Identification of the Technology Commercialization Strategies of High-tech Small Firms

by

Diana M. Hicks, Dirk P. Libaers, Alan L. Porter, David J. Schoeneck Search Technology, Inc. and Georgia Institute of Technology

for

under contract number SBAHQ05M0292

Release Date: December 2006

SBA Office of Advocacy
www.sba.gov/advo

Small Business Research Summary

Advocacy: the voice of small business in government

December 2006 No. 289

Identification of the Technology Commercialization Strategies of High-tech Small Firms

Diana Hicks, Dirk P. Libaers, Alan L. Porter, David J. Schoeneck,
Search Technology, Inc. Atlanta, GA ; 2006, [19 pages] Under contract SBAHQ-05-M-0292

Purpose

Small, highly innovative companies have been at the forefront in creating the industries of the future, but finding highly innovative companies is often difficult. This is because their innovations are sometimes found in their business strategies rather than the products they produce. The goal of this study was to capture the business models of innovative firms instead of classifying them by the goods or services they produce. A unique method was used to find these companies: searches of innovative terms on their websites.

Overall Findings

Searching for innovative terms on firms' websites determined that high-tech strategies could be found on known high-tech small firms but not on the control group. This indicates that this method holds promise for gaining insight into finding highly innovative firms.

Highlights

• "Serial innovators" (small firms with 15 or more patents in a five-year period) do not fit easily into business industry classifications.

• Small firm websites tend to be geared more to describing the firm, or creating an image, than seeking customers. This transparency about the firm enables researchers to evaluate their websites to determine the firms' business strategies

• High technology small firms tend to have common strategies, while their less innovative competitors tend to have varied strategies.

• Factor analysis showed that search terms could be created to categorize firms into the following groups:

- science-based product/service firm,
- R&D organization or contractor,
- product solutions provider,
- highly specialized component/module/ consumable or raw materials suppler, and
- specialized subcontractor.

• The web searching method presented in the paper holds promise for capturing more than just business strategies, and in some circumstances could be a cost-effective way to replace surveys and reduce respondent burden.

• Further research is needed to find out if this search method can determine categories for random individual firms.

Scope and Methodology

The study used six classes of firm technology strategies to evaluate which category firms match.

From 80 small firm websites, 67 terms were identified as being strategy indicator terms. Using Internet search programs, the terms were run on 407 serial innovators and a control group. The serial innovators, also referred to as high-tech small firms, were found by using a previous study identifying small firms highly involved in patenting. Hoover's Company Capsules was used to find competitors to those firms for the control group.

Of the 67 search terms, 43 keywords were useful, and the number of pages mentioning the terms was captured and normalized by size of website. Factor analysis was used to detect patterns of the keywords on the websites for both groups.

This report was peer reviewed consistent with the Office of Advocacy's data quality guidelines. More information on this process can be obtained by contacting the director of economic research at ***advocacy@sba.gov*** or (202) 205-6533.

Ordering Information

The full text of this report and summaries of other studies performed under contract with the U.S. Small Business Administration's Office of Advocacy are available on the Internet at ***www.sba.gov/advo/research***. Copies are available for purchase from:

National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
(800) 553-6847 or (703) 605-6000
TDD: (703) 487-4639
www.ntis.gov

Order number: PB2007-100690
Pricing Information:
Paper copy A03 ($26.50)
Microfiche A01 ($14.00)
CD-ROM A00 ($20.00)
Electronic Download A00 ($9.95)

For email delivery of Advocacy's newsletter, press, regulatory news, and research, visit ***http://web.sba.gov/list***. For Really Simple Syndication (RSS) feeds, visit ***www.sba.gov/advo/rsslibrary.html***.

Executive Summary

Highly innovative young firms are rare among the population of small firms, but such "serial innovators"[1] appear increasingly important in an economy undergoing structural change. The entrepreneurial moment is often studied, as is the relationship that small firms in general have with technology, but both types of study lack the specificity needed to understand the behavior and needs of this high value, elite group of innovative survivors. Our intent in this study was to focus attention on the serial innovators and take an important first step in understanding them by developing a taxonomy that outlines in broad terms key differences between firms. Unfortunately, existing classification schemes such as NAICS (North American Industry Classification System) or existing taxonomies of innovative firms are built from large firm data and emphasize continuity with the past, and so ill suit highly innovative small firms that more often than not are developing new business models and defining the industries of the future.

