

# WEB-BASED DIRECTORIES WITH SPATIAL SEARCH FUNCTIONALITY: AN ANALYSIS OF INTER- AND INTRA-DIRECTORY PERFORMANCE

James E. Wyse  
Faculty of Business Administration  
Memorial University

*The spatial search capabilities of web-based business directories provide locational information to directory-using customers on the whereabouts of businesses with consumer-targeted products and services. This article compares the performance of two web-based directories with respect to their provision of transaction-supporting locational information. Results reveal considerable variability in performance both within and across the two directories with respect to a group of 120 businesses in an industrial setting where 'location' is generally a material transactional concern. The article discusses the findings and identifies areas for further research.*

## **Business Directories: Past and Present**

Business directories appeared shortly after the telecommunications industry became a commercial reality. In the early 1880's, the Cheyenne Wyoming Telephone Company published a directory that not only listed the telephone numbers for its business subscribers but also included advertisements for goods and services. Because the printer had run out of white paper, the directory was printed on yellow paper. This unintentional action remains captured in what has arguably become the most widely used term for business directories in present times, the *yellow pages*. In the intervening century or so, yellow paper has done well; by the late 1990's there were over 360 million copies of over 6,000 print-based business directories in the U. S. alone with 60% of the U.S. population making use of business directories on a weekly basis (Eder, 1999). Although the emergence of the Internet may dampen the demand for yellow paper, it appears not to have a similar effect on the proliferation and evolution of business directories. Capturing both the unintended medium of business directory beginnings and the present day mode of business directory access, the term 'online yellow pages' refers to web-based directories whose content and functionality find extensive use in the wired world of recent times. Krauss (2001) reports that the web-based directory SuperPages.com conducts more than 20 million directory searches a month initiated by 4.5 million unique users. YellowPages.ca (2005) reports that it had 3.8 million visitors searching its online directories in July 2004. These reports indicate not only the extent of directory usage but also that directory-provided search functionality is used by a large and growing number of Internet-enabled consumers.

A recent development in directory-provided search functionality is the offering of geospatial search capability to directory users. Figure 1 contains a web page segment for SuperPages.ca showing "search options" for such geospatial functionality as "map-based" searching and "distance" searching. These "new" search capabilities support directory-based searches for goods and services within a geographic area of interest to a directory user. The area may be near a user's place of residence or it may be a place that the user wishes, or plans, to visit. Although the emergence of geospatial capabilities is a recent development, it is consistent with the long-held notion that yellow pages directories serve to connect local buyers and sellers (Krauss, 2001). Furthermore, the increasing adoption of geospatial search capabilities by web-based directories (Graham, 2002) indicates that this kind of directory-provided functionality is well on its way to becoming a standard, and indeed expected, directory feature.

**Figure 1**  
**Geospatial Search Functionality - SuperPages.ca**

Canadian Business Finder	
Simple Search	Search Options
<p><b>Category</b> (<a href="#">browse</a>)</p> <input type="text"/>	<p><a href="#">Phone Number (National)</a></p> <p><a href="#">Distance</a></p> <p><a href="#">Map Based Search (New)</a></p> <p><a href="#">Business Name (National)</a></p> <p><a href="#">Business Type</a></p> <p><a href="#">Simple Search</a></p> <p><a href="#">U.S. Businesses</a></p> <p>Make SuperPages.ca your <a href="#">HomePage</a>.</p>
<p><b>Or Business name</b></p> <input type="text"/>	
<p><b>City</b></p> <input type="text"/>	
<p><b>Prov/Territory</b> (required)</p> <input type="text" value="(choose)"/>	
<p><input type="button" value="Find It"/> <input type="button" value="Clear"/></p> <p><a href="#">Search Tips</a></p> <p><input type="checkbox"/> make Simple Search my default</p>	

(Accessed March 7, 2005)

Despite its increasing popularity, there are concerns that directory-reported locational information may inadequately meet the transactional support requirements of businesses and their customers (Wyse, 2004). Graham (2002) reports there are often substantial gaps in geographical coverage and significant inaccuracies in existing coverage. These shortcomings are somewhat unexpected given the apparent reliance placed on directory-provided locational information by directory users (Graham, 2002) and the level of effort expended by firms that specialize in its collection.<sup>1</sup> This brings us to the general area of concern for this report: the quality of locational information provided by web-based business directories. More specifically, this report addresses the issue of locational information quality through an examination of how well two prominent web-based ‘yellow page’ directories perform when reporting business locations (i.e., when ‘geo-placing’ a business). In what follows, the geo-placement performance of the two directories is determined, analyzed, and compared on the basis of observed positional discrepancies between actual and directory-reported locations for a selected group of businesses.

