

Search Engines: Research Tools or Dead-ends?

A Discussion Paper

by

Richard Giroday

MDDE615 – Group C

Search Engines: Research Tools or Dead-ends?

The Asian Tsunami of December 26, 2004, is a major disaster that captured global attention in a way never seen before in the world. Its global impact continues to unfold. This disaster impressed itself upon the hearts and minds of people throughout the world. From large communities to small communities, technology opened a two-way window: revealing the pain and suffering of humankind never before brought to light in such a way, nor on such a magnitude; and awakening the hearts of humanity to reach out with love, caring, and compassion. Many of the global community, who witnessed those events from afar, were and are deeply troubled, and moved to compassion, empathy, sadness, and sorrow for those who lived and continue to live through the impact of this disaster. The global community's response and its depth, raises a number of questions.

- What was the catalyst for the growth of this awareness and response?
- What factors contributed to the development of such a global awareness?
- How quickly did it develop?
- How long will the awareness and reaction to this disaster be sustained?
- Is the response waning?
- Why was awareness and response to this particular disaster so great? Other disasters throughout the world have been as tragic, if not more so, yet they have not had the level of impact and response.
- How great an impact did this disaster have on the world?
- What can be learned from examination of this development?
- What tool can be used to measure the development?

Measuring Global Awareness Development

An indicator of the rate of growth and decay of media coverage, interest, and presentation of the information surrounding this disaster may open awareness of how media coverage affects a global humanitarian response to disasters throughout the world. Measuring the rate of growth and decay of media coverage is a difficult challenge at best, impossible at worst. However, one medium lends itself to just such an analysis – the Internet. One indicator of the rate of growth and decay of media coverage is the number of web pages generated and/or modified daily in response to a particular disaster.

The following presents an analysis of the rate at which web pages have been created or modified on topics concerning the Asian Tsunami disaster. The data was gathered using the Altavista Search Engine. Using advanced search features, the number of web pages created for a particular topic or modified, on any given day, for the period indicated following the disaster, has been determined. The numbers were then processed and graphed using Excel spreadsheets, to provide a visual indicator of the rate at which web pages have been created and/or modified. The data was gathered on February 12, 2005 (see Appendix A and B), and then re-gathered again on March 24, 2005 (see Appendix C and D).

Observations

Examination of the raw data reveals a glaring discrepancy between the data gathered on the two dates. For example, consider the following data obtained on the two different dates, using the same search criteria, obtained through AltaVista's advanced Search Engine.

#	Date under scrutiny	Pages Created or Modified as recorded Feb 12, 2005 (from Appendix B)	Pages Created or Modified as recorded Mar 24, 2005 (from Appendix D)	Differential
1	December 30, 2004	100,000	68,100	Drop of 31,900
2	January 20, 2005	1,060,000	295,000	Drop of 765,000
3	February 10, 2005	209,000	318,000	Rise of 109,000

The inconsistency of the data begs the questions: Why?

It could be argued that in the case of #3 in the table above, that in the case of the Feb 12th data, the spiders that catalogue and “find” the pages on the Internet had only the time period from Feb 10 – Feb 12 to find pages created or modified on Feb 10th, whereas in the case of the Mar 24th data, the spiders had a window of opportunity to continue working that extended from Feb 10th – March 24th. A rise would be anticipated under such circumstances.

However, in the case of #1 and #2 above, it is hard to determine a reason for the significant drop in numbers. Where did these newly created or modified pages disappear to? How was the search engine “spiders” so quickly able to identify their disappearance? Why are the numbers so significantly different a month later?

