

# Analyzing the Prevalence of Sports-Related Terms among the Web Sites of Global Corporations

[2004, International Journal of Computer Science in Sport, 3(2): 5-18.]

*Arno Scharl • Larry Neale • Jamie Murphy*  
*{ascharl, lneale, jmurphy}@biz.uwa.edu.au*

*University of Western Australia, UWA Business School*  
*35 Stirling Highway, Crawley WA 6009, Australia*

## Abstract

This research investigates the prevalence of sports-related terms among the Web sites of the world's leading companies, the Fortune Global 500. An automated process copied about four gigabytes of textual data, around 70 million words, from their sites. The subsequent analysis revealed regional and industry differences in the distribution of sports-related terms, the popularity of tennis stars and few references to sports stars, especially in Asia.

**Keywords:** Sports Marketing, Sponsorship, Textual Analysis, Web Monitoring, Fortune Global 500

## Introduction

Sports sponsorship seeks favorable publicity for a company and its brands with a target audience (Bennett, 1999). Increased noise in print and broadcast media along with rising global interest in sports has pushed corporate sponsorship higher in recent years (Shank, 2002; Terrian, 2002). In Australia, corporate sport sponsorship doubled from 1996 to 2000, reaching US \$420 million. This figure omits an additional US \$420 million from the 2000 Sydney Olympics (Lloyd, 2000). Similar to the growth in sports sponsorship, there has been a corresponding increase in using the World Wide Web for internal and external corporate communication (Leichty & Esrock, 2001).

Researchers continue to study corporate sports sponsorship, but few investigate how corporate Web sites reinforce this sponsorship or sports in general. This study uses automated software tools to measure the frequency of sports-related terms among the Web sites of the world's top companies, the Fortune Global 500. After grouping the terms into meaningful associations, an exploratory analysis compares usage of these associations. Highlighting differences among Fortune Global 500 Web sites by country and industry give practitioners insights into the online presence of sports and give academics a basis for future research.

## Historical Development and Literature Review

Corporate sports sponsorship dates at least to the inaugural modern Olympic Games, Athens in 1896, when companies bought advertising space in the official Olympic program (Sandler & Shani, 1993). Regular sponsorship began in 1912, when Swedish companies acquired the permit to take photographs and sell souvenirs of the Stockholm Olympic Games (Papandropoulos, 2002). Coca-Cola was the first corporation to buy official Olympic sampling rights, at the St Moritz 1928 Winter Games (Stotlar, 1993).

Olympic sponsorship slumbered until the International Olympic Committee (IOC) and the city of Montreal lost money on the 1976 Summer Games. The US \$30 million deficit spurred the IOC to focus on sponsorship. Two factors hindered corporate sponsorship for the next summer games though, a US-led boycott and Moscow's communist environment. Sponsorship soared in 1984, helping the Los Angeles Summer Olympics earn a US \$225 million profit (Shaheeh, 1999; Stotlar, 1993).

Twenty years later, as the 2004 Summer Olympics returned to Athens, sponsorship was the main source of Olympic funding. As of August 2003, sponsorship revenues approached US \$500 million ([www.athens2004.com](http://www.athens2004.com)). The Athens 2004 sponsorship program offers specific rights and privileges depending on the category and size of the investment, such as Shell paying about US \$7 million to be the official fuel sponsor (Papandropoulos, 2002).

In addition to the Olympics, athletes across myriad sports benefit from increased corporate sponsorship. In May 2003, 18-year-old American high school basketball player LeBron James signed a seven-year Nike deal over US \$90 million. This falls short though, of Nike paying US \$100 million to golfer Tiger Woods. Tennis player Venus Williams has the women's sponsorship bragging rights, US \$40 million with footwear and apparel maker Reebok (Teather, 2003).

### *Corporate Sponsorship Objectives*

In exchange for sponsoring sports, corporations expect benefits. For example, one study shows a temporary boost in a company's stock price immediately after announcing stadium naming rights (Clark, Cornwell, & Pruitt, 2002). In practice, organizations use sports and sports stars in their marketing campaigns to reach some or all of the sponsorship objectives summarized in Table 1. While marketers debate the relative importance of these objectives, Sleight (1989) contends that personal objectives, such as management interest in the sport, are the least defensible reason for conducting a sport sponsorship campaign.

Table 1. Objectives of Corporate Sponsorship Campaigns (Pope, 1998)

<b>Corporate Objectives</b>	<b>Marketing Objectives</b>	<b>Media Objectives</b>	<b>Personal Objectives</b>
Public Awareness	Business Relations	Generate Visibility	Management Interest
Corporate Image	Reach Target Market	Generate Publicity	
Public Perception	Brand Positioning	Enhance Ad Campaign	
Community Involvement	Increase Sales	Avoid Clutter	
Financial Relations	Sampling	Target Specificity	
Client Entertainment			
Government Relations			
Employee Relations			
Competition			
Shareholder Wealth			

Sports researchers also differ on the benefits of corporate sponsorship (Pope, 1998). Some researchers argue that sponsorships should increase sales (Abratt, Clayton, & Pitt, 1987), while others argue for enhancing a company's image, product or brand (Armstrong, 1988; Javalgi, Traylor, Gross, & Lampman, 1994).

From a sponsorship perspective, sports and sports stars appeal to an international audience. Unlike competing entertainment such as cinema or music, international sports have standards and etiquette that transcend cultural, religious and linguistic barriers.

### *Modern Sports Marketing*

Shank (2002) defines sports marketing as applying marketing principles to products through association with sports. Estimates on the global value of sports marketing depend upon the variables included, such as sponsorships and revenue, but Shank estimates world sports marketing at approximately US \$350 billion in 2002. Thanks to the Internet, sports marketing takes an increasingly global perspective (Mullin, Hardy, & Sutton, 2000; Pope, Forrest, & Murphy, 1996; Summers, 2003).