In this project, we took the first steps towards a classification of the technology commercialization strategies of the 407 most innovative U.S. small firms, with successful innovators identified as those firms with a substantial portfolio of granted U.S. patents. In developing our method we placed a priority on cost-effective execution and on not placing a burden on the subject firms. Therefore, we explore the potential for gathering information relevant to firm strategy from firm websites. We argue that firm websites universally include aspirational statements of firm identity, and our method looks for systematic differences between these statements, using the differences to classify firm technology commercialization strategies.

In undertaking this research, web addresses were located and verified; then a theoretically informed reading of a sample of 80 websites was undertaken to identify keywords used by firms that related to known technology commercialization strategies; 89 potentially informative keywords related to business models were identified. We built a special "Innovative Firms Application Programming Interface" (IFAPI) to use Google's search capabilities. Using IFAPI, we obtained hit counts for each term on each firm website and normalized by size of website. Refining over the course of several iterations resulted in 407 innovative firms for which 43 keywords proved analytically useful. Factor analysis detected a pattern in use of the keywords, delineating six factors which we interpret as reflecting six high technology commercialization strategies used by small firms (Tables 2&3).

The six business models used by small firms are:

1. R&D organization or contractor
2. Science- based product/service firms
3. Highly specialized component supplier (high volume production)
4. Specialized subcontractor firm (one-offs or very low volume production)
5. Product solutions provider
6. Service solutions provider (technical consultants)

The same analysis was run on a group of control firms, identified in Hoovers as small firms who compete with the highly innovative small firms. This factor analysis failed to converge on a solution, that is no underlying factors could be identified that explained the pattern of keyword use across the control

[1] D. Hicks and D. Hegde, Highly Innovative Small Firms in the Markets for Technology, *Research Policy*, July 2005.

firms. We interpret this as the absence among control firms of a general pattern of use of the six technology commercialization strategies. This establishes that the highly innovative small firms pursue distinct business models to commercialize their technology.

In this project, we have established that the general approach holds promise for going beyond historical classifications to cost-effectively gain insight into large populations of innovative new players in a quickly evolving economic landscape. We used these techniques to propose a taxonomy of strategies used by innovative small firms. Further research is needed to classify individual firms into the categories and to use the scheme to understand the special needs of this elite group of small firms.

1. Introduction

Previous work for the Office of Advocacy established the existence of a growing cadre of highly innovative small firms (Hicks, 2002). These firms were identified in a comprehensive database of all U.S. firms issued 15 or more patents in the most recent 5-year period. The 15-patent threshold is quite high for a firm with fewer than 500 employees and is reached only by firms that have survived four or five years at least. We called such firms "serial innovators".[2]

In general, small firms have difficulty innovating; and even when they do innovate, they can find the patent system expensive to use. According to the NFIB Research Foundation, National Small Business Poll on innovation:

- 88% of small firms do not purposefully innovate;
- 16% of small firms avoid technology whenever possible,
- and only 5% of small firms own a patent used in the business.

These figures confirm that innovative, patenting small firms are highly unusual, and that the "serial innovators" are a very rare cadre of elite firms. Although uncommon, we can understand the existence of these firms within scholarly frameworks proposing an emerging division of labor in innovation (Arora, Fosfuri and Gambardella, 2001) and vertical disintegration or modularization within the semiconductor industry (Ernst, 2005). These frameworks point to an evolving industrial structure—one in which a greater variety of business models are viable, providing more opportunity for innovative small firms than was typical in the post-WWII environment dominated by highly vertically integrated large firms in traditional manufacturing industries. It is the need to understand this evolving structure and the opportunities it offers small firm innovators that motivate this study.

The fact that industrial structure is evolving and that the serial innovators are in the vanguard of this trend is starkly revealed when one attempts to classify these firms. Classification is a mundane activity, yet it proves extremely frustrating with these firms because even the most basic of traditional firm classification schemes fail to apply to these firms, let alone to provide insight. For example, noted Harvard University professor Michael Porter's classification of basic firm strategy options into cost leader and product differentiation is far too coarse grained, as all these firms use costly innovation to differentiate their offerings. "Manufacturing" and "service" designations proved impossible to implement, as these firms were found to implement varying and shifting combinations of these most basic firm operations. The NAICS classification of industrial sectors is useless, as many of these firms are found under "other" categories in this scheme, or cluster in areas not yet recognized with a designation in NAICS. The late University of Sussex professor Keith Pavitt's taxonomy of firms by sources of innovation proved to be the most useful. Still, this scheme was developed for large firms and then later adapted to small firms on the basis of very little detailed information about innovative small firms; this leaves us with just two

[2] In contrast with the generally recognized "serial entrepreneur" who after a firm they founded is bought or fails goes on to found another firm.

categories: new technology-based firms (in biotech, semiconductors etc.) and specialized suppliers (of machine tools, components, instruments and software) (Tidd et al., 1997). This scheme tells us little more than whether the firm operates in a new or an old industrial sector. To understand the serial innovators we need to probe more deeply. We believe that developing a taxonomy of these firms is the place to start in understanding their role in the evolving economy, and so we focus in this study on classifying the serial innovators.

