									
Representation of "Life" 0.00107 -0.00607 -0.001394 -0.001495 -0.00121 -0.001371 0.00107 -0.00607 -0.001394 -0.001395 -0.001391 -0.001311 0.00107 -0.00607 -0.001295 -0.001395 -0.001391 -0.001391 0.00107 -0.001205 -0.001395 -0.001391 -0.001391 -0.001391 0.00117 -0.001205 -0.001397 -0.001391 -0.001391 -0.001391 0.00114 -0.001205 -0.001397 -0.001397 -0.001397 -0.001397 0.00114 -0.001170 -0.001170 -0.001397 -0.001397 -0.001397 0.00114 -0.001515 -0.001107 -0.001170 -0.001397 -0.001397 0.00114 -0.001514 -0.001514 -0.001514 -0.001514 -0.001514 0.00114 -0.001514 -0.001514 -0.001514 -0.001514 -0.001514 0.00215 -0.001514 -0.001574 -0.001514 -0.001516 0.00216 <td< th=""><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th></td<>									
Respires entation of "Life" 0.0393 -0.021 -0.0394 -0.04945 -0.05371 0.01017 -0.0465 -0.0284 -0.0318 -0.0781 0.02139 -0.0224 -0.0318 -0.0786 -0.0384 0.02139 -0.02251 -0.0318 -0.0786 -0.0388 0.05139 -0.02264 -0.0318 -0.0786 -0.0388 0.05139 -0.02264 -0.0318 -0.0138 -0.0388 0.05130 -0.02264 -0.0318 -0.0138 -0.0388 0.05130 -0.0207 -0.001397 -0.03370 -0.03370 0.05140 -0.01126 0.02776 -0.03357 -0.030357 0.00814 -0.01126 0.02776 -0.010357 -0.030357 0.00817 0.04147 -0.01127 -0.01126 -0.010357 0.00817 0.04147 -0.01126 -0.010136 -0.010136 0.02815 -0.01286 -0.012146 -0.010136 -0.01216 0.02817 -0.01286 -0.0									
LTEE: 0.03982 -0.0238 -0.008452 -0.0221 -0.03371 0.001017 -0.0463 -0.0228 -0.0224 -0.02318 0.02118 0.001017 -0.0463 -0.0228 -0.0224 -0.02118 0.02111 0.001017 -0.04633 -0.0111 -0.0228 -0.0228 -0.0218 0.02118 0.02181 0.05323 -0.0223 -0.001749 -0.01118 0.02181 0.00319 0.05323 -0.0125 -0.01248 0.001118 0.00319 0.0111 -0.01749 -0.0118 0.00317 0.00317 0.0111 -0.02218 0.00317 0.00317 0.00317 0.0111 -0.01221 0.02218 0.00317 0.01031 0.02215 -0.01244 -0.01214 0.01031 0.02231 0.00231 0.02845 -0.01264 -0.01214 -0.01215 0.00231 0.00231 0.02845 -0.01264 -0.00121 -0.0214 0.01033 0.00231 0.02845 -	-	an	ree	oni	24	<u>nn</u>	of	" 7	<u>~</u> ″
LIFE: 0.0392 - 0.023 0.01344 -0.006452 -0.04221 -0.05371 0.07101 -0.0423 0.0234 0.0244 -0.0512 0.07213 0.0423 0.0215 0.00151 0.0244 0.05124 0.0275 0.0135 0.0155 0.00156 0.00359 0.05124 0.0255 0.01125 0.00155 0.00151 0.00359 0.0314 0.0155 0.01125 0.00157 0.00359 0.0301 0.001113 -0.02021 0.02476 0.00357 0.0305 0.00055 -0.00156 0.000157 0.00357 0.0305 0.00055 -0.00156 0.000157 0.00357 0.0305 0.00055 -0.00156 0.000157 0.00357 0.0305 0.00055 -0.00156 0.000157 0.00357 0.0305 0.00055 -0.00156 0.00157 0.000357 0.0305 0.00055 -0.00156 0.00157 0.000357 0.0305 0.00055 -0.00156 0.00157 0.00157 0.0305 0.00055 -0.00156 0.00157 0.000357 0.0305 0.00055 -0.00156 0.00157 0.00157 0.0305 0.00056 0.00151 0.00164 0.01215 0.00895 0.00056 0.00157 0.00157 0.00151 0.00156 0.00357 0.00151 0.00156 0.00157 0.00157 0.00158 0.00537 0.00151 0.00156 0.00157 0.00157 0.00158 0.00536 0.00056 0.00156 0.00157 0.00157 0.00156 0.00536 0.00056 0.00157 0.00157 0.00154 0.00156 0.00536 0.00056 0.00156 0.00165 0.00157 0.00154 0.00546 0.00058 0.00157 0.00159 0.00154 0.00154 0.00056 0.00156 0.00165 0.00157 0.00154 0.00154 0.00056 0.00156 0.00165 0.00157 0.00154 0.00154 0.00056 0.00158 0.00157 0.001594 0.00154 0.00154 0.00056 0.00158 0.00157 0.001594 0.00154 0.00154 0.00056 0.00158 0.00157 0.000754 0.00154 0.00154 0.00056 0.00158 0.00155 0.00155 0.00155 0.00156 0.00516 0.00154 0.00055 0.00165 0.00157 0.00154 0.00516 0.00154 0.00155 0.00165 0.00157 0.00754 0.00164 0.00516 0.00154 0.00155 0.00155 0.00155 0.00155 0.00155 0.00154 0.00056 0.00158 0.00155 0.00155 0.00156 0.00516 0.00154 0.00056 0.00156 0.00155 0.00155 0.00156 0.00516 0.00156 0.00156 0.00155 0.00155 0.00156 0.00516 0.00156 0.00156 0.00157 0.00756 0.00516 0.00156 0.00156 0.00157 0.00756 0.00156 0.00156 0.00156 0.00156 0.00157 0.00756 0.00156 0.00156 0.00156 0.00156 0.00157 0.00756 0.00516 0.000156 0.00156 0.00157 0.00756 0.000156 0.00156 0.00156 0.00156 0.00157 0.00756 0.00516 0.00056 0.00156 0.00155 0.000556 0.00157 0.000516 0.000156 0.00056 0.000550 0.00		<u>ep</u>	65						<u> </u>