The Atlanta Games of 1996 were the first to embrace Internet technology, and subsequent Olympics have continued this practice. As an example, the entire official Web site for the Athens Games ([www.athens2004.com](http://www.athens2004.com)) is available in Greek, English and French, illustrating modern sports marketing techniques and the importance of multilingual content to reach an international audience.

The Internet gives teams, leagues, fans and consumers a two-way communication platform, independent of time and location. For example, the US-based National Basketball Association (NBA) invited online fans, regardless of their country, to select the 1996 All-Star team. In addition to English, the NBA provided French, Spanish and Italian versions of the Web site ([www.nba.com](http://www.nba.com)). This multi-lingual initiative helps explain why one third of NBA's Web traffic during the All-Star selection originated outside the United States (Mullin et al., 2000).

While one expects large sports organizations such as the Olympics, NBA and the National Football League ([www.nfl.com](http://www.nfl.com)) to establish Web sites, research has neglected investigating the prevalence of sports-related terms on their Web sites.

## Objectives

This study investigates the use of sporting related terms on the Web sites of large multinational corporations. Three research questions guide the data collection:

1. How prevalent are sports-related terms (sports and the names of sports stars) among the Web sites of leading global companies?
2. Are there industry differences in the frequency of sports-related terms on the Web sites of leading global companies?
3. Are there regional differences in the frequency of sports-related terms on the Web sites of leading global companies?

## Methodology

To investigate the prevalence of sports-related terms on the Web sites of major corporations, this research studied the Web sites in the 2002 edition of the Fortune Global 500 ([www.fortune.com](http://www.fortune.com)). Researchers have used Fortune Magazine's rankings of the world's leading companies in disciplines such as business ethics (Morf, Schumacher, & Vitell, 1999; Reicher, Webb, & Thomas, 2000; Weaver, Treviño, & Cochran, 1999), health care (Montenegro-Tores, Engelhardt, Thamer, & Anderson, 2001), quality management (Baker, DeTienne, & Smart, 1998; Lawler III, Mohrman, & Ledford Jr., 1992), and international business (Gabba, Pan, & Ungson, 2002). Studies have also investigated the Web sites of Fortune-ranked companies from perspectives including content (Perry & Bodkin, 2000), marketing (Palmer & Griffith, 1998), global usage patterns (McManis, Ryker, & Cox, 2001), customer relationship management (Romano Jr., 2002-3), and email use (Leichty & Esrock, 2001).

### *Sports and Sports Stars*

Given the international focus of this study, the preliminary list of sports stemmed from those recognized by the IOC ([www.olympic.org](http://www.olympic.org)). Due to the preponderance of US companies in the Global 500 and English content in the remaining Web sites, popular US sports, general sports-related terms and popular sports in English-speaking countries such as cricket and rugby augmented the list of Olympic sports. Finally, given the exploratory nature of this study, the researchers added a few terms related to leisure (see Appendix).

The list of sports stars stemmed from the Laureus World Sports Awards ([www.laureus.com](http://www.laureus.com)). This annual event honors the world's best sportsmen and -women across sports and countries. This research used the names of 131 athletes nominated between 2000 and 2003 for the following six individual Laureus categories: Sportsman of the Year, Sportswoman of the Year, Newcomer of the Year, Comeback of the Year, Sports-person of the Year with a Disability, and Alternative Sportsperson of the Year.

### *Gathering Web Content*

Since the 1700s and across myriad media, scholars have used content analysis to deduce a medium's subject matter (Krippendorff, 1980). They have applied this technique to Web sites in general (McMillan, 2000; Scharl, 2000) and sports Web sites in particular (Pope et al., 1996). Web sites reflect industry trends and competitive strategies, but

methodological problems of objectivity and reliability hinder content analysis of Web sites (McMillan, 2000) and textual data (Rourke, Anderson, Garrison, & Archer, 2001).

Human coding, common on Fortune Web site studies (Leichty & Esrock, 2001; McManis et al., 2001; Palmer & Griffith, 1998; Perry & Bodkin, 2000), is time consuming, suffers from reliability issues and usually analyzes just the home page rather than the whole Web site (McMillan, 2000). Automating the coding process (Bauer & Scharl, 2000; Scharl, 2000; Thelwall, 2002) helps address this limitation, quickly and reliably processing large samples of Web sites.

Mirroring entire sites of major corporations, however, is resource-intensive. As information towards the top of a Web site reflects common use, most content analyses use just the site's home page (McMillan, 2000). Based on experiences from past research (Bauer & Scharl, 2000; Scharl & Bauer, 2004; Scharl, Pollach, & Bauer, 2003), this study used a limit of 10 megabytes to help manage available storage space and compare sites of heterogeneous size, but is by no means limited to this parameter. A robot started at the home page and then followed a site's hierarchical structure until amassing 10 megabytes of text. A site's markup code and embedded scripts guided the mirroring process to capture documents and build a hierarchical document tree.

The robot then wrote the textual content into one single text file for further processing. The size of this file can never reach or exceed the limit of ten megabytes, as the robot removes all the tags and scripts from the original set of documents. As a rule of thumb, ten megabytes of markup code result in about three to five megabytes of plain text.

Due to changes in the Fortune Global 500 since its publication, mergers for example, this April 2003 study began with 493 of the 500 companies. The robot could not process 77 Web sites for several reasons such as little textual information, inaccessibility, and parsing difficulties for technical reasons (e.g. applets or complex scripting elements). Mirroring the remaining 416 sites yielded almost four gigabytes of textual Web data, representing more than 270,000 documents with 70 million words.