However, to classify small firms, we need detailed information about them. This is more difficult to obtain than it is for large firms. To overcome the limits of existing data, in our previous studies we built an innovative patent database to identify the serial innovators. This study pursues this type of methodological innovation again. In particular, we seek to gather detailed information on small firms from publicly available sources in order to avoid imposing on already overburdened small firm entrepreneurs. In addition, we seek a method that overcomes the limits of alternative means of gathering firm data, namely small sample size, and focus on one sector (interview based data) or prohibitive cost and low data quality (surveys with low response rates). In this study, we develop and test a method for large-scale website content analysis. Websites are publicly available information so our analysis places no burden on subject firms; therefore, a very high percentage of target firms will be analyzed (much higher than is typical within the limits of response rates). As we access the websites algorithmically, costs do not scale with the number of firms examined (as they do with interview based studies), so large numbers of geographically dispersed firms can be cheaply analyzed.

In the rest of this report, we describe our taxonomy of serial innovators and the large-scale website content analysis method. We begin by briefly reviewing the literature on content analysis of firm websites, arguing that the text on firm web sites expresses firm strategy in an informal sense. We then describe our methodology: the selection of small firms to study, keyword identification and construction of firm-by-keyword matrices recording the frequency of occurrence of each word on each firm's website. To analyze this data we ran a factor analysis. The analysis identified six strong factors underlying the pattern of keyword use across the websites. We interpret this as a six-category classification of serial innovator strategy.

2. The meaning in firm websites

The advent of universally accessible, ubiquitous organizational websites has provided scholars with an obvious opportunity to apply methods of content analysis to a new communication medium, and researchers have taken advantage. To judge by literature reviews in McMillan and in Opoku, a preponderance of website content analyses is conducted by scholars of communications or marketing who seek to assess the functionality of corporate websites. Almost without exception, corporate websites are found wanting. Marketing and communication scholars are deeply disappointed that firm websites fall so short on metrics of communication and interaction. For example, Perry & Bodkin (2000) note that companies are aggressive in communicating with many audiences. Unfortunately, this means that the information about a firm has no specific intended audience, and as the nature of a text should differ by its intended audience, this compromises communication effectiveness. Pollach concludes that: "Instead of helping readers to learn more about the company in an interactive way, they tend to offer rather static pages engaging readers in one-way communication only" (Pollach, 2005, p. 296.). Ellinger et al. (2003) find a disappointing absence of many features on websites such as interactivity and decent navigation, especially for small firms. Perry & Bodkin conclude that: "in aggregate, firms are only scratching the surface in terms of marketing communication", exhibiting "a naïve appreciation for the information requirements of customers."

The studies conclude that the websites fail at communication; that is, they fail effectively to impart information or ideas in an exchange of information. Furthermore, these studies locate this failure in the design of websites. That is, the analysts point to many well-accepted design principles absent from firm websites, as opposed to assessing the actual impact of websites through customer focus groups

(which might find material was designed using sound principles and yet still failed in empirical testing of audience impact). Since the failure is one of design, one is left to wonder whether the failure is intentional, meaning that the firm designing the website meant it to serve a purpose other than communication in the above sense. We suggest that the elements of corporate websites, especially the "about us" sections, that fail to meet criteria of good design for communication, are instead designed to articulate. By articulate we mean: expressing in coherent verbal form, giving words to, or fitting together into a coherent whole. Judged against this standard, the websites might not prove so disappointing.

If we accept this thought however, we are left with the question -- what are the websites meant to articulate? Corporate identity is the most likely answer. The British Standards Institution in 1995 defined corporate identity as what an organization is, what it stands for, what it does and how it goes about its business. Ellinger et al. (2003) compared websites of large and small firms in the transportation industry. There were many differences and the small firm sites lacked many features; nevertheless, the mission statement/about section of a website was almost universally present. Perry & Bodkin's examination of corporate websites led them to conclude that the sites focused on institutional advertising. Sullivan (1999) asked the question: What are the functions of corporate home pages? He sought to answer it in a cross-national comparison of corporate websites, and found that website function varied somewhat with local circumstances; for example U.S. websites included more legal information than German or Japanese sites due to the more litigious environment in this country. Nevertheless, he concluded that "image creation is the most important function" of corporate websites.

