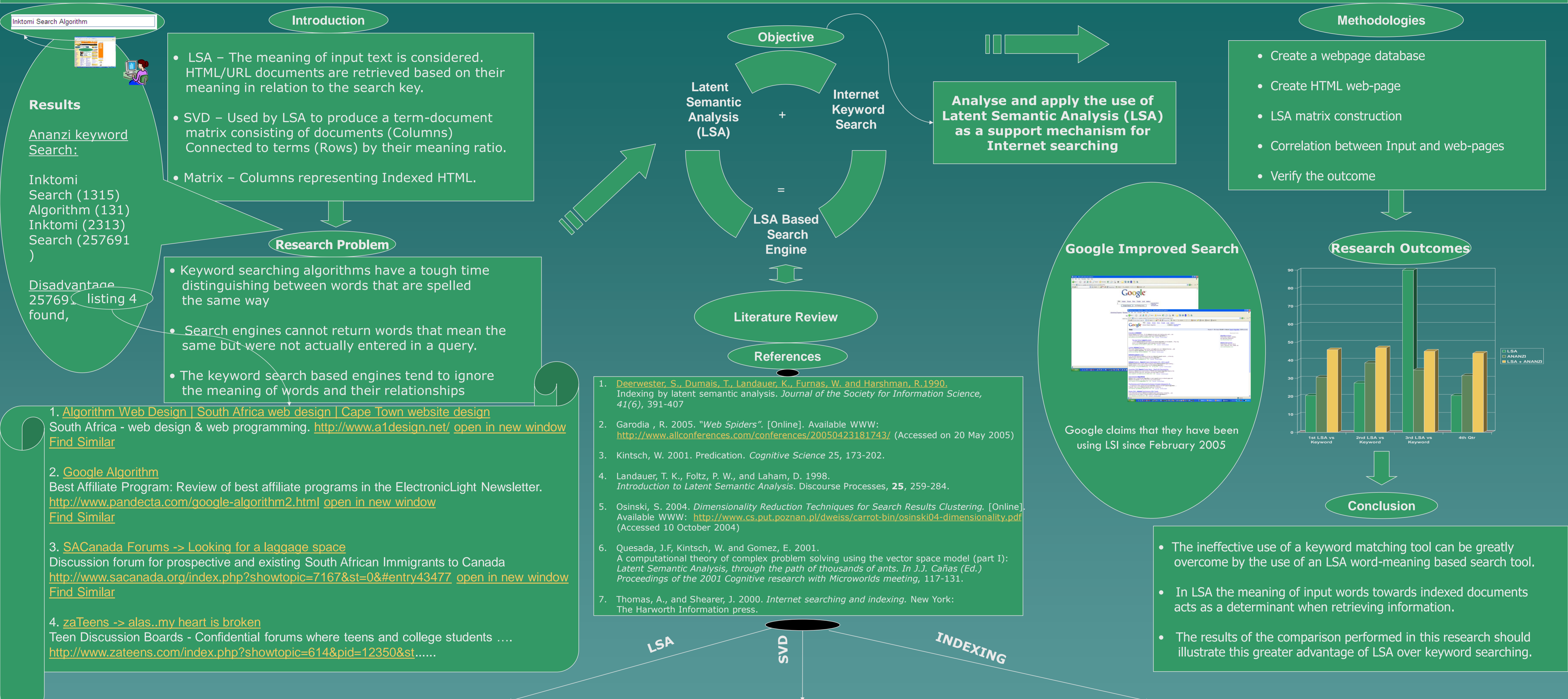


Investigating the effective use of LSA matrix comparison for Internet search

Mr. Songezo Nkukwana (FBI), Profs. Melius Weideman (FBI) and Jevon Davies (EE)