Over two out of five (43%) of the remaining 416 sites represented US companies. Most other companies had their headquarters in Europe (31%) or Asia (24%). The predominant industry in these 416 sites was finance and insurance (24%), followed by resources (9%) and food/beverage/tobacco (8%). Each of the remaining sites belonged to one of the following industries: automotive, electronics, energy, engineering, information technology, media, pharmaceuticals, paper/freight, retail, telecommunications, travel, and wholesale.

### *Analyzing Web-based Corpora*

Corpora are collections of recorded content used for descriptive analysis. This research investigated and visualized regularities in the mirrored text by applying and extending methods from corpus linguistics and textual statistics (Biber, Conrad, & Reppen, 1998; Lebart, Salem, & Berry, 1998; McEnery & Wilson, 2001).

Quantitative textual analysis of Web documents necessitates three steps in order to yield a machine-readable representation (Lebart et al., 1998). The first step *converts* hypertext documents into plain text. The second step *segments* the textual chain into minimal units by removing coding ambiguities such as punctuation marks, the case of letters, hyphens, or points in abbreviations. The third step, *identification*, groups identical units and counts their occurrences – i.e., creating an inventory of words. This exhaustive in-

dex uses decreasing frequency of occurrence as the primary sorting criterion and lexicographic order as the secondary criterion.

Our perception of language relies on the recognition of words as units. Aligning grammar and vocabulary, words are the primary unit of lexical meaning (Sinclair, 2004). Despite a lack of contextual information, researchers use word frequencies to analyze both traditional (Leech, Rayson, & Wilson, 2001; McEnery & Wilson, 2001) and electronic (Meyer, Grabowski, Han, Mantzouranis, & Moses, 2003; Scharl & Bauer, 2004) corpora. For Web content analyses of large document collections (Sinclair, 1991) from multiple sources (Barnbrook, 1996), word frequencies are particularly useful.

This study used a plain-text corpus, as annotated corpora are less readily updated or expanded and therefore difficult to handle when automatically analyzing dynamic Web resources. Moreover, word-based and category-based approaches such as corpus annotation and tagging address different questions and often reinforce each other. “A reluctance, on theoretical grounds, to use categories that already exist in linguistics has led to a word-based practice of corpus investigation, which in turn has led to a revised theory of what language is like” (Hunston, 2002, p93).

A sample as culturally heterogeneous as the Fortune Global 500 necessitates identifying the language(s) used. Several techniques tackle this issue, usually based on trigrams and common short words (Hull & Grefenstette, 1996). Trigrams compare a document’s frequency of three-letter sequences with a particular language’s distribution of these same three-letter sequences. Similarly, common short words such as determiners, conjunctions and prepositions help divine a language. Both methods produce similar results for chunks of text larger than ten words (Grefenstette, 1995), so this research used the computationally lighter short-word technique to classify content within each Web site.

The 416 sites’ use of English dwarfed content in four other West European languages. English content was 11 times more prevalent than French, 16 times more common than German, and used 33 times more often than Spanish or Italian. After detecting the document languages, statistical tests analyzed differences by region and by industry among the corporations’ use of sports-related terms on their sites.

## Results

The initial analysis revealed difficulty interpreting some terms. The ambiguous words *golf* and *marathon*, for example, also showed up as a Volkswagen car model and an oil company. Terms such as *health*, *swimming*, *climbing* or *running* also had several meanings, popular in colloquial non-sporting phrases such as ‘climbing the corporate ladder’. Although the purpose of this study was to examine how corporations relate to sports on their Web sites, these ambiguities highlight the pervasive role of sports.

After eliminating ambiguous terms, the most popular term – appearing on 70% of sites – was *sport* or *sports*, hereinafter referred to as *sport(s)*. The term *olympic(s)* appeared on one third of the sites, but only 1% of the sites (five corporations) included the term *paralympics*. Selecting the most popular sport proved difficult due to ambiguous use of the terms *football* and *soccer* in different countries. *Football* was on 29% of the sites compared to *soccer* at 21%. *Baseball* (21%) and *basketball* (20%) closely followed in popularity.

Just 11 corporate sites included the term *sport(s) marketing* – most frequently mentioned by the Web site of the Massachusetts Mutual Life Insurance Company, followed

by Hyundai Motor, Hyundai, Anheuser Bush, McDonalds, Samsung Electronics, United Parcel Service, Pepsi Cola, News Corporation and Bank of America.

Of those industries with at least five corporations, the automotive sector led in using the terms *sport(s)*. The results of a one-way ANOVA test showed significant differences across industries in the use of the term *sport(s)*;  $F(27, 388) = 2.915, p < .0001$ , but no differences in the use of *olympic(s)*;  $F(27, 388) = 0.453, p = .992$ .

### *Regional Web Coverage of Sports*

To investigate sporting term usage rather than trading sports-related products, the next analysis ignored retail companies that sell sports products, such as Wal-Mart or Kmart. Similarly, Marriott featured recreational mountain biking and golfing, while Berkshire Hathaway promoted a work environment 'better than golfing or fishing'. The analysis also ignored companies with less than ten percent English content, and eight companies not based in Europe, North America or Asia.

Of the remaining 292 companies, over half (54%) were based in North America, 32% in Europe and 14% in Asia. One out of five (23%) companies was in finance and insurance, followed by companies in food, beverage, tobacco (10%) and resources (9%).

Due to difficulties in interpreting term occurrences and the low frequency of some expressions, the analysis grouped terms such as *soccer* with *football* and *rugby*, and *canoeing* with *kayaking* and *rowing*. To ensure comparability across Web sites, the analysis used term frequencies relative to the total number of words contained in the corpus. To increase the stability of results, the analysis ignored terms that appeared on less than ten Web sites. Figure 1 shows the regional distribution of sports terms, presented as a percentage of the global number of occurrences indicating a particular sport. North American companies made up over half the sample, and about one out of seven companies was Asian. Had there been no regional differences in sports, the results should have reflected this distribution. This was not the case.