Why is this? Sullivan directly addresses this point by comparing the explanatory power of several theories in relation to website content (Sullivan, 1999). He concludes that the universal image creation function of corporate websites supports McLuhan's global village model:

> *McLuhan's Global Village model stressed a developed industrial world almost overwhelmed by information flows fostered by improvements in electronics and communication technology. In this perspective, corporations, as they compete with each other and with millions of individuals to deliver messages to stakeholders, find their task almost impossible. They soon learn to focus primarily on what McLuhan called "pattern recognition" rather than on specific content that will not be attended to. According to the model, corporate home pages always will focus mostly on image creation, and the use of the Web for transactions, disclosure, and information transfers will remain secondary goals. (Sullivan, 1999, p. 208)*

Thus to Sullivan, although the design of firm websites may not be ideal for purposes of substantive communication, the design is realistic in accepting that the individual firm must adapt to customers who are unwilling to invest cognitive effort in assessing product and other corporate messages (Sullivan, 1999, p. 209). Topalian expands this idea, contrasting our current world with previous stable times when relationships could be built up over time, face to face. Then firms were embedded in local cultures, and their founders and works were known in their communities. He points out that today we face unstable times, 'when reality has no time to "settle".' and competitive pressures constantly shift (Topalian, 2003). Glaser argues that such fast pacing is inherent in an information intensive age (Glaser, 1991). Websites are perhaps the quintessential emblem of this information intensity. Firms desire to be known by people who could be important to the business but at best have fleeting contact not with the firm, but with the firm's information. Glaser notes that in an information age, power shifts to buyers. To Sullivan it is "a realistic acceptance of the difficulty of doing anything else" combined with the urge to be known in this environment that expresses itself as firms universally articulating a corporate identity.[3]

[3] Anecdotal support for this proposition is provided by the curious fact that many small firm websites display a picture of their (usually unimpressive) building on their website. Topalian notes that a generation ago,

In articulating an identity, firms apparently take full advantage of the rhetorical potential and don't limit themselves to verifiable facts but venture into the aspirational arena. Pollach (2005) analyzed "about" website sections (not product pages or transaction pages) of 20 most admired multinational firms. She chose the firms because they have achieved wide admiration, yet surprisingly Pollach still found the firms to be making outsize claims. Firms made claims aiming to position themselves among the "great and good." For example, their use of verb tense was detached from past or present to create a sense that they exist in perpetuity. They made outrageous, unsupported claims of size and scope. They used passive tense used to present claims of positive characteristics in a way that implied universal recognition. They positioned themselves as agents of positive global change and good corporate citizens to create an impression that they were saving the world. Claims of leadership were universal. Pollach harshly judges the companies on the basis that such text lacks credibility and so viewers will reject the message. On the other hand, Pollach's findings clearly establish the aspirational nature of the web text.

We argue then that while firm websites may vary in the extent to which they actually achieve the myriad communication and e-commerce goals possible, corporate websites universally articulate an aspirational corporate identity. In this study, we seek to classify firms based on this corporate identity.

We have found one other study that seeks to exploit the potential of websites to assess identity - Opoku's study of brand "personality" of US business schools (Opoku, 2005). Opoku developed a dictionary of terms associated with each of the five dimensions of brand personality proposed in Aaker's theoretical framework (sincerity, excitement, competence, sophistication and ruggedness). The frequency of the words on the websites of 30 US business schools was then obtained and the pattern of occurrence of the words was assessed to classify the business schools by their brand "personality" according to the Aaker framework. Our study is similar in style to Opoku's.

3. Method

In undertaking this research, web addresses were located and verified; then a theoretically informed reading of a sample of 80 small firm websites was undertaken to identify keywords used by firms that related to known technology commercialization strategies. 89 potentially informative keywords related to business models were identified. We built a special "Innovative Firms Application Programming Interface" (IFAPI) to use Google's search capabilities. Using IFAPI, we obtained hit counts for each keyword on each firm's website and normalized by size of the website. Refining over the course of several iterations resulted in 407 innovative firms for which 43 keywords proved analytically useful. Factor analysis detected a pattern in use of the keywords, delineating six factors that we interpret as six high technology commercialization strategies used by small firms (Tables 2&3).