11 1 0.03992 -0.0394 -0.006452 -0.0214 0.00101 -0.0394 -0.0242 -0.0214 0.00110 -0.0463 -0.1228 -0.0214 0.0113 -0.0214 -0.0314 -0.0118 0.0113 -0.02118 -0.0244 -0.0118 0.0113 -0.02118 -0.0118 -0.0118 0.0114 -0.01472 -0.0118 -0.0118 0.0115 -0.01472 -0.0118 -0.0118 0.0116 -0.01126 -0.0227 -0.0119 0.0116 -0.01126 -0.0227 -0.01197 -0.00357 0.0101 -0.01126 -0.0227 -0.0221 -0.0223 0.0101 -0.01126 -0.0221 -0.0223 -0.02357 0.00891 -0.01647 -0.01846 -0.02142 -0.010357 0.02891 -0.01526 -0.0221 -0.02247 -0.0121 0.0281 -0.01626 -0.02142 -0.01335 -0.01614 0.0281									
0.0392 -0.023 0.01394 -0.006452 -0.04221 -0.05371 0.01397 -0.0425 0.02184 -0.02284 -0.05242 0.02218 0.02730 0.04236 0.0138 -0.0138 -0.0139 0.05139 -0.0252 0.00139 -0.0139 0.02841 0.0149 0.05139 0.0027 -0.00178 -0.0139 0.00510 0.0539 0.05114 -0.0155 0.01126 0.00759 -0.00487 0.005137 0.03011 -0.01113 -0.0221 0.02759 -0.00487 0.00357 0.0301 -0.01113 -0.0221 0.02759 -0.00487 0.00357 0.0305 0.000017 -0.00184 0.00189 -0.00187 0.00357 0.0455 0.0404 -0.00189 0.00187 -0.00357 0.0305 0.00011 -0.0184 0.00187 -0.00187 0.0305 0.00017 -0.0184 0.00184 0.0113 0.02828 -0.00291 0.00017 -0.00184 0.01215 0.02828 0.00359 -0.00184 -0.0184 0.01215 0.0284 0.00359 -0.0184 0.00181 -0.0221 -0.0244 0.00319 0.0111 0.0121 -0.0184 0.0211 -0.0184 0.0283 0.00051 -0.0121 -0.00184 0.01215 0.0284 0.00278 -0.01121 -0.00184 0.0211 -0.0244 0.00319 0.01118 0.02208 -0.01121 -0.0358 0.00154 -0.01519 0.00317 0.01118 0.02208 -0.0121 -0.0358 0.00154 -0.01519 0.00346 0.00278 -0.0121 -0.0358 0.00154 -0.00154 0.00516 0.00468 0.0148 -0.00184 0.0224 -0.0154 0.00516 0.00468 0.0148 -0.00184 0.0214 -0.0154 0.00516 0.00468 0.0148 -0.00184 0.0214 -0.0154 0.00516 0.00468 0.0148 -0.00150 -0.0124 -0.0154 0.00516 0.00468 0.0148 -0.00550 -0.0157 -0.02146 0.00516 0.00468 0.0148 -0.00184 0.0224 -0.0154 0.00516 0.00468 0.0148 -0.00184 0.0214 -0.01044 0.00746 0.00516 0.00448 -0.00578 0.00164 0.00746 0.00146 0.00746 0.00516 0.00448 -0.00578 0.00164 0.00746 0.00516 0.00449 -0.00150 0.00174 -0.01644 0.00746 0.00516 0.00449 0.00578 -0.01668 0.00164 0.00746 0.00516 0.00449 0.00578 -0.01668 0.00164 0.00746 0.00516 0.00449 0.00578 -0.01668 0.00164 0.00746 0.00516 0.00449 0.00578 -0.01668 0.00464 0.00746 0.00516 0.00498 0.00164 0.000796 0.00164 0.00746 0.00516 0.00498 0.00164 0.000796 0.00164 0.00746 0.00516 0.00498 0.00165 0.00167 0.00746 0.00548 0.000548 0.00164 0.00746 0.00548 0.000548 0.00164 0.00746 0.00164 0.00746 0.00548 0.000548 0.00164 0.00057 0.00746 0.00548 0.000548 0.00164 0.00057 0.000746 0.00164	LIFE:								
0.001017 0.0466 0.0.0228 0.0.0234 0.0512 0.02118 0.0213 0.0425 0.013 0.013 0.013 0.0131 0.0241 0.0213 0.0425 0.000428 0.0014 7.0.00438 0.01788 0.0512 0.0252 0.00178 0.0215 0.00040 0.01393 0.0514 0.0017 0.00178 0.0515 0.01014 0.001397 0.00516 0.00007 0.00178 0.0515 0.00140 0.01357 0.0055 0.00040 0.00149 0.0011 0.00357 0.0455 0.00040 0.00149 0.0011 0.00357 0.0455 0.00040 0.00145 0.00110 0.0022 0.00221 0.0258 0.00040 0.00156 0.00121 0.00231 0.000371 0.0258 0.000591 0.0136 0.00151 0.00231 0.000371 0.0258 0.000591 0.0136 0.00151 0.00231 0.000371 0.00153 0.0258 0.00152 0.00156 0.00231 0.000371 0.00153 0.0258 0.00152 0.00156 0.00231 0.000371 0.00153 0.00258 0.000591 