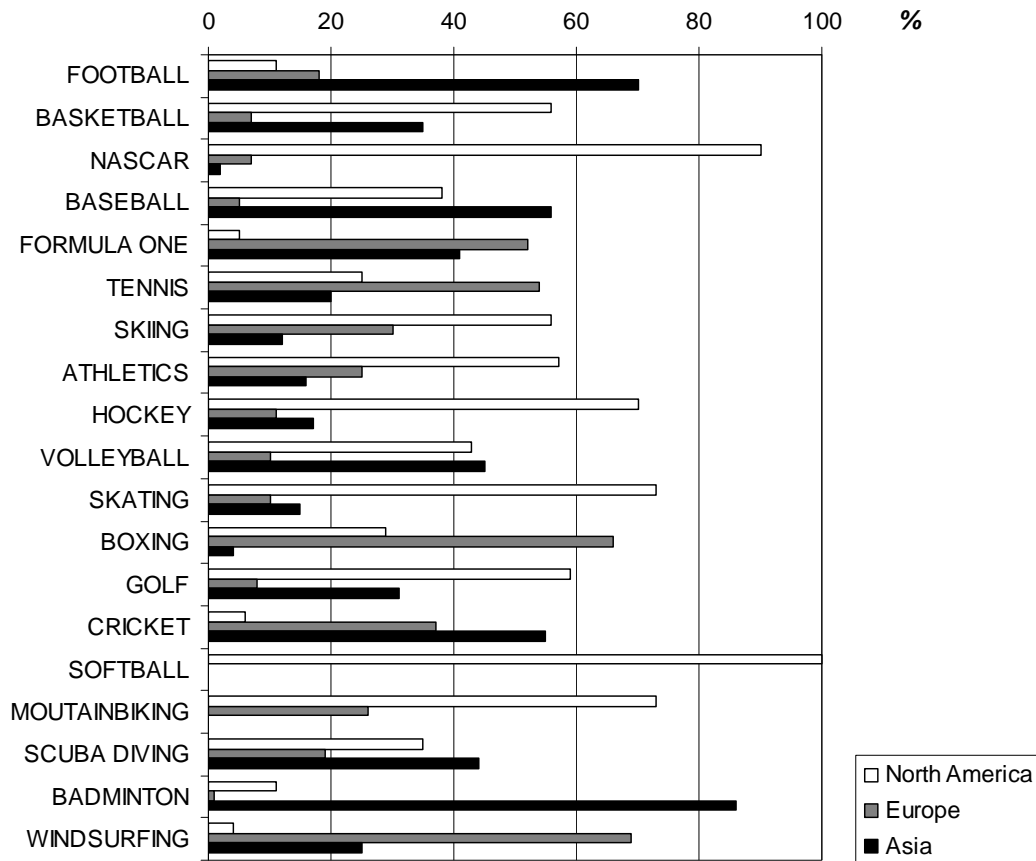


Figure 1. Regional Distribution of the Top 20 Sports

The terms *basketball*, *nascar*, *skiing*, *athletics*, *hockey*, *skating* (ice skating, figure skating), *golf* (pga, lpga), *softball*, and *mountain biking* showed a North American bias. *Formula one*, *tennis*, *boxing* and *windsurfing* had a stronger European presence. Finally, the Asian sites had a proportionally greater use of the terms *football* (rugby league, rugby union and soccer), *baseball*, *cricket*, *scuba diving* and *badminton*.

#### Web Coverage of Sports by Industry

Web content represents multidimensional data with so many variables that an orthogonal visual structure is often insufficient. Subdividing a problem or computationally reducing the dimensionality leads to results that are faster to interpret, convey more information, and cause fewer interpretive problems (Lebart et al., 1998). Correspondence analysis, for example, identifies lower-dimensional approximations of the original data and portrays general patterns without the need to specify a priori assumptions (Blasius, 1994; Van der Heijden, Mooijart, & Takane, 1994).

The perceptual map of Figure 2 reports the Web coverage of sports by industry, based on term frequencies relative to the total number of words contained in the corpus. It represents the rows and columns of a contingency table in a joint plot (Hair, Anderson, Tatham, & Black, 1998). The square markers represent the Top 10 sports; the round markers denote the position of a particular industry in the computationally created two-dimensional space.

The vertical axis suggests a distinction between individual sports (*formula one, nascar*) and team sports (*basketball, volleyball, baseball*). The horizontal axis is more ambiguous. Participating sports targeting white-collar interests such as *tennis* and *skiing* dominate the right part of the diagram and are most popular with the travel and tourism industry. Besides sponsorship, travel and tourism companies often include tennis and skiing in their offerings (e.g. holiday packages, hotel facilities, location close to well-known skiing regions), which helps explain the spatial proximity. As expected, *nascar* and *formula one* are the preferred sports of the automotive industry. Together with football, they tend to be spectator sports that often appeal to blue-collar interests.