**Methodology and Results**

Observations of the positional discrepancies between actual and directory-reported locations were obtained using a six-step method devised by Wyse (2004) in his investigation of the geo-placement of small enterprises by web-based directories: (1) a group of businesses was selected for which ‘location’ would often be considered a material transactional concern, (2) web-based directories were identified with comparable geospatial functionality and with business listings that purport to cover the region containing the selected businesses, (3) the geographical coordinate values for the location of each business were extracted from the two directories, (4) the coordinate values for the actual location of each business were measured in the field, (5) the positional discrepancies between actual and directory-reported locations were calculated, and (6) the geo-placement performance of the two directories was compared based on an analysis of the observed positional discrepancies.

**Business Selection:** Location is typically a material transactional concern for businesses providing hospitality services. For this study, a list of such businesses was obtained from a travel guide published by the tourism authority for a province in eastern Canada. A convenience

<sup>1</sup> Graham (2002) reports that Chicago-based NavTech “supplements its satellite-based information with 400 staffers who drive across the USA checking [locational] data.”

sample of 120 businesses was selected from the travel guide based on the feasibility of traveling to business sites to obtain measurements of the geographical position of each business in the sample.

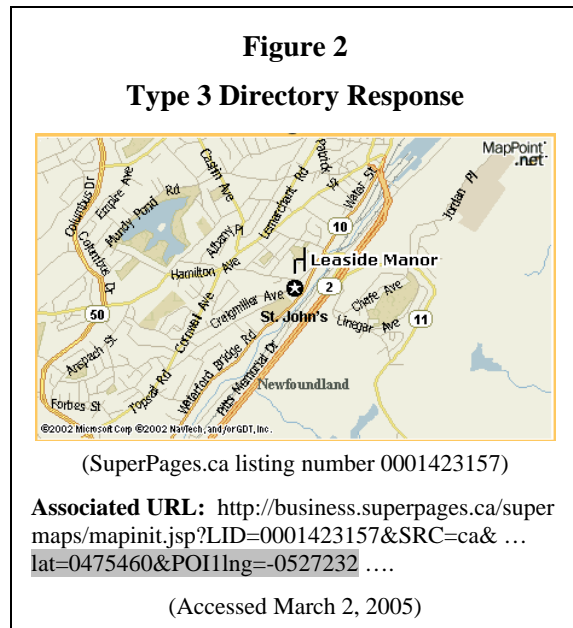
**Directory Identification:** Superpages.ca (“SP”) and YellowPages.ca (“YP”) are web-based business directories with user-oriented geospatial functionality that purport to provide listings in the region containing the sample of businesses. The two directories are also comparable in several other respects: both are ‘general’ business directories in that they neither exclusively nor primarily focus on the hospitality industry; both advertise that their listings cover all the regions of Canada; and both have affiliated ‘411’ websites.<sup>2</sup>

**Extraction of Directory-Reported Locations:** The two directories were queried for information on each of the selected businesses. Each query was met with one of three possible responses: (1) no listing was found for the business about which information is requested (Type 1 Directory Response), (2) a listing was found but the business is not ‘geo-placed’ on a map by the directory (Type 2 Directory Response), or (3) a listing was found with the location of the business explicitly indicated on a location-centric map (Type 3 Directory Response). A sample Type 3 Directory Response is shown in Figure 2. This response is the result of a query to the SuperPages.ca directory about Leaside Manor, an 8-unit urban-based inn located in eastern Canada. Associated with each location-centric map is a URL (partially shown in the lower portion of Figure 2) that bears the geographical coordinates used by the directory to determine the location of the mapped business.

Table 1 summarizes the directory responses that resulted from querying the two directories about each of the 120 businesses. Figures for Type 3 responses show that both directories geo-place some portion of the 120 businesses by showing maps centered on their locations: 68% are geo-placed by SP while 49% are geo-placed by YP. Type 2 response figures reveal that 28% of the businesses are listed with YP but not geo-placed. No Type 2 responses resulted from queries to the SP directory; thus, whenever a business was listed with SP it was always geo-placed. Type 1 responses show that neither directory lists all 120 businesses. The results in Table 1 indicate that YP’s directory coverage exceeds that of SP; however, the extent to which SP geo-places the 120 businesses exceeds that for YP.