0.0136 0.00151 0.00153 0.00153 0.00258 0.000591 0.0136 0.00151 0.00153 0.00153 0.00258 0.000591 0.0136 0.00151 0.00153 0.00153 0.00251 0.00278 0.00156 0.00231 0.000370 0.00339 0.00254 0.00258 0.00156 0.00151 0.00153 0.00153 0.00254 0.00058 0.00156 0.00157 0.00153 0.00153 0.00254 0.00058 0.00156 0.00157 0.00153 0.00153 0.00258 0.00058 0.00158 0.00157 0.00153 0.00153 0.00258 0.00058 0.00158 0.00157 0.00153 0.00153 0.00258 0.00058 0.00158 0.00157 0.00158 0.00156 0.00158 0.00058 0.00158 0.00157 0.00158 0.00156 0.00158 0.00058 0.00158 0.00158 0.00157 0.00128 0.00258 0.00058 0.00158 0.00158 0.00158 0.00158 0.00158 0.00058 0.00158 0.00158 0.00158 0.00158 0.00158 0.00158 0.00058 0.00158 0.00158 0.00158 0.00158 0.00158 0.00014 0.00158 0.00058 0.00058 0.00058 0.00158 0.00158 0.00058 0.00058 0.00058 0.00058 0.00058 0.00058 0.00158 0.00158 0.00058 0.00058 0.00058 0.00058 0.00058 0.00058 0.00058 0.00158 0.00158 0.00158 0.00158 0.00058 0.00058 0.00058 0.00058 0.00058 0.00058 0.00058 0.00058 0.00058 0.00	0.03992	-0.023	0.01394	-0.006452	-0.04221	-0.05371			
0.02730 0.0223 0.02252 0.002254 0.00747 0.007481 0.02539 0.0255 0.00747 0.007491 0.05549 0.00551 0.00134 0.02155 0.009691 0.03491 0.05149 0.00517 0.00178 0.0418 0.00151 0.00339 0.0114 0.0155 0.01126 0.0276 0.00387 0.00357 0.00357 0.003557 0.0485 0.04640 0.00469 0.00817 0.00257 0.0485 0.04640 0.0181 0.001537 0.0495 0.004617 0.0181 0.00184 0.01215 0.0497 0.00181 0.01126 0.0213 0.00174 0.0285 0.00117 0.0152 0.0213 0.00174 0.0285 0.00011 0.01152 0.0213 0.00174 0.0285 0.00151 0.01152 0.0213 0.00174 0.0285 0.00151 0.01152 0.01181 0.00184 0.01125 0.04972 0.00181 0.01182 0.0118 0.00181 0.001533 0.00231 0.00178 0.01182 0.00181 0.00183 0.001533 0.00284 0.00278 0.01121 0.00181 0.00184 0.01184 0.00178 0.01181 0.0218 0.00181 0.00184 0.00153 0.00284 0.00278 0.01121 0.00181 0.00184 0.00183 0.00284 0.00278 0.01121 0.01184 0.00184 0.00183 0.00184 0.00278 0.01121 0.01184 0.00184 0.00183 0.00184 0.00278 0.01181 0.00184 0.00184 0.00184 0.00184 0.00278 0.01181 0.00184 0.00184 0.00184 0.00278 0.01181 0.00184 0.00184 0.00184 0.00278 0.00185 0.00184 0.00182 0.00184 0.00278 0.00185 0.00184 0.00182 0.00184 0.00278 0.00185 0.00185 0.00184 0.00184 0.00078 0.00185 0.00184 0.00182 0.00184 0.00078 0.00185 0.00185 0.00184 0.00184 0.00078 0.00185 0.00185 0.00184 0.00084 0.00078 0.00185 0.00185 0.00184 0.00084 0.00078 0.00185 0.00185 0.00184 0.00185 0.00084 0.00078 0.00185 0.00185 0.00185 0.00185 0.00084 0.00078 0.00185 0.00185 0.00185 0.00185 0.00184 0.00084 0.00078 0.00185 0.00185 0.00185 0.00185 0.00185 0.00184 0.00084 0.00078 0.00185 0.00185 0.00185 0.00185 0.00184 0.00084 0.00078 0.00185 0.00185 0.00185 0.00185 0.00185	-0.001017	-0.04663	-0.0228	-0.02284	-0.0542	0.02118			
0.02139 -0.0252 0.002628 -0.007437 -0.003438 0.01788 0.05392 0.02553 -0.01334 0.02155 0.000581 0.03431 0.05395 0.02553 -0.01349 0.02155 0.000581 0.03431 0.05316 0.00527 -0.00178 -0.04187 0.005319 0.0301 0.00113 -0.0221 0.02578 0.00837 0.003537 0.0401 0.00113 -0.0221 0.02578 0.00837 0.003537 0.0403 0.00113 -0.00145 0.00101 -0.0522 0.05393 0.0027 0.00417 0.04143 -0.00184 -0.0216 0.01003 0.05285 0.00299 -0.01362 0.02317 -0.00373 0.00133 0.05285 0.00299 -0.01362 0.02317 -0.00373 0.00339 0.00286 0.00278 -0.01121 -0.05185 0.00217 -0.00373 0.00399 0.00286 0.00278 0.00121 -0.05185 0.00217 -0.00373 0.00399 0.00286 0.00278 0.00121 -0.05185 0.00213 -0.00353 0.00286 0.00278 0.00121 -0.05185 0.00213 -0.00353 0.00286 0.00278 0.00121 -0.05185 0.00213 -0.00374 0.00133 0.00286 0.00278 0.00121 -0.05185 0.00213 -0.00134 0.00278 0.00128 0.00128 -0.00126 0.00373 0.00133 0.00286 0.00278 0.00121 -0.05185 0.00246 0.00133 0.00286 0.00278 0.00128 -0.00186 0.00146 0.00134 0.00286 0.00278 0.00128 -0.00168 0.00146 0.00134 0.00286 0.00278 0.00128 -0.00168 0.00146 0.00136 0.00286 0.00278 0.00128 -0.00168 0.00146 0.00136 0.00286 0.00278 0.00128 -0.00168 0.00146 0.00134 0.00216 0.00286 0.00188 -0.00168 0.00146 0.00136 0.00286 0.00278 0.00128 -0.00168 0.00168 0.00218 -0.00136 0.00286 0.00278 0.00128 -0.00168 0.00168 0.00137 -0.00186 0.00286 0.00278 0.00128 -0.00168 0.00168 0.00218 -0.00136 0.00286 0.00278 0.00128 -0.00168 0.00168 0.00146 0.00136 0.00286 0.00286 0.00188 -0.00168 0.00168 0.00168 0.00168 0.00218 -0.00168 0.000168 0.000168 0.00168 0.000168 0.000168 0.00168 0.00168 0.00168 0.00168 0.00168 0.00168 0.00168 0.00168 0.00168 0.00168 0.00168 0.000168 0.000168 0.000168 0	0.02703	0.04236	0.013	0.0151	-0.01519	0.02841			
0.0552 0.0253 0.0253 0.02134 0.02155 0.000901 0.03491 0.05146 0.00527 0.00179 0.04187 0.04187 0.001511 0.00339 0.0114 0.0155 0.01126 0.02576 0.00387 0.003557 0.0455 0.04641 0.004699 0.00817 0.06322 0.05233 0.0455 0.04641 0.00169 0.00817 0.06222 0.02231 0.0452 0.008417 0.04143 0.00184 0.02142 0.01003 0.02585 0.005991 0.0152 0.0233 0.00231 0.003537 0.02585 0.005991 0.0152 0.0233 0.00231 0.001533 0.02585 0.005991 0.0152 0.0233 0.005734 0.00153 0.02576 0.00591 0.0118 0.00181 0.00153 0.00153 0.02578 0.00157 0.0118 0.00181 0.00153 0.00153 0.02578 0.00157 0.0118 0.01214 0.001514 0.00153 0.02578 0.00157 0.0118 0.01214 0.00151 0.00153 0.00251 0.00259 0.0112 0.0158 0.02417 0.00573 0.01533 0.00254 0.00258 0.0112 0.0158 0.01513 0.01533 0.01533 0.00254 0.00258 0.0112 0.0158 0.01513 0.01513 0.01513 0.00254 0.00258 0.0114 0.01558 0.01524 0.01516 0.02518 0.00468 0.0118 0.01558 0.01524 0.00154 0.00514 0.00258 0.0148 0.01558 0.01524 0.00154 0.00544 0.00278 0.0158 0.01558 0.01524 0.00154 0.00544 0.00278 0.0158 0.01558 0.01524 0.00154 0.00544 0.00278 0.0158 0.01558 0.0152 0.01544 0.00154 0.00544 0.00278 0.0158 0.01558 0.0152 0.01544 0.00154 0.00544 0.00278 0.0158 0.0158 0.0158 0.00458 0.01545 0.00544 0.00278 0.0057 0.01558 0.0154 0.00544 0.00278 0.0057 0.01558 0.00544 0.00154 0.00544 0.00054 0.00570 0.01558 0.00545 0.01554 0.00545 0.00544 0.00554 0.00558 0.00450 0.0152 0.00545 0.00544 0.00554 0.00558 0.00450 0.00520 0.00555 0.000545 0.00554 0.00558 0.00550 0.00555 0.01545 0.01555	-0.02139	-0.0252	0.002628	-0.007437	-0.003438	0.01788			
0.005196 0.00527 0.001789 -0.04187 0.005131 0.005329 0.0314 -0.01513 -0.0227 0.005759 -0.04084 -0.01537 0.0301 -0.0113 -0.02271 0.02575 -0.040387 0.003557 0.03934 0.0055 -0.01056 0.00817 0.003857 0.003557 0.03934 0.0055 -0.00584 -0.0113 -0.04644 0.0113 0.05939 0.00591 -0.0113 -0.00513 -0.00521 -0.0227 0.05251 -0.02999 -0.01562 0.02317 -0.005734 0.001533 0.005991 0.01528 -0.00151 -0.001513 -0.005734 0.001533 0.00599 0.01562 0.02817 -0.00573 -0.00573 -0.0053 0.00537 0.01128 -0.01563 -0.00181 -0.00153 -0.00154 0.02537 0.0118 0.02288 -0.00182 -0.00153 -0.00154 0.02545 -0.00258 -0.01287 -0.00158 -0.01546 -0.00154 0.02545 -0.00258 -0.0128 -0.00184 -0.00153 -0.00154 0.02545 -0.00258 -0.0128 -0.00168 -0.00154 -0.00154 0.02546 -0.00278 -0.0128 -0.00168 -0.00154 -0.00154 0.02546 -0.00278 -0.0128 -0.00168 -0.00164 -0.00154 0.02546 -0.00278 -0.0128 -0.00168 -0.00164 -0.00154 0.00154 -0.00278 -0.0168 -0.00168 -0.00164 -0.00154 0.00254 -0.00258 -0.0165 -0.01668 -0.0164 -0.00164 -0.00154 0.00254 -0.00254 -0.00168 -0.00168 -0.00164 -0.00164 -0.00164 -0.00154 0.00154 -0.00278 -0.0155 -0.01668 -0.0164 -0.00164 -0.00154 0.00254 -0.00258 -0.0165 -0.01668 -0.00164 -0.00154 0.00254 -0.00258 -0.00168 -0.00168 -0.00164 -0.00154 0.00254 -0.00258 -0.00168 -0.00164 -0.00164 -0.00164 -0.00154 0.00254 -0.00258 -0.00168 -0.00168 -0.00164 -0.00164 -0.00154 0.00254 -0.00258 -0.0165 -0.01668 -0.00164 -0.00154 -0.00164 -0.00164 -0.00154 0.00254 -0.00258 -0.00158 -0.00168 -0.00164 -0.00164 -0.00164 -0.00154 -0.00164	-0.05929	0.02553	-0.01334	0.02155	0.009091	0.03491			