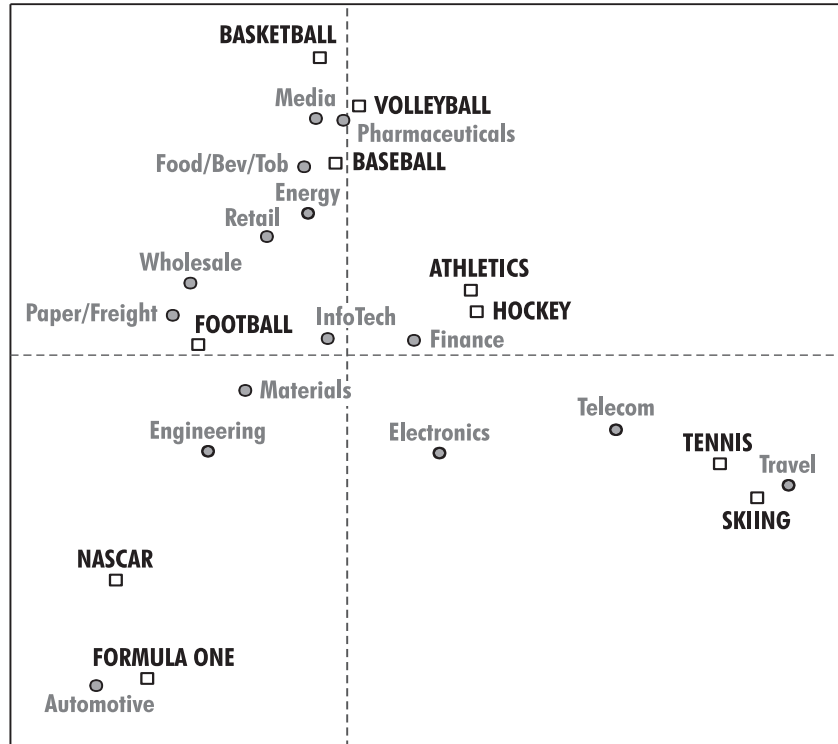


Figure 2. Perceptual Map Contrasting the Top 10 Sports with the Fortune Global 500 Industries

### Web Coverage of Renowned Athletes

Of the initial list of 131 sports stars, only 28 names appeared on the 416 Global 500 Web sites. Two of the biggest sports stars in recent times, *Michael Jordan* and *Tiger Woods* topped the list, each appearing on 11 sites. *Tiger Woods* earns 10% of his annual income from tournament play and 90% from corporate endorsements (Roberts, 2002). Athletic apparel maker Nike’s choice of basketballer *Michael Jordan* to lead their endorsement campaign in the mid 1980’s probably saved the company (Hatfield, 2003). Jordan’s sustained on-court success and public appeal ensures he remains one of the highest profile athletes in the world today, even after retirement (Anonymous, 2003).

Formula One driver *Michael Schumacher* was third with eight references, and Brazilian footballer *Ronaldo* was fourth with six references. Reflecting the regional appeal of cycling, fifth-placed American cyclist *Lance Armstrong* appeared three times on European sites and only twice on US sites. Three or more sites mentioned the following athletes:

*Marion Jones, Serena Williams, Kelly Clark, Jenson Button, Janica Kostelic and David Hall.*

On an aggregate level, the names of athletes in only 12 sports appeared. Seven of the 28 names on corporate Web sites were tennis players, five were golfers, four were track and field stars, three were auto racers, two played basketball, and two were wheelchair athletes. The remaining sports – cycling, soccer, snowboarding, skateboarding, downhill skiing and American football – appeared just once. Perhaps surprisingly, *Ronaldo* was the only representative of the self-labeled “World Game” soccer.

Regionally, the results showed a marked imbalance. Over a quarter of the North American and European corporate Web sites mentioned sports stars, compared to only 4% in the Asian region. This may reflect the individualism/collectivism dimension differences between Western cultures that tend to promote the achievements of the individual and Asian cultures where individuals subordinate their needs to those of the group (Hofstede, 1980). The percentages for Australian and Mid/South American sites mentioning sports stars were larger, albeit from a smaller base of corporations.

Table 2. Regional Distribution of Sites Mentioning Renowned Athletes

<b>Regions</b>	<b>Corporations</b>	<b>Occurrences</b>	<b>Percentage</b>
N. America	178	46	26%
Europe	127	33	26%
Asia	100	4	4%
Australia	5	2	40%
Mid/S. America	5	4	80%
<b>Totals</b>	<b>415</b>	<b>89</b>	<b>21%</b>

With the Soccer World Cup held in Japan and North Korea in 2002, and the surge in popularity of Chinese basketball star Yao Ming in North America, it would seem that Asian corporate Web sites reflect growing corporate sponsorship of sports. Little use of sports star names on these Web sites, however, contradicts this view. Despite a tradition of corporate-sponsored teams since the 1950s, the erosion of corporate sports sponsorship in Japan results from efforts to cut costs in light of the economic recession (Terukazu, 2001). In other Asian countries, it is often governments that push the sports agenda, not the private sector (Granitsas, 2002).

## **Conclusion and Future Research**

Sponsorship literature is often normative, suggesting best practices in sponsorship management (Olkkonen, 2001). This paper adds a positivistic dimension by investigating the textual content on 416 Web sites. In contrast to traditional analysis with human coders, this method considers the dynamic nature of the World Wide Web by collecting the data within hours, thereby improving the comparability of results.

The method complements but does not replace qualitative approaches. This study adds an important dimension to the analysis of Web coverage by gathering and analyzing more than 270,000 documents, a text corpus impossible to analyze manually – ignoring the methodological issues of inter- and intracoder reliability.

Seventy percent of Fortune Global 500 companies mentioned *sport(s)* on their sites. As just one in three sites mentioned the Olympics, some companies may sponsor the Olympics but fail to promote their sponsorship through their Web sites. These results indicate

rich sponsorship opportunities for the International Olympic Committee and suggest addressing large transnational organizations that do not yet sponsor the Olympics.

The findings also showed geographic and industry differences in the presence of sports-related terms. These results give sports marketers insights into current trends. Softball could be an untapped market outside of North America, for example, while windsurfing could be a North American opportunity. Alternatively, this lack of geographic representation could signal an unpopular sport. Marketers can add these Web-based results to their own information in order to improve the evaluation of opportunities.

While this research provides a snapshot of the online presence of sports and sports stars with the world's leading corporation, future research should add a theoretical perspective. Diffusion of Innovations (Rogers, 1995) is one possible avenue for exploring the adoption of information technology. Organizations often adopt innovations due to fashion and fad rather than for strategic purposes (Abrahamson & Rosenkopf, 1993), such as corporate Web sites that fail to reinforce sports marketing sponsorship.