Keyword development

We began with theoretically informed immersion in the text of a sample of the websites of 80 highly innovative small firms to develop a set of keywords that are used by firms and which also relate to strategies analyzed in the innovation literature. We took care to develop a set of keywords that not only reflects the activities of the firms, but also the literature on strategy and innovation. We began with Michael Porter's concept of a generic value chain and conceived of a simplified version (Porter, 1980):

R&D ⇨Manufacturing ⇨Distribution ⇨Sales ⇨After Sales

The manual examination of the 80 firm websites sought to identify for each firm which activities in the value chain were internalized within the small firm's boundaries and which ones were outsourced.

---

when reality was more stable, firms' "facilities were visible realities." The picture of the building perhaps harkens back to that simpler time, when customers knew the firms and had physical experience of them.

Terms chosen were those that seemed to be broadly used among firms pursuing a particular strategy, but not shared among strategies. The pages examined on firm websites were:

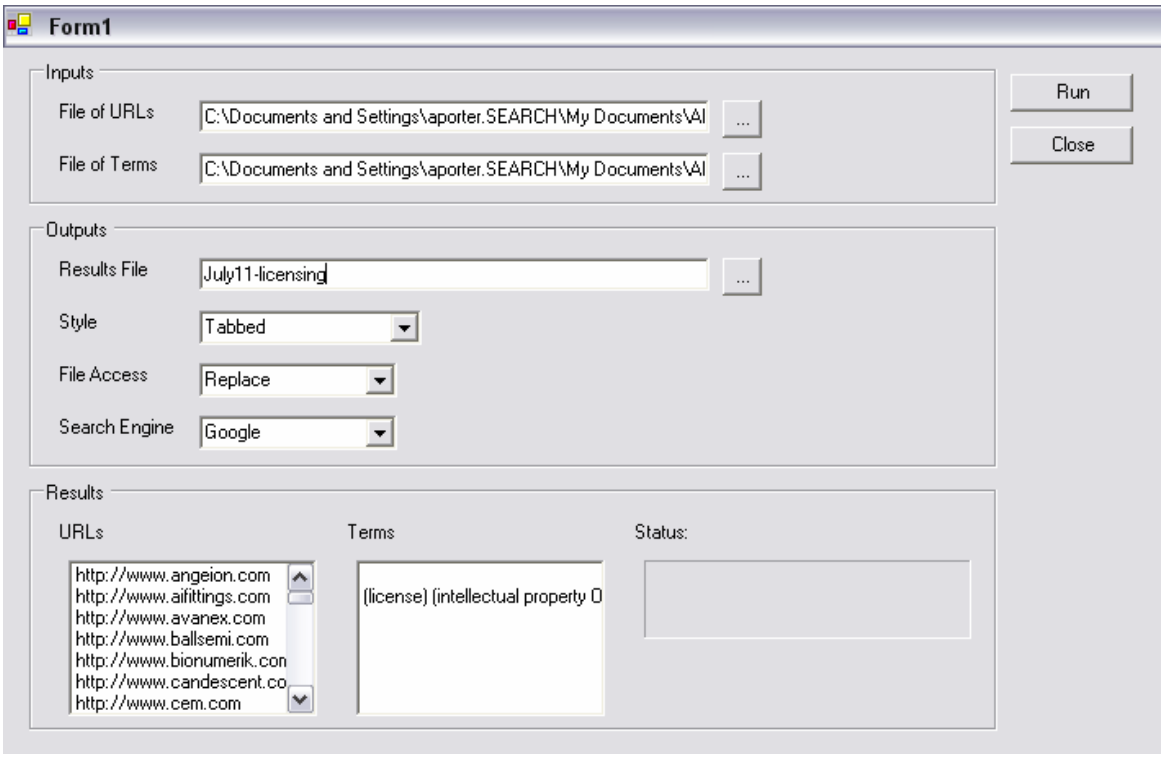
1. Introductory page
2. Page that describes product/service portfolio offerings
3. Page that explicitly highlights Research & Development (R&D) activities performed by the firm. This section is very common on corporate websites active in biotechnology, pharmaceutical, medical devices, semiconductor, and R&D contracting.
4. Page explicitly describing the intellectual property assets held by the firm e.g. patents, trade secrets, and perhaps copyrights. Also scientific publications for biotech, pharmaceutical, and medical device firms.
5. Page that describes a firm's historical background and milestones.
6. Page that describes employment opportunities, e.g. R&D scientists, Product Development Engineers, system integration specialist, etc.
7. Page that describes the management team and board of directors and that might include keywords such as inventor, spinout, professor, etc.