---

<sup>2</sup> SP is affiliated with 411.ca while YP is affiliated with Canada411.ca. Although the 411 web sites are visually and functionally different from their affiliated yellow page directories, they draw upon the same locational information resources as their affiliates. Thus, for purposes of this investigation, the 411 web sites were not considered business directories that are separate and distinct from their respective yellow page affiliates.



**Table 1**  
**Directory Response Distributions**

Response Type	<u>YP</u>	<u>SP</u>
1. No listing provided.	23% (28)	33% (39)
2. Listing with no locational info.	28% (33)	0% (0)
3. Location-centric map displayed.	49% (59)	68% (81)

Note: Rounding may result in percentages across the four response types not adding to 100.

**Actual Location Determination:** The goal for determining the actual location of each business was one of obtaining coordinate values for a position that would place a customer with no previous visitation experience within ‘human-eye’ visual range (20 meters or less) of a business’s service-providing location. Whenever the location of a business was not visually available (by, for example, being situated at the end of a long pathway and/or obscured by trees), then a location with clearly displayed signage was used as a proxy for the business’s location. The GPS coordinates for each business were obtained by an observer whose self-positioning in the vicinity of a business was guided by the ‘no previous visitation’ assumption.

**Displacement Determination:** The displacement of a business’s actual location from its directory-reported location was determined as the great circle distance between the two locations.<sup>3</sup> Table 2 provides a selection of descriptive statistics on the observed displacements for businesses with directory-reported geographical coordinates (i.e., those businesses associated with a Type 3 Directory Response). Thus, displacements could only be determined for 59 of the businesses in YP’s case and for 81 of the businesses in SP’s case. Given the sensitivity of average displacements to small numbers of highly displaced businesses, percentile values were felt be appropriate indicators of the extent to which each directory tends to displace a business. The displacement percentiles shown in Table 2 indicate that 50% of YP’s displacements exceed 239 metres while 50% of SP’s displacements exceed 776 metres. The 90<sup>th</sup> percentile values show that 10% of the YP’s displacements exceed 8,948 metres while 10% of SP’s displacements exceed 4,243 metres. Figures in the last row of Table 2 show the percentage of directory-reported locations that are displaced by more than one kilometre from their actual locations: 37% (or 22) for YP and 46% (or 37) for SP. Results here indicate that displacement variability is greater for YP than for SP; however, YP displaces fewer of the selected businesses beyond one kilometre than does SP.

<sup>3</sup> The great circle distance is essentially the length of the shortest path over the earth’s surface between the two locations. A discussion on the determination of great circle distances may be found in Bonham-Carter (1994).

Table 3 compares the two directories on the basis of the extent to which directory-reported locations lie within a specified radius of their actual locations. The first row reports the percentages and numbers of businesses whose directory-reported locations are within a 100-metre radius of their actual locations. Figures here show that 41% or 24 of the businesses are displaced 100 metres or less by YP while 32% or 26 are displaced 100 metres or less by SP. An examination of figures associated with the other radii in Table 3 reveal that for radii under 5 kilometres, YP places proportionately more of the businesses closer to their actual locations, while for the 5 and 10 kilometre radii SP places more of the businesses closer to their actual locations. Although transactions between a business and its customers may not be unduly affected for businesses within the 100 metres radius, the results associated with the other radii suggest that there are likely to be many businesses for which transactions could, to varying degrees, be placed at risk by the locational information reported by the two directories.

**Table 2**  
**Selected Statistics on**  
**Business Location Displacements**  
 (Type 3 Directory Responses.)

	<u>YP</u>	<u>SP</u>
N	59	81
Average Displacement (metres)	8,294	2,276
50 <sup>th</sup> Percentile (metres)	239	776
90 <sup>th</sup> Percentile (metres)	8,948	4,243
Displacements greater than one kilometre	37% (22)	46% (37)