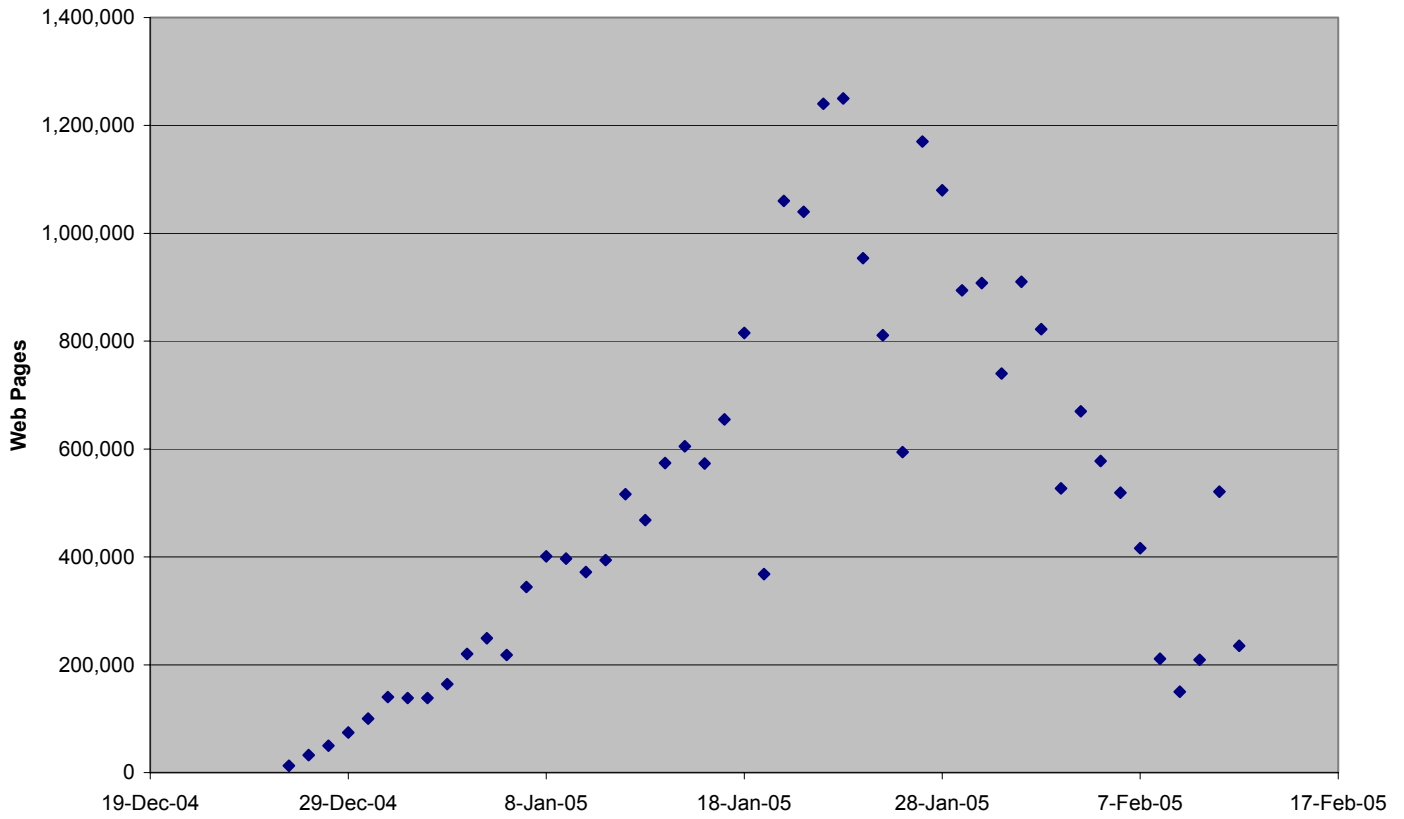
Conclusion

As powerful and useful as Search Engines are, in the access and identification of information on the Internet, their usefulness as an investigative tool is truly suspect. Instead of using the AltaVista Search Engines advanced search by date feature, perhaps recording a daily searching over the same time frame would have been better. However, in light of the changing nature of Search Engine databases, such a method would also be highly suspect in gathering data. It had been hoped that a day by day analysis of the changing face of the Internet as viewed by specific Tsunami topic keyword searches, would provide a weather vane indicator of the rate of growth and decay of media coverage, and how that coverage could be or may be affecting the global humanitarian response to disasters throughout the world. The results indicate that Search Engines have yet to develop attributes that will lend them to substantive investigative research of this nature.