0.0114 -0.0155 0.01126 0.02759 -0.04004 -0.01597 0.0405 0.0011 0.00111 -0.02201 0.02476 0.003837 0.003057 0.0405 0.04604 -0.004699 0.00017 -0.06222 -0.02303 0.04057 0.00841 0.04143 -0.01184 -0.02146 0.01003 0.02215 0.00841 -0.0162 0.0233 -0.00221 -0.02247 0.02845 -0.00299 -0.0152 0.0233 -0.00231 -0.00231 -0.00153 0.02728 0.00891 -0.0116 -0.00184 -0.02146 -0.01153 0.02531 -0.0299 -0.0124 -0.00184 -0.0213 -0.00153 0.02531 -0.0299 -0.0124 -0.0181 -0.00153 -0.01339 0.00284 0.00278 -0.01121 -0.05159 -0.01264 -0.00156 0.02535 0.00468 0.01148 -0.0158 0.01468 0.04268 0.02548 0.00248 -0.0148 -0.0156 0.01264 -0.00156 0.02548 0.00468 0.0148 -0.0158 0.01426 0.04268 0.02548 0.00468 0.0148 -0.0158 0.01426 0.04268 0.02548 0.00468 0.0148 -0.0156 0.01264 0.00116 0.00116 0.00468 0.0148 -0.0156 0.0126 0.0126 0.00116 0.00468 0.0148 -0.0156 0.0126 0.0126 0.00116 0.00468 0.0148 -0.0156 0.0126 0.0126 0.00116 0.00168 0.0148 -0.0156 0.0126 0.0126 0.00116 0.00168 0.0148 -0.0156 0.0126 0.01023 0.0016 0.00259 0.0027 0.0168 0.0426 0.00126 0.01053 0.00068 0.00076 0.00059 0.00050 0.00057 0.01073	-0.005196	0.009027	-0.001789	-0.04187	0.006131	0.005329			
0.0301 0.001113 -0.02021 0.02076 0.003837 0.003837 0.00385 0.00385 0.003054 0.000017 -0.00222 -0.0233 0.00384 0.00355 -0.008064 -0.01813 -0.004644 0.01213 0.00427 0.00427 0.004064 0.01844 0.01214 0.01003 0.02885 -0.00394 -0.0131 0.00184 -0.0214 0.01003 0.02885 -0.00394 -0.0131 0.00184 -0.00121 -0.00214 0.01003 0.00384 -0.00278 -0.01121 -0.00181 -0.00537 -0.00339 0.00346 -0.00278 -0.01121 -0.05185 -0.01264 -0.001316 0.00357 0.01181 0.02186 -0.001813 -0.00131 -0.00139 0.00346 -0.00278 -0.01121 -0.05185 -0.01264 -0.001316 0.02365 -0.00248 -0.00184 -0.00186 0.001842 0.00246 0.02618 0.00368 0.01484 -0.00186 0.001842 0.00246 0.00316 -0.01249 -0.0165 0.01647 0.001842 0.001846 0.00364 0.00278 -0.0165 0.01675 0.001746 0.00364 0.00278 -0.0165 0.01678 0.00784 -0.010440 0.00284 0.00239 0.01484 -0.00189 0.00184 0.00184 0.00184 0.00248 -0.00185 0.01648 0.00184 0.001846 0.00186 0.01948 0.00187 0.00186 0.00184 0.00184 0.000186 0.01948 0.00187 0.00186 0.00184 0.00184 0.000186 0.01948 0.00187 0.00186 0.00184 0.00184 0.00184 0.00278 0.00187 0.00186 0.00184 0.00184 0.00184 0.00278 0.00187 0.00186 0.00184 0.00184 0.00184 0.00184 0.00218 0.00185 0.00185 0.00185 0.00184 0.00184 0.00218 0.00186 0.00185 0.00185 0.00184 0.00184 0.00184 0.00186 0.00186 0.00186 0.00185 0.00185 0.00186 0.00184 0.00186 0.00186 0.00186 0.00186 0.00185 0.00185 0.00186 0.00184 0.00186 0.00186 0.00186 0.00186 0.00185 0.00185 0.00186 0.00185 0.00186 0.00186 0.00186 0.00186 0.00185 0.00185 0.00186 0.00186 0.00186 0.00186 0.00186 0.00186 0.00185 0.00186 0.00185 0.00186 0.00186 0.00186 0.00186 0.00186 0.00186 0.00186 0.00186 0.00186 0.00186 0.00186 0.00186 0.00186 0.00186 0.00185 0.00186 0.000186 0.00186 0.000186 0.000186 0.00186 0.000	0.0114	-0.01655	0.01126	0.05759	-0.04004	-0.01597			