The results in Table 4 address the impact of the observed displacements on the integrity of the location-centric maps displayed by the directories. (Figure 2 shows an example.) The issue here is whether a displayed map, which explicitly indicates the location of a business, truly contains the actual location of the business *anywhere* in the area depicted by the map. The first row of Table 4 indicates that 78% of the businesses listed with YP are actually located *somewhere* on the displayed map; thus 22% of the businesses are not in fact anywhere on the maps that explicitly indicate that they are. The equivalent figures for SP are 63% or 51 business actually located somewhere on the map with 37% of the businesses not anywhere on the displayed map. Furthermore, user functionality associated with the directory-displayed maps provides a directory user with the option of ‘zooming in’ to obtain a map covering a smaller area (i.e., a larger-scale map). Whenever this option is chosen, the name of the business remains visually associated with the larger-scale (i.e., zoomed-in) map and the location of the business remains explicitly indicated at the center of the map. A user would easily be left with the impression that the ‘zoomed-in’ map contains the sought-after business location somewhere within map’s boundaries (and likely at its center). Results shown in the second row of Table 4 indicate that this impression would be correct for only 53% or 31 of the businesses for YP and 46% or 37 of the businesses for SP. Although YP’s performance in this respect may be somewhat better than SP’s; these results suggest that both directories provide map-based locational information that is of questionable integrity.

**Table 3**  
**Radial Displacements**

Location within:	YP (N = 59)	SP (N = 81)
100 m	41% (24)	32% (26)
500 m	53% (31)	44% (36)
1 km	63% (37)	54% (44)
5 km	83% (49)	90% (73)
10 km	90% (53)	96% (78)

**Table 4**  
**Directory Map Integrity**

Actual location is ...	YP (N = 59)	SP (N = 81)
... somewhere on the map initially displayed.	78% (46)	63% (51)
... somewhere on the largest scale map.	53% (31)	46% (37)

**Common Subset Analysis:** Comparing the two directories based on the results shown in Tables 2, 3, and 4 is somewhat confounded by the use of differing subsets of the businesses in the selected sample. The results shown in Tables 5, 6, and 7 are derived from a subset of the sample consisting of the 54 ‘map-associated’ businesses that are common to both directories. Results here reveal levels of misplacement similar to those revealed in the results seen in Tables 2, 3, and 4; however, the divergence between directories is considerably reduced. The results in Table 5, like those seen in Tables 2, indicate that YP shows greater displacement variability than SP; however, unlike the results in Table 2, YP no longer displaces fewer of the businesses beyond one kilometre and the divergence between YP and SP is now reduced from 9% (37% vs. 46%) to 2% (37% vs. 35%). Figures in Table 6, like those in Table 3, compare the directories on the basis of the extent to which directory-reported locations lie within a specified radius of their actual locations. A comparison of Tables 6 and 3 shows a convergence of the extents of radial displacement across the two directories for most of the selected businesses in the common subset. A similar outcome for map integrity is seen from a comparison of the results in Tables 4 and 7 with the difference between the two directories reduced in the ‘map initially displayed’ case and completely eliminated in the ‘largest scale map’ case. Despite the reduction in the differences between the map integrity figures, results here also suggest that both directories provide map-based locational information that is of questionable integrity.

**Table 5**  
**Selected Statistics**  
**(Common Subset)**

	YP	SP
N	54	54
Average Displacement (metres)	7,002	1,274
50 <sup>th</sup> Percentile (metres)	239	338
90 <sup>th</sup> Percentile (metres)	8,246	3,938
Displacements greater than one kilometre	37% (20)	35% (19)

**Table 6**  
**Radial Displacements**  
**(Common Subset)**

Location within:	YP (N = 54)	SP (N = 54)
100 m	41% (22)	39% (21)
500 m	54% (29)	52% (28)
1 km	63% (34)	65% (35)
5 km	83% (45)	93% (50)
10 km	91% (49)	100% (54)

**Table 7**  
**Map Integrity**  
**(Common Subset)**

Actual location is ...	YP (N = 54)	SP (N = 54)
... somewhere on the map initially displayed.	78% (42)	70% (38)
... somewhere on the largest scale map.	54% (29)	54% (29)

**Discussion of Geo-Placement Performance**

The design of the study does not support generalizations of the geo-placement performance findings to larger populations of directory-listed businesses nor does it permit the drawing of any statistical inference about the significance of geo-placement differences between directories. However, the results provide limited evidence that (1) overall inter-directory differences in geo-placement performance are minimal and (2) that many directory-listed businesses are misplaced to an extent that may jeopardize transactions between these businesses and their customers. The first outcome (minimal inter-directory differences) may be largely attributable to locational approximations used by geocoding methodology (i.e., the methods by which business locations are associated with specific values of latitude and longitude). Geocoding methodologies often misplace locations but do so in a consistent fashion that tends to contribute to a convergence in locational misplacement across the two directories. The second outcome (transaction-jeopardizing locational misplacement) may be attributable (as discussed below) to a combination of factors, including the locational approximations used by geocoding methodology.