Appendix A

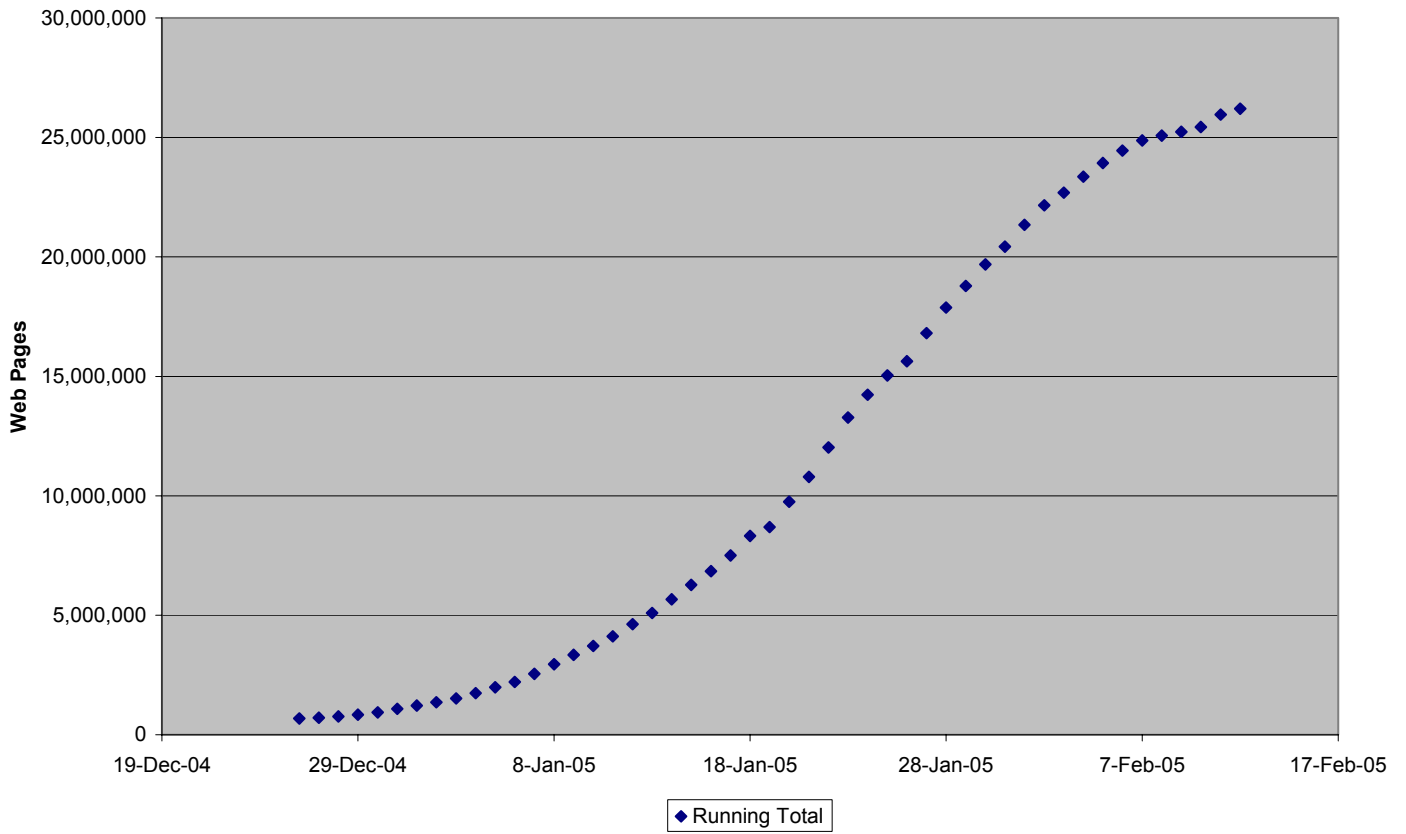
Graph of Data recorded February 12, 2005

Number of Web Pages Added or Modified Each Day
Keyword Search = "Tsunami"



Graph of Data Recorded February 12, 2005

Web Pages Added or Modified - Cumulative Total
Keyword Search = "Tsunami"