0.0495 0.04604 0.008699 0.008017 0.06222 0.0223 0.09894 0.00255 0.0085064 0.00181 0.004648 0.01215 0.04272 0.008417 0.04143 0.001844 0.022142 0.01003 0.02855 0.002991 0.0113 0.0233 0.00231 0.00221 0.02247 0.02521 0.0299 0.01562 0.0231 7 0.005734 0.001533 0.002844 0.00278 0.01121 0.00191 0.001573 0.001939 0.002844 0.00278 0.01211 0.01285 0.01244 0.04154 0.02375 0.01118 0.2205 0.01281 0.01284 0.04268 0.02356 0.01484 0.00278 0.01484 0.01584 0.04268 0.02544 0.00278 0.01484 0.001954 0.001954 0.010468 0.01554 0.00468 0.01484 0.001954 0.001954 0.010468 0.00584 0.00278 0.01484 0.0158 0.01594 0.010468 0.00584 0.00168 0.01484 0.001954 0.001954 0.010468 0.00584 0.00184 0.001954 0.00185 0.00184 0.00584 0.00185 0.01484 0.01284 0.01284 0.00584 0.00184 0.001954 0.00185 0.00184 0.00258 0.01484 0.001954 0.00184 0.00184 0.00258 0.00185 0.00185 0.00185 0.000184 0.00258 0.00258 0.00420 0.00178 0.00185 0.000856 0.00258 0.00040 0.00178 0.0148	0.0301	0.001113	-0.02021	0.02676	0.003837	0.000355			
0.009894 0.0035 -0.003664 -0.0119 -0.00464 0.01215 0.002894 0.0035 -0.003664 -0.01219 -0.00464 0.01215 0.02285 -0.003941 -0.0143 0.02313 -0.00221 -0.00231 0.02521 -0.02299 -0.01555 0.02417 -0.003734 0.00333 0.008991 -0.01518 -0.01653 -0.00373 -0.00373 -0.00353 0.00884 -0.002788 -0.0121 - 0.05155 -0.01516 0.03373 -0.01118 0.02304 -0.05159 0.01462 -0.001516 0.03521 -0.026278 -0.0121 - 0.00195 -0.001575 -0.02216 0.01551 -0.03624 -0.00589 0.01462 0.00716 0.05518 -0.03624 -0.0155 0.01475 0.00794 -0.01064 0.05518 -0.03249 -0.0155 0.01475 0.00264 -0.00464 0.05818 0.00284 -0.0250 0.01475 0.00264 -0.00785 0.005818 -0.03294 -0.0250 0.04606 0.04044 0.00716 0.05518 -0.03294 -0.0155 0.00470 0.03792 -0.00164 0.05818 0.00384 0.00231 0.04000 0.02794 -0.01064 0.05818 0.00384 0.00251 0.00400 0.03795 -0.02186 0.000818 0.003518 -0.0255 0.004070 0.03795 -0.02185	0.0495	0.04604	-0.004699	0.008017	-0.06222	-0.05293			
0.04272 0.008417 0.04143 -0.001844 -0.02142 0.01003 0.0285 -0.00299 -0.0143 0.0233 -0.00271 -0.0221 -0.02247 0.02521 -0.0299 -0.01562 0.02417 -0.005734 0.001533 0.008991 0.01278 -0.01512 -0.00191 -0.00573 -0.01536 0.00284 -0.00278 -0.01211 -0.05159 -0.01264 -0.001516 0.02375 0.01181 0.02248 -0.01580 -0.03572 -0.02166 0.02545 1.01282 0.0227 -0.01688 0.00462 0.04268 0.0255 1.01282 0.01284 -0.001895 0.001372 -0.02166 0.00516 1.01248 -0.0168 0.001875 0.00734 -0.01064 0.00516 1.01248 -0.0155 0.00734 -0.01646 0.00518 -0.01264 -0.0218 0.0464 0.00734 0.01064 0.00518 -0.01264 -0.0218 0.0464 0.00734 0.01064 0.00518 -0.0126 -0.0218 0.0464 0.00734 0.01064 0.00518 -0.0126 -0.0218 0.0464 0.00734 0.01064 0.00018 -0.0123 0.0464 0.00734 0.01073 0.00734 0.01064 0.00018 0.00528 0.0255 0.00047 0.00738 -0.0116	0.009894	0.00355	-0.005064	-0.01819	-0.004684	0.01215			
0.0288 - 0.00394 -0.0143 0.0233 -0.00221 -0.02247 0.02521 -0.02278 -0.0125 0.0247 -0.005734 0.00333 0.00899 1.01218 -0.01653 -0.008171 -0.00873 -0.00356 0.00384 -0.002788 -0.0121 - 0.05195 -0.01244 -0.00156 0.03373 -0.0118 0.02304 -0.05195 0.01424 -0.00156 0.03575 -0.04228 0.02518 0.004688 0.01484 -0.03588 0.03462 0.04268 0.03582 -0.06228 -0.0125 0.01475 0.00746 0.05181 -0.03244 -0.0155 0.01475 0.00744 -0.01064 0.05182 0.00254 -0.0250 0.04268 0.04044 0.007456 0.03584 0.00628 -0.0230 0.04668 0.04094 -0.07585 0.03585 0.03298 -0.0255 0.04208 0.04268 -0.07585 0.00085 0.03298 -0.0255 0.04208 0.04268 -0.07585 0.00087 0.03298 -0.0255 0.04208 0.04268 -0.07585	-0.04272	0.008417	0.04143	-0.001864	-0.02142	0.01003			
0.02521 -0.02099 -0.01562 0.02217 -0.003734 0.001333 0.008991 0.012578 -0.01121 -0.001891 -0.00373 -0.00399 0.002844 -0.00278 -0.01121 -0.05189 -0.01264 -0.001316 0.02375 0.01181 0.02304 -0.01388 0.03462 0.04268 0.02618 0.008468 0.01484 -0.007956 -0.03572 -0.02196 0.02547 -0.04262 0.02277 -0.01688 0.00464 0.00746 0.05184 1.01249 -0.0165 0.01688 0.00164 0.007954 0.01064 0.05264 0.00264 -0.023 0.0466 0.03752 0.00233 0.04264 0.00278 -0.01378 0.04692 0.04923 0.04264 0.0494 -0.00378 0.04692 0.04923 0.04924 0.0494 -0.00378 0.04692 0.04923 0.04924 0.0494 0.00378 0.04692 0.04923	-0.02885	-0.003961	-0.0143	0.02333	-0.000221	-0.02247			