Research has shown that organizational factors such as size, industry and management relate to the effective implementation of new technologies (Fichman, 2000; Fichman & Kemerer, 1999), including Web sites (Murphy, Olaru, Schegg, & Frey, 2003). Qualitative studies could analyze corporate strategy on the one hand, and marketing and sponsorship that correspond to this strategy on the other hand.

Future research would benefit from longitudinal studies of changes in Fortune Global 500 Web sites' use of sports-related terms. The analysis should coincide with events such as the Soccer World Cup or the Olympic Games to examine fluctuations during major global championships. For studies using specific US terms such as the Super Bowl, Southeastern Conference (SEC) or NCAA, the Fortune 1000 is a better database as it contains twice the number of corporations, and only those in the US.

Two limitations of this study are the 10 megabyte mirroring limit, and the reliance on term frequencies. While the limit ensures comparability, it also means ignoring content of larger sites. Statistics based on frequencies fail to consider the usage context of a particular term. Therefore, future research should (i) explore and compare alternative settings for the crawling agent and (ii) determine the semantic orientation of a site towards specific concepts.

One method to compute the semantic orientation is measuring the strength of a term's association with a set of positive words (taken from a tagged dictionary), minus the strength of its association with a set of negative words (Scharl et al., 2003), or multi-word units of meaning (Danielsson, 2004). Yet the lexis of Web content only partially determines its semantic orientation. Future extensions of the current system architecture should therefore employ shallow parsing (Hammerton, Osborne, Armstrong, & Daelemans, 2002) to process syntactic and semantic ambiguities, and capture meaning-making processes at levels beyond lexis.

Analyzing the popularity of corporate Web sites that include a large proportion of sporting terms with a positive semantic orientation may highlight the importance of sports sponsorship for successful online strategies. Does the use of sports-related terms influence a Web site? Future research should compare sports-rich Web sites and sites poor in the use of sports-related terms by studying attitudes towards corporate Web sites and consumer beliefs such as awareness, trustworthiness, aesthetic appeal, community interest and entertainment.

## Appendix – List of Sports-Related Terms

**Sports terms, single word:** aquatics, archery, athletics, badminton, baseball, basketball, biathlon, bicycling, bmx, bobsleigh, boxing, canoe, celebrity, climbing, crew, cricket, curling, cycling, diving, dressage, épée, equestrian, fencing, fitness, football, freediving, golf, gymnastics, handball, hockey, judo, karate, kayak, kitesurfing, lacrosse, luge, marathon, motorcycling, nascar, olympic, olympics, paralympics, rowing, running, sabre, sailing, scuba, shooting, skateboarding, skating, skiing, slalom, soccer, softball, spectators, sponsorship, sport, sports, stadium, surfing, swimming, taekwondo, tennis, triathlon, volleyball, walking, waterskiing, weightlifting, windsurfing, wrestling.

**Sports terms, multiple word:** australian rules football, beach volleyball, body boarding, collegiate football, cross country skiing, downhill skiing, field hockey, figure skating, formula 1, formula one, grand prix, grand slalom, horse racing, ice dancing, ice hockey, ice skating, indy car, modern pentathlon, motorcycle racing, mountain biking, professional football, rhythmic gymnastics, road racing, roller blading, roller skating, rugby league, rugby union, sail boarding, ski jumping, slalom skiing, snow boarding, sports marketing, street luge, super g, super-g, synchronised swimming, synchronized swimming, table tennis, track and field, track cycling, trick skiing, wake boarding, water polo, water skiing, whitbred round the world, winston cup, world cup.

**Sports stars:** hank aaron, andre agassi, troy aikman, muhammed ali, heidi andreasen, lance armstrong, aaron baddely, layne beachley, david beckham, albert belle, ole einar bjoerndalen, bonnie blair, terry bradshaw, kobe bryant, bob burnquist, jenson button, jennifer capriati, anne-caroline chausson, kelly clark, kim clijsters, derrick coleman, earle connor, shea cowart, tara dakides, john daly, lindsay davenport, inge de bruijn, stacy dragila, heike dreschler, john elway, ludmila engqvist, brett favre, juan carlos ferrero, george foreman, brian frasure, cathy freeman, will gad, pierre-luc gagnon, sergio garcia, steven gerrard, steffi graf, maurice greene, wayne gretzky, ken griffey, david hall, mia hamm, daniela hantuchova, tonya harding, tony hawk, justine henin, beatrice hess, grant hill, martina hinges, mat hoffman, evander holyfield, mike horn, tara ipinski, michael irvin, goran ivanisevic, magic johnson, marion jones, michael jordan, tanja kari, davo karnicar, nancy kerrigan, janica kostelic, janica kostelic, anna kournikova, michelle kwan, bernhard langer, vinny lauwers, brett lee, mario lemieux, lennox lewis, hermann maier, dan marino, michael milton, yao ming, joe montana, juan pablo montoya, david nalbrandian, joe namath, shaquille o'neal, shaun palmer, travis pastrana, lee pearson, umberto pelizzari, chantal petitclerc, scottie pippen, dean potter, paula radcliffe, steve redgrave, elena repko, jerry rice, cal ripken, andy roddick, dennis rodman, ronaldo, wayne rooney, pete rose, marat safin, pete sampras, louise sauvage, gerd schönfelder, paul schulte, michael schumacher, beat schwarzenbach, monica seles, o.j. simpson, emmitt smith, annika sorenstam, john stamstad, daryl strawberry, picabo street, martin strel, gabriela szabo, michael teuber, ian thorpe, dara torres, mike tyson, joachim uytdehaage, franziska van almsick, pieter van den hoogenband, esther vergeer, kurt warner, karrie webb, serena williams, venus williams, tiger woods, kristi yamaguchi, steve young.