Appendix B

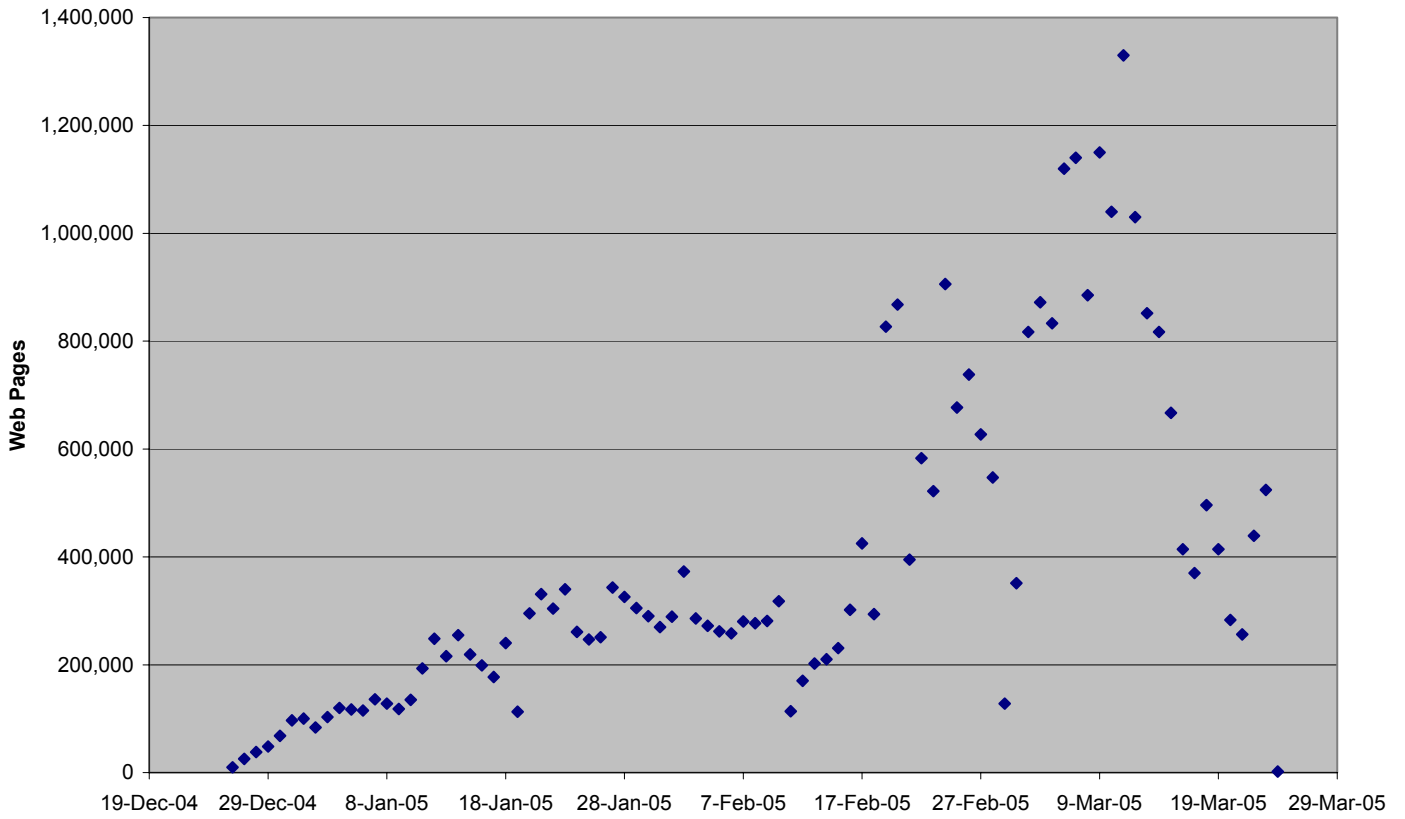
Raw Data Recorded February 12, 2005

Keywords Searched:	Tsunami	Running Total
Date	Pages added daily	
Pre 26-Dec-04	669,000	669,000
26-Dec-04	12,500	681,500
27-Dec-04	32,500	714,000
28-Dec-04	49,800	763,800
29-Dec-04	74,400	838,200
30-Dec-04	100,000	938,200
31-Dec-04	140,000	1,078,200
1-Jan-05	138,000	1,216,200
2-Jan-05	138,000	1,354,200
3-Jan-05	164,000	1,518,200
4-Jan-05	220,000	1,738,200
5-Jan-05	249,000	1,987,200
6-Jan-05	218,000	2,205,200
7-Jan-05	344,000	2,549,200
8-Jan-05	401,000	2,950,200
9-Jan-05	397,000	3,347,200
10-Jan-05	372,000	3,719,200
11-Jan-05	394,000	4,113,200
12-Jan-05	516,000	4,629,200
13-Jan-05	468,000	5,097,200
14-Jan-05	574,000	5,671,200
15-Jan-05	605,000	6,276,200
16-Jan-05	573,000	6,849,200
17-Jan-05	655,000	7,504,200
18-Jan-05	815,000	8,319,200
19-Jan-05	368,000	8,687,200
20-Jan-05	1,060,000	9,747,200
21-Jan-05	1,040,000	10,787,200
22-Jan-05	1,240,000	12,027,200
23-Jan-05	1,250,000	13,277,200
24-Jan-05	954,000	14,231,200
25-Jan-05	811,000	15,042,200
26-Jan-05	594,000	15,636,200
27-Jan-05	1,170,000	16,806,200
28-Jan-05	1,080,000	17,886,200
29-Jan-05	894,000	18,780,200
30-Jan-05	908,000	19,688,200
31-Jan-05	740,000	20,428,200
1-Feb-05	910,000	21,338,200
2-Feb-05	822,000	22,160,200
3-Feb-05	527,000	22,687,200
4-Feb-05	670,000	23,357,200
5-Feb-05	578,000	23,935,200
6-Feb-05	519,000	24,454,200
7-Feb-05	416,000	24,870,200
8-Feb-05	211,000	25,081,200
9-Feb-05	150,000	25,231,200
10-Feb-05	209,000	25,440,200
11-Feb-05	521,000	25,961,200
12-Feb-05	235,000	26,196,200