0.008991 0.01218 -0.01653 -0.008171 -0.008733 -0.00873 0.00884 -0.002788 -0.01212 -0.05195 -0.01244 -0.001316 0.03375 0.01118 0.02304 -0.03598 0.03462 0.04268 0.03567 -0.04268 0.01484 -0.03988 0.03572 -0.02186 0.03567 -0.04822 0.00057 -0.001968 0.00404 0.00716 0.03561 -0.03247 -0.01616 0.01475 0.000744 -0.0164 0.03628 0.00624 -0.023 0.04666 0.0352 -0.00264 0.03528 0.00627 0.0239 0.0466 0.04204 -0.07585 0.00386 0.03298 -0.0255 0.04206 0.0379 -0.02186 0.00328 0.02598 -0.0255 0.000470 0.03793 -0.02185	0.02521	-0.02099	-0.01562	0.02417	-0.005734	0.001533			
0.002844 -0.00278 -0.01121 -0.03195 -0.01244 -0.001316 0.0375 0.01118 0.02304 -0.03588 0.03462 0.04248 0.02515 0.008468 0.01484 -0.007926 -0.03572 -0.02196 0.03557 -0.04252 0.02277 -0.001668 0.001044 0.007166 0.00318 1-0.0349 -0.0155 0.01675 0.00734 -0.01064 0.00216 0.01494 -0.00377 0.04462 0.03732 0.40233 0.04241 0.00464 -0.0377 0.04462 0.03734 -0.01064 0.04241 0.00494 0.00377 0.04462 0.03734 -0.01064 0.04021 0.01494 0.00377 0.04462 0.03745 -0.07943 0.00074 0.01494 0.00377 0.04462 0.00745 -0.07943	-0.008991	0.01218	-0.01653	-0.008191	-0.006373	-0.03939			
0.03375 0.01118 0.02304 -0.03583 0.03462 0.04268 0.03567 0.002668 0.01848 -0.03583 0.03572 -0.02196 0.03567 -0.04822 0.02627 -0.001688 0.004044 0.007165 0.03561 0.03624 0.0165 0.01875 0.000794 -0.01064 0.03524 0.00624 0.0.023 0.04866 0.0352 0.00823 0.045848 0.003528 0.0023 0.04686 0.03528 0.00823 0.00007 0.03598 0.02558 0.000400 0.03798 -0.02185 0.00007 0.03598 0.02558 0.000400 0.03798 -0.02185	-0.002844	-0.002278	-0.01121	-0.05195	-0.01264	-0.001516			
0.02618 0.009468 0.01484 -0.007906 -0.01572 -0.02196 0.03567 -0.04822 0.02627 -0.001668 0.004044 0.007416 0.058181 -0.03249 -0.0155 0.01675 0.007934 -0.01064 0.03624 0.00626 -0.023 0.04696 0.0392 0.008223 0.008286 0.008249 0.003373 0.046208 0.029245 -0.07855 0.008087 0.03298 -0.02558 0.000470 0.03795 -0.02188 0.00030 0.01181 -0.0156 0.004260 0.03279 -0.0218	0.03375	0.01118	0.02304	-0.03583	0.03462	0.04268			
0.03567 -0.04822 0.02627 -0.003648 0.00404 0.007156 0.03567 -0.02427 0.0165 0.01675 0.000794 -0.01064 0.03524 0.00628 -0.023 0.04866 0.0392 0.00823 0.03504 0.00249 0.00337 0.04608 0.02445 -0.07585 0.003067 0.03298 -0.02558 0.000470 0.03793 -0.02186 0.003067 0.03298 -0.02558 0.000470 0.03793 -0.02185	-0.02618	0.009468	0.01484	-0.007926	-0.03572	-0.02196			
0.05181 0.03249 -0.0155 0.01375 0.017954 -0.01664 0.03624 0.00626 -0.023 0.04962 0.0392 0.008223 0.002816 0.03494 0.003373 0.04268 0.02454 -0.07585 0.003087 0.03299 -0.02558 0.000470 0.03793 -0.02148 0.00370 0.0118 -0.01516 0.02466 -0.004573 0.02147	-0.03567	-0.04822	0.02627	-0.001668	0.004044	0.007416			
0.03824 0.00626 -0.023 0.04962 0.0392 0.008223 0.002816 0.007545 0.002816 0.007545 0.002067 0.03298 -0.02058 0.000470 0.03793 -0.02148 0.00017 0.01318 -0.01361 0.07342 -0.008473 0.01437	0.005181	-0.03249	-0.0165	0.01675	0.007954	-0.01064			
0.002816 0.03494 0.003373 0.04208 0.02945 -0.07585 0.003087 0.03298 -0.02058 0.000470 0.03793 -0.02148 0.0017 0.01518 -0.01361 0.02762 -0.0008555 0.0147	0.03624	0.00626	-0.023	0.04962	0.0392	0.008223			
0.003087 0.03298 -0.02058 0.000470 0.03793 -0.02148 0.04032 0.01518 -0.01361 0.02362 -0.0008575 0.03437	-0.002816	0.03494	0.003373	0.04208	0.02945	-0.07585			
0.04032 0.01518 =0.01361 0.02362 =0.0008575 0.03437	-0.003087	0.03298	-0.02058	0.000470	0.03793	-0.02148			
	-0.04032	0.01518	-0.01361	0.02362	-0.0008575	0.03437			
0.02592 0.01731 -0.06542 -0.02625 -0.009007 0.01611	0.02592	0.01731	-0.06542	-0.02625	-0.009007	0.01611			