## References

- Abrahamson, E., & Rosenkopf, L. (1993). Institutional and Competitive Bandwagons: Using Mathematical Modelling as a Tool to Explore Innovation Diffusion. *Academy of Management Review*, 18(3), 487-517.
- Abratt, R., Clayton, B., & Pitt, L. (1987). Corporate Objectives in Sports Sponsorship. *International Journal of Advertising*, 6, 351-362.
- Anonymous. (2003). At the Top of the Sports Marketing Game. *The Oregonian* (May 23, 2003), A01.
- Armstrong, C. (1988). Sports Sponsorship: A Case Study Approach to Measuring its Effectiveness. *European Research*, 16(2), 97-103.
- Baker, W. H., DeTienne, K., & Smart, K. L. (1998). How Fortune 500 Companies Are using Electronic Résumé Management Systems. *Business Communication Quarterly*, 61(3), 8-19.
- Barnbrook, G. (1996). *Language and Computers*. Edinburgh: Edinburgh University Press.

- Bauer, C., & Scharl, A. (2000). Quantitative Evaluation of Web Site Content and Structure. *Internet Research: Networking Applications and Policy*, 10(1), 31-43.
- Bennett, R. (1999). Sports Sponsorship, Spectator Recall and False Consensus. *European Journal of Marketing*, 33(3/4), 32-48.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics - Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Blasius, J. (1994). Correspondence Analysis in Social Science Research. In M. Greenacre & J. Blasius (Eds.), *Correspondence Analysis in the Social Sciences* (pp. 23-52). London: Academic Press.
- Clark, J. M., Cornwell, T. B., & Pruitt, S., W. (2002). Corporate Stadium Sponsorships, Signalling Theory, Agency Conflicts and Shareholder Wealth. *Journal of Advertising Research*, 42(6), 16-32.
- Danielsson, P. (2004). Automatic Extraction of Meaningful Units from Corpora. *International Journal of Corpus Linguistics*, 8(1), 109-127.
- Fichman, R. G. (2000). The Diffusion and Assimilation of Information Technology Innovations. Framing the Domains of IT Management. In R. W. Zmud (Ed.), *Projecting the Future Through the Past* (pp. 105-128). Cincinnati: Pinnaflex Publishing.
- Fichman, R. G., & Kemerer, C. F. (1999). The Illusory Diffusion of Innovation: An Examination of Assimilation Gaps. *Information Systems Research*, 10(3), 255-275.
- Gabba, V., Pan, Y., & Ungson, G. R. (2002). Timing of Entry in International Market: An Empirical Study of U.S. Fortune 500 Firms in China. *Journal of International Business Studies*, 33(1), 39-55.
- Granitsas, A. (2002). Nationalism In The Driving Seat: Government, Not Private Enterprise, Promotes Sport In Asia. *Far Eastern Economic Review* (April 11, 2002), <http://www.feer.com/>.
- Grefenstette, G. (1995). *Comparing Two Language Identification Schemes*. Paper presented at the 3rd International Conference on Statistical Analysis of Textual Data, Rome, Italy.
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate Data Analysis*. Upper Saddle River N.J.: Prentice Hall.
- Hammerton, J., Osborne, M., Armstrong, S., & Daelemans, W. (2002). Introduction to Special Issue on Machine Learning Approaches to Shallow Parsing. *Journal of Machine Learning Research*, 2(3), 551-558.
- Hatfield, S. (2003). What Makes Nike's Advertising Tick? *The Guardian* (June 17, 2003), p. 15.
- Hofstede, G. (1980). *Culture's Consequences: International Differences in Work-Related Values*. Beverly Hills: Sage.
- Hull, D. A., & Grefenstette, G. (1996). *Querying Across Languages: A Dictionary-based Approach to Multilingual Information Retrieval*. Paper presented at the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Zurich, Switzerland.
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Javalgi, R. G., Traylor, M. B., Gross, A. C., & Lampman, E. (1994). Awareness of Sponsorship and Corporate Image: an Empirical Investigation. *Journal of Advertising*, XXIII(4), 47-58.
- Krippendorff, K. (1980). *Content Analysis: An Introduction to Its Methodology*. Beverly Hills: Sage.