The methods utilized by geocoders often operate on the address code (zip code, postal code) that is associated with a business’s address. Geo-coding transforms the address code for the *area* within which the business is actually located into positional coordinates for a *point* at which the business is assumed to be located. Not infrequently, some centroidal point in the area is arbitrarily taken as the address’s location. In certain geographical settings (e.g., highly urbanized, densely populated city blocks), an area that corresponds to an address code is generally smaller than the area that corresponds to an address code in certain other geographical settings (e.g., sparsely populated, rural areas). Thus, the positional coordinates resulting from the geocoding of addresses tend to misplace business locations to a greater extent for address codes associated with larger areas than for address codes associated with smaller areas. An examination of the locational data extracted from directory-reported URLs revealed instances where the same locational (point) coordinates are assigned to several businesses whose individual actual locations were not in close proximity to one another but whose postal codes are the same. Although stronger statements about the effect of geocoding methodology on business location misplacement must await further research, the consistency and extent of the misplacements suggest that geocoding methods account to an appreciable extent for business misplacement as well as for the minimal misplacement differences between the two directories.

In addition to misplacement attributable to geocoding methods, an examination of the data revealed instances of business location misplacement that may be attributable to (at least) three other factors: (1) *data entry errors* arising from inadequate quality control (for example, an inadvertent transposition of digits during recording or transcribing positional measurements or address codes), (2) *place-name misinterpretations* whereby a business in one location is mistakenly assumed to be in another similarly named location (for example, a business in Placentia, NL, Canada is taken to be located in Placentia, CA, USA), and (3) *address mis-designations* wherein the address reported by the directory is indeed associated with the business but the address does not correspond to a business's service-providing location (for example, a when directory listing shows an off-season urban address for a small rural inn that only operates during the summer).

Instances of all three misplacement factors appear in the data collected for the study. YP's maximum observed discrepancy (177 km) appears attributable to a data-entry error (possibly the entry of an incorrect address code). Several place-name misinterpretations are suspected, with businesses being misplaced to locations with addresses (or parts thereof) similar to the correct address of the business. Instances of address mis-designation were observed and seemed to take two forms: (a) mis-designation attributable to the use of off-season contact addresses and (b) mis-designation attributable to the use of post box addresses (which result in a business appearing on a map at the location of a postal station). Although instances of locational misplacement attributable to each of these possibilities appear in the data collected, study limitations prevent assessments of the extent to which each misplacement possibility accounts for the results obtained and/or whether there exist other misplacement possibilities.

The two directories appear to be aware that their geospatial functionality produces locational information that often misplaces listed businesses. Both directories provide web site functionality that invites and accommodates the provision of corrected locational information. YP has a "My map is incorrect" category listed on its "Report an error" web page. SP offers to make corrections "if your business location is mapped in the wrong spot in a location map" on its "Map Tips" web page.<sup>4</sup> The observation of minimal inter-directory differences and high levels of locational misplacement suggest that listed business infrequently respond to the invitation to provide corrections. This may be a reflection of poor awareness on the part of businesses about the quality of the locational information associated with their directory listings. Informal contact with businesses in the sample suggests that business operators are only vaguely aware of their listings in web-based business directories. Other factors may also contribute to low correction submission rates: (a) skill inadequacies in the collection and submission of corrected location information, (b) a view on the part of businesses that web-based directories are simply unimportant, (c) insufficient exposure of the invitation by directories to submit corrected locational information, (d) absence of directory functionality to assist businesses with a proper determination of their locational coordinates, and (e) other possible factors. Here as well, the design of the study does not support the identification of factors in this respect.