Results

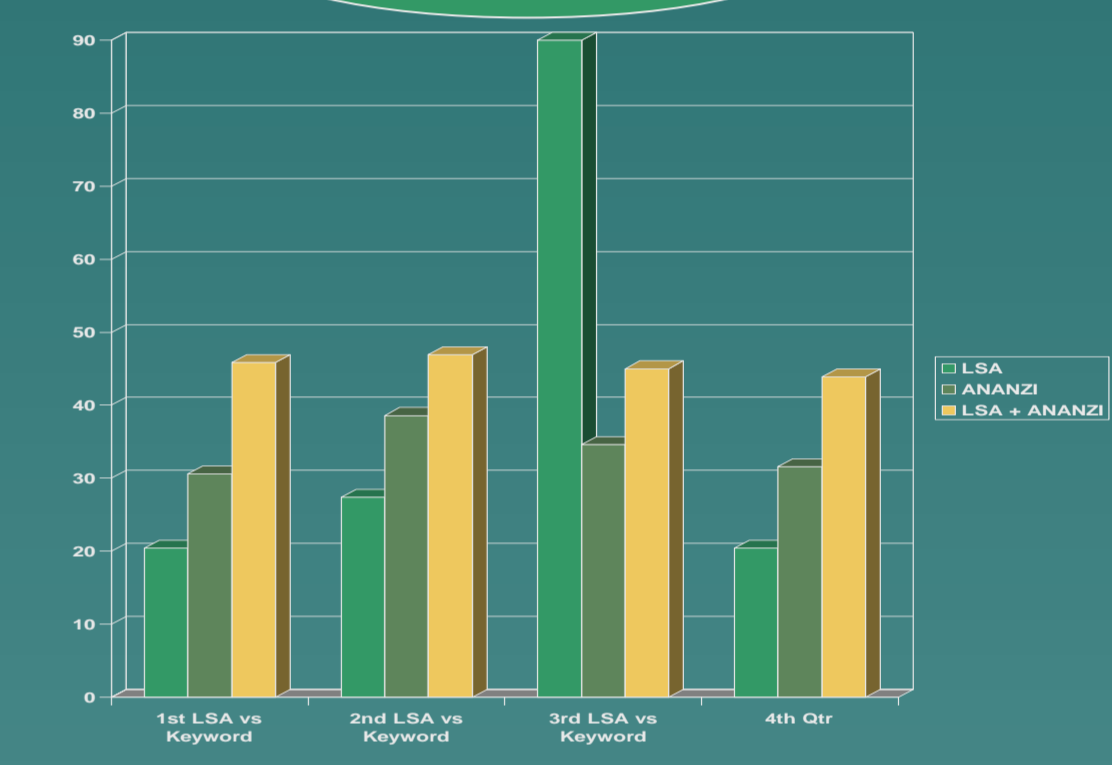
Ananzi keyword Search:

Inktomi Search (1315)
Algorithm (131)
Inktomi (2313)
Search (257691)

Disadvantage 25769 listing 4 found,

1. [Algorithm Web Design | South Africa web design | Cape Town website design](#)
South Africa - web design & web programming. <http://www.a1design.net/> open in new window Find Similar
2. [Google Algorithm](#)
Best Affiliate Program: Review of best affiliate programs in the ElectronicLight Newsletter. <http://www.pandecta.com/google-algorithm2.html> open in new window Find Similar
3. [SACanada Forums -> Looking for a laggage space](#)
Discussion forum for prospective and existing South African Immigrants to Canada <http://www.sacanada.org/index.php?showtopic=7167&st=0&#entry43477> open in new window Find Similar
4. [zaTeens -> alas..my heart is broken](#)
Teen Discussion Boards - Confidential forums where teens and college students <http://www.zateens.com/index.php?showtopic=614&pid=12350&st.....>

1. Deerwester, S., Dumais, T., Landauer, K., Furnas, W. and Harshman, R. 1990. Indexing by latent semantic analysis. *Journal of the Society for Information Science*, 41(6), 391-407
2. Garodia, R. 2005. "Web Spiders". [Online]. Available WWW: <http://www.allconferences.com/conferences/20050423181743/> (Accessed on 20 May 2005)
3. Kintsch, W. 2001. Predication. *Cognitive Science* 25, 173-202.
4. Landauer, T. K., Foltz, P. W., and Laham, D. 1998. *Introduction to Latent Semantic Analysis*. Discourse Processes, 25, 259-284.
5. Osinski, S. 2004. *Dimensionality Reduction Techniques for Search Results Clustering*. [Online]. Available WWW: <http://www.cs.put.poznan.pl/dweiss/carrot-bin/osinski04-dimensionality.pdf> (Accessed 10 October 2004)
6. Quesada, J.F, Kintsch, W. and Gomez, E. 2001. A computational theory of complex problem solving using the vector space model (part I): *Latent Semantic Analysis, through the path of thousands of ants*. In J.J. Cañas (Ed.) *Proceedings of the 2001 Cognitive research with Microworlds meeting*, 117-131.
7. Thomas, A., and Shearer, J. 2000. *Internet searching and indexing*. New York: The Harworth Information press.