	Meening	_	-f	"l ifa"
	NGennig			
0.62	lives	47	0.45	other
0.58	lifetime	10	0.45	
0.56	everyday	***	0.45	sour
0.55	death	52	0.44	sustain
0.54	living	53	0.43	pattison
0.52	loneliness	57	0.43	almost
0.50	bereavement	58	0.43	comfort
0.50	pleasures	63	0.43	own
0.49	alive	64	0.43	who
.49	despair	65	0.42	born
0.49	dying	67	0.42	on
0.49	existence	68	0.42	20
0.48	happiness	00	0.42	- the Hanne
0.48	die	69	0.42	challenge
0.48	dies	70	0.42	eternity
0.48	comforts	74	0.42	years

76 0.42 77 0.42 79 0.41 80 0.41 81 0.41 82 0.41 83 0.41 85 0.41

0.47 0.47 0.47 0.46 0.46 0.46

Latent Semantic Analysis

- Implementation: Four Basic Steps
 - Term by document (context) matrix
 - Convert matrix entries to weights
 - Singular Value Decomposition (SVD) performed on matrix

- Reduce Rank of matrix

- all but the k highest singular values are set to 0
- produces k-dimensional approximation of the original matrix (in least-squares sense)
 this is the "semantic space"

Latent Semantic Analysis

• Singular Value Decomposition



• Dimension Reduction





Latent Semantic Analysis

- · A theoretical model of cognitive phenomena
- A practical tool for measuring cognitive artifacts based on semantic information
- Provides measures of the semantic relatedness, quality, and quantity of information contained in discourse
- · Automatic and fast

Summary

• Some Issues

- SVD Algorithm complexity O(n^2*k^3)
 - n = number of terms
 - k = number of dimensions in semantic space (typically small ~50 to 350)
 - for stable document collection, only have to run once
 - not so good for large, dynamic document collection

Summary

- Some issues
 - Finding optimal dimension for semantic space
 - precision-recall improve as dimension is increased until hits optimal, then slowly decreases until it hits standard vector model
 don't want to run SVD too may times to find
 - optimal dimension
 - research being done on this

Summary

- Some issues
 - SVD assumes normally distributed data
 - term occurrence is not normally distributed
 - matrix entries are weights, not counts, which may be normally distributed even when counts are not