A long list of 89 keywords was produced in the manual phase.

Website search

To obtain the frequency of keyword occurrence on the websites of some 400 highly innovative small firms, we used the functionality of Google. Specifically, we searched for each of the keywords on each of the 400 websites and recorded the number of hits returned as well as the total number of pages on the website. This process was facilitated by an algorithm written by us expressly for the purpose. The algorithm was developed using Google's SOAP (Simple Object Access Protocol) API (Application Programming Interface) which allows users to write scripts to conduct automated searches. We developed the software, referred to hereafter as the "Inventive Firms API" ("IFAPI"), in C++. The development process involved determining how to enable searches on complex Boolean strings and multiple URLs, retrieving separate counts. The IFAPI Window is shown as Figure 1. It is simple to use, enabling several of us to "parallel process" the search operations. Results are returned as a tab-delimited file (either replacing or appending if the file already exists). IFAPI also provides an additional results output that shows which of the URL searches returned "exact" and which "approximate" results (at Google's discretion).

Figure 1. The Inventive Firms API Window



To obtain a measure of frequency independent of the size of each firm's website, we required the total number of pages on each website. By entering a blank line in our search terms file, IFAPI search returns an approximate (sometimes exact) count of the total webpages. For instance, in a sample of 30 firms, the total counts for smaller sites show single-digit precision (e.g., 508 pages or 179 for), but rounded for larger sites (e.g., 17800 rather than 17856). For our focus on small firms and intended normalization uses, this seems fine.

Google provides documentation on preparing APIs and on the search method because algorithm-driven searches are not identical to regular Google searching. Building the best possible search strategy involved learning the details of Google's API-driven search methodology. For instance, Boolean OR must be capitalized; "AND" is assumed; "*" does not wildcard because Google API searching includes built-in stemming. By running alternative term variations, we decided that this was quite usable for our purposes.

Table 1 shows search results for a small sample that affirms use of search terms such as "research." Google searching returns approximate counts. This means that our results have a margin of error. Repeat runs on the same searches may yield somewhat different results. In most cases, the results are very similar; however, there are cases with very different results returned and we are unable to explain those.

Table 1. Sample Google API Results

Firm	research	researcher	researchers	"researchers"	researching	(research OR researcher OR researchers OR researching)
A	262	2	7	7	3	264
B	37		8	8		37
C	8					9
D	38		3	3		46
E	33		1	1	1	35
F	5					5
G	462	9	11	11	6	465

Using the IFAPI, the analysis was conducted as follows:

- We began with the 89 keywords.
- Prior to testing, URLs were needed. These were collected in manual Google searches for each of the small firms in the prior CHI study that still existed as independent, non-bankrupt firms in the summer of 2006. We identified URLs for 420 firms.
- During July and early August 2006 the IFAPI was used to obtain the frequency of occurrence of all keywords on all websites.
    - Terms were dropped from the analysis if they proved to be very infrequently used or if they were so broadly used that they did not distinguish business models effectively—"patent" and "intellectual property" were dropped for the latter reason.
    - Firms were dropped from the analysis if their websites were less than 10 pages in size because such small sites could produce outlier word frequency results. Our statistical technique, factor analysis, is highly affected by outliers.
    - The analysis finalized on a firm by word matrix of 407 firms by 67 terms, 43 of the terms proved analytically useful in delineating six of the proposed business models.
- The raw firm by term matrix was transformed into a matrix of firm by word frequency by dividing the number of hits on a term by the number of pages on a website. As Google returns the number of occurrences of a word, not the number of pages on which it occurs, the resulting "term intensity" measure can be greater than 100%.
- Factor analysis in SPSS was used to look for patterns in word usage across firms.

Factor analysis

Factor analysis is an analytical method often used in psychology. According to Darlington:

> *[Factor analysis] is used to study the patterns of relationship among many dependent variables, with the goal of discovering something about the nature of the independent variables that affect them, even though those independent variables were not measured directly* (Darlington, 2006)

In our application, we are studying the patterns of relationship among many terms used on serial innovator websites. Our goal is to discover something about the firm's business models (independent

variables not observed directly), and we argue that the business models are reflected in the way firms describe their business on the website, and so in the terms they use.