- The ineffective use of a keyword matching tool can be greatly overcome by the use of an LSA word-meaning based search tool.
- In LSA the meaning of input words towards indexed documents acts as a determinant when retrieving information.
- The results of the comparison performed in this research should illustrate this greater advantage of LSA over keyword searching.

- > **LSA is a machine-learning model that induces representations of meaning of words by analysing the relation between words and passages in large bodies of representative text.**
- > **The implication of relevancy with LSA is the relationship between words extracted from a corpus of text and their meaning, in a given context or word space.**
- > **LSA has been effectively used as an Essay marking tool.**
- > **Google now claims to be using Latent Semantic Indexing to improve their searching algorithm.**
- > **SVD is a dimensional reduction tool used by LSA to produce three matrices of greater meaning or word relationships deduced from the original keyword/term matrix.**

(X) =	c1	c2	c3	c4	c5	m1	m2	m3	m4
human	1	0	0	1	0	0	0	0	0
interface	1	0	1	0	0	0	0	0	0
computer	1	1	0	0	0	0	0	0	0
user	0	1	1	0	1	0	0	0	0
system	0	1	1	2	0	0	0	0	0
response	0	1	0	0	1	0	0	0	0
time	0	1	0	0	1	0	0	0	0
EPS	0	0	1	1	0	0	0	0	0
survey	0	1	0	0	0	0	0	0	1
trees	0	0	0	0	0	1	1	1	0
graph	0	0	0	0	0	0	1	1	1
minors	0	0	0	0	0	0	0	1	1

Original Matrix

c1: **Human** machine **interface** for ABC **computer** applications
 c2: A **survey** of **user** opinion of **computer** system **response** time
 c3: The **EPS** user interface management system
 c4: System and human **system** engineering testing of **EPS**
 c5: Relation of user perceived **response time** to error measurement

m1: The generation of random, binary, ordered **trees**
 m2: The intersection **graph** of paths in **trees**
 m3: **Graph minors** IV: Widths of trees and well-quasi-ordering
 m4: **Graph minors**: A **survey**

	c1	c2	c3	c4	c5	m1	m2	m3	m4
human	0.16	0.4	0.38	0.47	0.18	-0.05	-0.12	-0.16	-0.09
interfac	0.14	0.37	0.33	0.4	0.16	-0.03	-0.07	-0.1	-0.04
e	0.15	0.51	0.36	0.41	0.24	0.02	0.06	0.09	0.12
comput	0.26	0.84	0.61	0.7	0.39	0.03	0.08	0.12	0.19
er	0.45	1.23	1.05	1.27	0.56	-0.07	-0.15	-0.21	-0.05
respons	0.16	0.58	0.38	0.42	0.28	0.06	0.13	0.19	0.22
e	0.16	0.58	0.38	0.42	0.28	0.06	0.13	0.19	0.22
time	0.22	0.55	0.51	0.63	0.24	-0.07	-0.14	-0.2	-0.11
EPS	0.1	0.53	0.23	0.21	0.27	0.14	0.31	0.44	0.42
survey	-0.06	0.23	-0.14	-0.27	0.14	0.24	0.55	0.77	0.66
trees	-0.06	0.34	-0.15	-0.3	0.2	0.31	0.69	0.98	0.85
graph	-0.04	0.25	-0.1	-0.21	0.15	0.22	0.5	0.71	0.62
minors									

Final SVD Matrix

Strongly correlated (r = .9)
 Weakly correlated (r = -.9)

- An index is a search engine component which contains a copy of website information harvested by web spiders.
- Indexing is the process through which web spiders gather and index web data to search engine's database.
- Latent Semantic Indexing acts as a process of representing the indexed data by producing an SVD term-document matrix containing indexed documents as columns as search keys and rows.