- Lawler III, E. E., Mohrman, S., & Ledford Jr., G. E. (1992). The Fortune 1000 and Total Quality. *The Journal for Quality and Participation*, 15(5), 6-10.
- Lebart, L., Salem, A., & Berry, L. (1998). *Exploring Textual Data* (Vol. 4). Dordrecht: Kluwer Academic Publishers.
- Leech, G., Rayson, P., & Wilson, A. (2001). *Word Frequencies in Written and Spoken English*. Harlow: Pearson.
- Leichty, G., & Esrock, S. (2001). Change and Response on the Corporate Web Site. *American Communication Journal*, 5(1).
- Lloyd, S. (2000). The Games Are Over: Let the Sponsorship Begin. *Business Review Weekly* (Nov 10, 2000), <http://www.brw.com.au/>.
- McEnery, T., & Wilson, A. (2001). *Corpus Linguistics* (2nd ed.). Edinburgh: Edinburgh University Press.
- McManis, B. L., Ryker, R., & Cox, K. C. (2001). An Examination of Web Usage in a Global Context. *Industrial Management and Data Systems*, 101(8/9), 470-478.
- McMillan, S. J. (2000). The Microscope and the Moving Target: The Challenge of Applying Content Analysis to the World Wide Web. *Journalism and Mass Communication Quarterly*, 77(1), 80-98.
- Meyer, C. F., Grabowski, R., Han, H.-Y., Mantzouranis, K., & Moses, S. (2003). The World Wide Web as Linguistic Corpus. In P. Leistyna & C. F. Meyer (Eds.), *Corpus Analysis - Language Structure and Language Use* (pp. 241-254). Amsterdam: Rodopi.
- Montenegro-Tores, F., Engelhardt, T., Thamer, M., & Anderson, G. (2001). Are Fortune 100 Companies Responsive to Chronically Ill Workers? *Health Affairs*, 20(4), 209-219.
- Morf, D. A., Schumacher, M. G., & Vitell, S. J. (1999). A Survey of Ethics Officers in Large Organizations. *Journal of Business Ethics*, 20(3), 265-271.
- Mullin, B. J., Hardy, S., & Sutton, W. A. (2000). *Sport Marketing 2nd ed.* Champaign, IL: Human Kinetics.
- Murphy, J., Olaru, D., Schegg, R., & Frey, S. (2003). The Bandwagon Effect: Swiss Hotels' Website and E-mail Management. *Cornell Hotel and Restaurant Administration Quarterly*, 44(1), 71-87.
- Olkkonen, R. (2001). The Network Approach to International Sport Sponsorship Arrangement. *Journal of Business and Industrial Marketing*, 16(4), 309-329.
- Palmer, J. W., & Griffith, D. A. (1998). An Emerging Model of Web Site Design for Marketing. *Communications of the ACM*, 41(3), 44-51.
- Papandropoulos, A. C. (2002). Sponsorship: The Financial Blood of the Olympic Games. *Trade with Greece*, 25, 40-45.
- Perry, M., & Bodkin, C. (2000). Content Analysis of Fortune 100 Company Web Sites. *Corporate Communications: An International Journal*, 5(2), 87-96.
- Pope, N. (1998). Overview of Current Sponsorship Thought. *Cyber-Journal of Sport Marketing*, 2(1), <http://www.ausport.gov.au/fulltext/1998/cjism/>.
- Pope, N., Forrest, E. J., & Murphy, J. (1996). *Sports Marketing on the Internet*. Paper presented at the Australia New Zealand Marketing Academy, Auckland, New Zealand.
- Reicher, A. K., Webb, M. S., & Thomas, E. G. (2000). Corporate Support for Ethical and Environmental Policies: A Financial Management Perspective. *Journal of Business Ethics*, 25(1), 53-64.
- Roberts, R. (Writer) (2002). Tiger Woods Endorsement Deals [Television]. New York: ABC News: World News This Morning (Oct 9, 2002).
- Rogers, E. M. (1995). *Diffusion of Innovations* (4th ed.). New York: The Free Press.

- Romano Jr., N. C. (2002-3). Customer Relationship Management for the Web-Access Challenged: Inaccessibility of Fortune 250 Business Web Sites. *International Journal of Electronic Commerce*, 7(2), 61-117.
- Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (2001). Methodological Issues in the Content Analysis of Computer Conference Transcripts. *International Journal of Artificial Intelligence in Education*, 12(1), 8-22.
- Sandler, D. M., & Shani, D. (1993). Sponsorship and the Olympic Games: the Consumer Perspective. *Sport Marketing Quarterly*, 2(3), 38-43.
- Scharl, A. (2000). *Evolutionary Web Development*. London: Springer.
- Scharl, A., & Bauer, C. (2004). Mining Large Samples of Web-Based Corpora. *Knowledge-Based Systems*, 17(5-6), 229-233.
- Scharl, A., Pollach, I., & Bauer, C. (2003). Determining the Semantic Orientation of Web-based Corpora. In J. Liu (Ed.), *Intelligent Data Engineering and Automated Learning, 4th International Conference, IDEAL-2003 (Lecture Notes in Computer Science, Vol 2016)* (pp. 840-849). Berlin: Springer.
- Shaheeh, N. (1999). *The Olympics Supermarket: Money is the Chief Player*. Retrieved May 27, 2003, from <http://www.islamonline.net/iol-english/dowalia/money-30-Sep-2000/money1.asp>
- Shank, M. D. (2002). *Sports Marketing: A Strategic Perspective*. Upper Saddle River, NJ: Prentice Hall.
- Sinclair, J. (1991). *Corpus Concordance Collocation*. Oxford: Oxford University Press.
- Sinclair, J. (2004). *Trust the Text - Language, Corpus and Discourse*. London: Routledge.
- Sleight, S. (1989). *Sponsorship*. Maidenhead: McGraw Hill.
- Stotlar, D. (1993). Sponsorship and the Olympic Winter Games. *Sport Marketing Quarterly*, 2(1), 35-43.
- Summers, J. (2003). Sport Marketing. In J. R. McColl-Kennedy (Ed.), *Services Marketing: A Managerial Approach*. Milton QLD: John Wiley.
- Teather, D. (2003). *Sponsor Passes £55m to School Basketball Star Before His First Game*. Retrieved May 27, 2003, from <http://www.guardian.co.uk/international/story/0,3604,962553,00.html>
- Terrian, J. (2002). *Career Paths in Sports Marketing and Corporate Sponsorships Today*, Marquette University, from <http://classwork.busadm.mu.edu/classwork/Terrian/SportFebruary16.ppt>
- Terukazu, S. (2001). Corporate Sports Sponsorship Falls Upon Hard Economic Times. *Japan Quarterly*, 48(1), 17-24.
- Thelwall, M. (2002). Methodologies for Crawler Based Web Surveys. *Internet Research: Electronic Networking Applications and Policy*, 12(2), 124-138.
- Van der Heijden, P. G. M., Mooijart, A., & Takane, Y. (1994). Correspondence Analysis and Contingency Table Models. In M. Greenacre & J. Blasius (Eds.), *Correspondence Analysis in the Social Sciences* (pp. 79-111). London: Academic Press.
- Weaver, G. R., Treviño, L. K., & Cochran, P. L. (1999). Corporate Ethics Practices in the Mid-1990's: An Empirical Study of the Fortune 1000. *Journal of Business Ethics*, 18(3), 283-294.