Statistical techniques are usually used to generalize from a sample to a population. We do not do that here. Rather, we have a population of firms; specifically, serial innovators defined as all firms with fewer than 500 employees and with 15 or more U.S. patents in the five years preceding 2002. These are highly unusual firms and we do not believe that our conclusions will generalize to other firms. However, we do argue that it is vitally important to understand this set of firms in its own right. Towards this end, we use factor analysis to help identify business models that innovative small firms seem to use.

The factor analysis technique is a "complex procedure with few absolute guidelines and many options" (Costello and Osborne, 2005). In making our choices, we followed many of the recommendations of Costello & Osborne. We chose factor analysis over the default principal components analysis because factor analysis better suits our purpose of revealing "any latent variables that cause the manifest variables to covary" (Costello & Osborne, p. 2). Because our data violate conditions of normality, we used "alpha extraction." Among the choices available for extraction using non-normal variables, alpha extraction considers the variables in the analysis (terms, in our case) to be a sample from the universe of potential variables. This is a sensible approach to take to our data. Using the default, we retained factors with eigenvalues greater than 1.0, though Costello & Osborne recommend retaining fewer and experimenting with this to gain a best solution. We also used the default varimax rotation that produces uncorrelated factors. Our subject (firm) to item (word) ratio was 6. A high ratio (and so large sample size) is needed for this technique to work. Like most researchers, we did not reach the recommended ratio of 10. Nevertheless, our data passed the default Kaiser-Meyer-Olkin Measure of Sampling Adequacy, with a value of 0.826 (0.60 is the minimum acceptable value). Table 2 displays the solution obtained with the above set of choices. The table highlights factor communalities greater than 0.40 as recommended. Crossloading items (those with 0.32 communalities on more than one factor) are retained. The solution is good because there are six factors with more than three items and most have five or more items loading strongly, i.e. 0.50 or better. These are the markers of a "solid factor" (Costello & Osborne, p. 5).

To establish that the result we obtained is not purely a result of a statistical technique seeking a factor solution and always producing one, we constructed a sample of control firms matched to the serial innovators by industry but not among the highly innovative firms with 15 or more patents. The control sample was constructed as follows:

1. For each highly innovative small firm, we consulted the Hoover's Company & Capsules Database, which provides brief information on 40,000 public and non-public companies (Capsules) and 225,000 key executives. The Profiles include an overview and history of company operations as well as key officers, competitors, number of employees and selected historical financial data (seven years).
2. For each innovative firm, Hoover's yielded a list of direct competitors, along with their HQ location and work force size.
3. From the list of competitors, we selected one firm incorporated and headquartered in the US, and having a workforce of 500 employees or less.
4. When the list of competitors did not yield a firm that met these two criteria, a list of competitors from a competing firm was checked; i.e. a competitor of a competitor, often termed "indirect competitor". The same procedure was followed, and care was taken that the firm that met the two criteria outlined above and was active in the same sector as our focal firm.
5. The procedure highlighted in points 2, 3 and 4 was repeated for each of the innovative firms, resulting in a control sample of small firms.

The same analysis was run on the control sample; that is, the same terms were searched on the same days using the same algorithm. The resulting control firm-term matrix was submitted to the factor analysis described above, and passed the same test of sampling adequacy. However, on the control firms, the factor analysis failed to converge on a solution as some items "retained communalities greater than 1". This suggests that the serial innovators exhibit a pattern of word use not found among their competitors. We interpret this as meaning that the high tech small firms pursue some common strategies that their less innovative competitors do not pursue.

4. Results and discussion of the factor analysis output

The solution obtained in the factor analysis of highly innovative firm website term usage is shown in Table 2. As mentioned above, it is a good solution as judged by the solidity of the factors identified; and it is not obvious that a good solution would have been found, as no solution was found for the control sample. We believe that the factors can be plausibly identified with six business models. In this section, we explore the relationship between the factors and a proposed taxonomy for small, highly innovative firms. For each factor, we list the associated keywords, describe the underlying strategy the words reflect, and offer a sample firm's self description taken from a website. Finally we offer a summary table that elaborates on each category in the proposed taxonomy.