## Further Research

Of the many areas where further work could be productively undertaken, three emerge from the foregoing discussion as areas where more should be known: (1) geo-placement performance assessments in other geographical and industrial settings, (2) examinations of the reasons and their relative impacts on business location misplacement, and (3) investigations of the factors

---

<sup>4</sup> YP's web page in this respect may be accessed at <http://www.contactus.yellowpages.ca/yellowpages/electronic/03.html> while that for SP may be accessed at <http://business.superpages.ca/supermaps/tips.jsp?SRC=ca&STYPE=S&source=map>. [Both pages accessed on March 2, 2005.]



contributing to the non-participation of directory-listed businesses in the review and correction of their locational information.

**Assessments in Other Settings:** Studies in other geographical and industrial settings will indicate whether the results obtained here would be realized in a wider range of circumstances or whether they are largely artifacts of the businesses selected or the geographical setting in which the sample is situated. The region used for this investigation would likely be regarded by many as relatively isolated. Thus, business directories may be disinclined to invest the resources required to bolster the veracity of their locational information for directory-listed businesses in such regions. Furthermore, the businesses used in the study were for the most part very small enterprises and they were taken from a single industrial setting. Also, they were selected based on the feasibility of traveling to their actual locations to obtain positional readings. Consequently, it is important before proceeding with research in other respects that substantive levels of misplacement are shown to exist in settings other than the specific geographical and industrial setting used here.

**Reasons for Misplacement:** An important basis of knowledge for improving a directory's geo-placement performance is an understanding of the reasons for misplacement and their relative importance. A variety of misplacement reasons (geo-coding approximation, data entry error, geographical mis-interpretation, and address mis-designation) were identified and discussed above and it is entirely possible that there are other misplacement reasons. A knowledge of the reasons will help determine appropriate courses of corrective action: misplacements attributable to geocoding approximation will focus corrective action on improvements in (or selective use of) geocoding methodology, misplacements attributable to data entry errors will focus corrective action on quality controls in data capture processes, and so on. Thus, research focused on misplacement reasons, both their revelation and their relative impact, will identify actions that would most effectively improve a directory's geo-placement performance.

**Non-Participation in Corrective Procedures:** If all directory-listed businesses were to review their locational information and submit any required corrections, a significant improvement in geo-placement performance would likely be realized. Furthermore, from a directory's perspective, such an occurrence would be a highly cost effective way to improve geo-placement performance since the substantial cost associated with obtaining valid positional measurements is effectively transferred from the directory to its listed businesses. Although directories invite corrections from listed businesses, there is much to suggest that businesses are not responding. Previous discussions in this respect explored some of the factors that might account for this non-participation with simple lack of awareness seen as a major factor. Further research in this respect could assess the importance of the factors identified, reveal other factors, and suggest appropriate actions to bolster the participation of listed businesses in an activity with considerable potential to cost effectively improve a directory's geo-placement performance and thereby enhance the value of its locational information in facilitating transactions between location-constrained businesses and their customers.

## References

- Bonham-Carter, G. *Geographic Information Systems for Geoscientists: Modelling with GIS*. Pergamon, 1994.
- Eder, P. "Big Changes Ahead for Telephone Directories." *The Futurist* 33 No. 4 (April 1999): 51-4.
- Graham, J. "Online Mapmakers: Popular, Prolific, Not Perfect." *USA Today* (Oct. 7, 2002.), pg. D.03.
- Krauss, M. "Once-Stogy Yellow Pages Now Mark Digital Leading Edge." *Marketing News* 35 No. 15 (July 16, 2001): 8.
- Wyse, J. "Badly Off in All Directions: The Geo-Placement of Small Enterprises by Web-Based Business Directories." *International Journal of Electronic Commerce*. Vol. 8., No. 4 (Summer 2004), 145-159.
- YellowPages<sup>TM</sup>.ca. "Advertise with us." [web page]. [http://www.yellowpages.ca/findLinks.do?bodyOfHtml=/fragments2/AWU-high\\_traffic\\_web\\_site.html&isBodyHTML=true&leftsidebar=/jsp2/component/minibox-advertise.jsp&isLeftsidebarJSP=true&subTitle=yellowPages&sub=highTraffic&top=/fragments2/topAdWithUs.html](http://www.yellowpages.ca/findLinks.do?bodyOfHtml=/fragments2/AWU-high_traffic_web_site.html&isBodyHTML=true&leftsidebar=/jsp2/component/minibox-advertise.jsp&isLeftsidebarJSP=true&subTitle=yellowPages&sub=highTraffic&top=/fragments2/topAdWithUs.html). [Accessed 1 Mar. 2005].