Appendix C

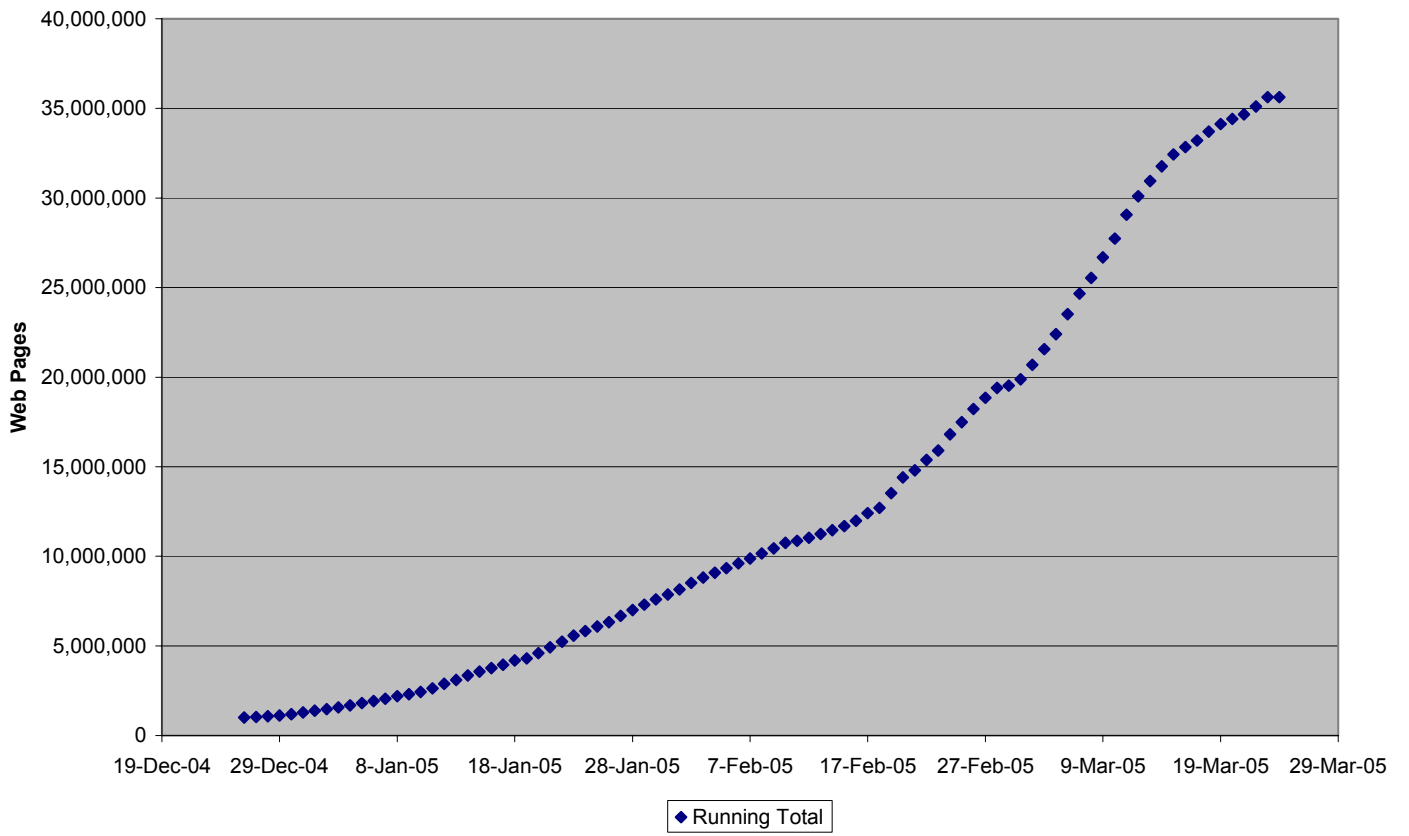
Graph of Data Recorded March 24, 2005

Number of Web Pages Added or Modified Each Day
Keyword Search = "Tsunami"



Graph of Data Recorded March 24, 2005

Web Pages Added or Modified - Cummulative Total
Keyword Search = "Tsunami"



Appendix D

Raw Data Gathered March 24, 2005

Keywords Searched:	Tsunami	Running Total
Date	Pages added daily	
Pre 26-Dec-04	997,000	997,000
26-Dec-04	10,000	1,007,000
27-Dec-04	25,200	1,032,200
28-Dec-04	38,100	1,070,300
29-Dec-04	48,400	1,118,700
30-Dec-04	68,100	1,186,800
31-Dec-04	96,900	1,283,700
1-Jan-05	100,000	1,383,700
2-Jan-05	83,800	1,467,500
3-Jan-05	103,000	1,570,500
4-Jan-05	120,000	1,690,500
5-Jan-05	117,000	1,807,500
6-Jan-05	115,000	1,922,500
7-Jan-05	136,000	2,058,500
8-Jan-05	128,000	2,186,500
9-Jan-05	118,000	2,304,500
10-Jan-05	135,000	2,439,500
11-Jan-05	193,000	2,632,500
12-Jan-05	248,000	2,880,500
13-Jan-05	216,000	3,096,500
14-Jan-05	255,000	3,351,500
15-Jan-05	219,000	3,570,500
16-Jan-05	199,000	3,769,500
17-Jan-05	177,000	3,946,500
18-Jan-05	240,000	4,186,500
19-Jan-05	113,000	4,299,500
20-Jan-05	295,000	4,594,500
21-Jan-05	331,000	4,925,500
22-Jan-05	304,000	5,229,500
23-Jan-05	340,000	5,569,500
24-Jan-05	261,000	5,830,500
25-Jan-05	247,000	6,077,500
26-Jan-05	251,000	6,328,500
27-Jan-05	343,000	6,671,500
28-Jan-05	326,000	6,997,500
29-Jan-05	305,000	7,302,500
30-Jan-05	290,000	7,592,500
31-Jan-05	270,000	7,862,500
1-Feb-05	289,000	8,151,500
2-Feb-05	373,000	8,524,500
3-Feb-05	286,000	8,810,500
4-Feb-05	272,000	9,082,500
5-Feb-05	262,000	9,344,500
6-Feb-05	258,000	9,602,500
7-Feb-05	280,000	9,882,500
8-Feb-05	277,000	10,159,500
9-Feb-05	281,000	10,440,500
10-Feb-05	318,000	10,758,500
11-Feb-05	114,000	10,872,500
12-Feb-05	170,000	11,042,500

13-Feb-05	202,000	11,244,500
14-Feb-05	210,000	11,454,500
15-Feb-05	231,000	11,685,500
16-Feb-05	302,000	11,987,500
17-Feb-05	425,000	12,412,500
18-Feb-05	294,000	12,706,500
19-Feb-05	827,000	13,533,500
20-Feb-05	868,000	14,401,500
21-Feb-05	395,000	14,796,500
22-Feb-05	583,000	15,379,500
23-Feb-05	522,000	15,901,500
24-Feb-05	906,000	16,807,500
25-Feb-05	677,000	17,484,500
26-Feb-05	738,000	18,222,500
27-Feb-05	627,000	18,849,500
28-Feb-05	547,000	19,396,500
1-Mar-05	128,000	19,524,500
2-Mar-05	351,000	19,875,500
3-Mar-05	817,000	20,692,500
4-Mar-05	872,000	21,564,500
5-Mar-05	833,000	22,397,500
6-Mar-05	1,120,000	23,517,500
7-Mar-05	1,140,000	24,657,500
8-Mar-05	885,000	25,542,500
9-Mar-05	1,150,000	26,692,500
10-Mar-05	1,040,000	27,732,500
11-Mar-05	1,330,000	29,062,500
12-Mar-05	1,030,000	30,092,500
13-Mar-05	852,000	30,944,500
14-Mar-05	817,000	31,761,500
15-Mar-05	667,000	32,428,500
16-Mar-05	414,000	32,842,500
17-Mar-05	370,000	33,212,500
18-Mar-05	496,000	33,708,500
19-Mar-05	414,000	34,122,500
20-Mar-05	283,000	34,405,500
21-Mar-05	256,000	34,661,500
22-Mar-05	439,000	35,100,500
23-Mar-05	524,000	35,624,500
24-Mar-05	2,030	35,626,530