Table 2 – Factor Analysis results*

Interpretation	Search term	1	2	3	4	5	6
Factor 1 science-based product/service firm	research	0.79					
	(R&D OR research and development)	0.76	0.30				
	Research and Development AND (Company OR Corporation OR Co)	0.75	0.31				
	novel	0.74					
	(collaboration OR collaborating)	0.71			0.35		
	clinical trial	0.70					
	discovery	0.69					
	pipeline	0.63					
	professor	0.60					
	commercialize AND research	0.59					
	research and development AND organization	0.58					
	university OR universities	0.57					
	expertise	0.56					
	proprietary	0.47					
	science	0.33					
Factor 2 R&D organization or contractor	R&D AND Testing		0.93				
	Contract AND Research		0.81				
	Technology Development		0.72				
	Research AND Testing	0.35	0.55				
	machines OR machining		0.45				
	project		0.40				
Factor 3 product solutions provider	(total AND solution)			0.70			
	(system AND integration) OR (system AND solution)			0.65			
	Solution			0.57	0.49		
	product AND performance			0.55			
	cost AND effective			0.54			
	capability			0.51			
Factor 4 service solutions provider	enabling				0.64		
	consulting		0.59		0.63		
	(module OR modular)				0.62		
	integration			0.32	0.61		
	provider				0.56		
	government				0.42		
	packaging				0.41		
Factor 5 highly specialized component/ module/consumable or raw materials supplier	maintenance					0.66	
	(contract OR subcontract)		0.49			0.61	
	OEM OR (original AND equipment)					0.59	
	improve					0.54	
	unmatched					0.41	
	accessories					0.41	
	turnkey					0.38	
	supplier					0.31	
Factor 6 specialized subcontractor	we						0.82
	integrated			0.35			0.65
	custom					0.32	0.64
	efficiency					0.38	0.58
	component			0.34			0.54

*Rotated factor matrix; extraction method: alpha factoring; rotation method: varimax with Kaiser normalization; rotation converged in 11 iterations. 34.3% of total variance explained by the six factors displayed; correlations over 0.30 displayed, over 0.40 highlighted, factors with greater than three items displayed.

Factor 1 – This factor has a factor loading of 7.39 and accounts for 11.02 % of the total variation in the data. We propose that this factor represents a **science-based product/service** strategy. The following keywords are correlated with this factor (correlation coefficient in parentheses):

1. Research (0.79)
    a. R&D OR Research and Development (0.76)
        i. Research and Development AND company (0.75)
        ii. Research and Development AND organization (0.58)
2. Clinical trial (0.70)
3. Novel (0.74)
4. Discovery (0.69)
5. Professor (0.60)
6. Collaboration OR Collaborating (0.71)
7. Pipeline (0.63)
8. University OR Universities (0.57)
9. Expertise (0.56)
10. Commercialize AND Research (0.59)
11. FDA (0.48)
12. Proprietary (0.47)

In the following section, the keywords in the factor analysis are typed in bold.

Science-based product/service firms often use the term **research and development company or corporation** in their 'About Us' section of their corporate website. The term **research and development** and its derivative forms **R&D** and **research** are widely used, since these firms are very research-intensive and spend a large portion of their revenue (provided the firm is revenue-generating) or funds raised from third parties on **R&D**. These firms can be biotechnology, biopharmaceutical, pharmaceutical, nanotechnology or agrobiotechnology firms. Science-based product/service firms are often engaged in drug or gene **discovery**, searching for **novel** compounds that can be tested against known disease targets or bugs that may harm agricultural crops.

These drugs or other active compounds must be tested in **clinical trials** and approved by the Food and Drug Administration (**FDA**). Biotechnology or pharmaceutical firms in this category often use the term drug **pipeline** to indicate how many compounds are in development or clinical trial, an issue closely watched by the financial community that use it as a proxy to value the firm. These science-based product/service firms have rich intellectual property bases and develop **proprietary** technology or products, backed by patents. The industries in which these firms operate are characterized by strong **collaboration** patterns between small firms and large, incumbent firms with very high degrees of vertical integration. Just like the large, established corporations in their industry, the small science-based product/service firm have strong linkages with the science base i.e. **universities**. Science-based product/service firms may commercialize their proprietary technology or **expertise** by means of licensing arrangements with large firms in their industry, without having to build manufacturing capacity, distribution channels or a sales force. Finally, a substantial number of firms in this category have been founded by a **professor** or have elected professors to their Board of Director committees, another indication of strong ties to the academic world.

Here is an excerpt from the website of a science based product/service firm:

The firm is a clinical-stage pharmaceutical company that is developing a rich pipeline of synthetic peptide-based drugs against clinically validated targets for the treatment of kidney diseases and cancer. Our mission is to accelerate the drug discovery and development process through innovative

technologies based on novel applications of chemistry and biology. This will expedite advancement to clinical development and commercialization. In August 2001, a collaboration of several venture firms created the firm as an independent company—a spin-out from a large pharmaceutical firm. The Company has leveraged its exceptional foundation to become a powerful developer of peptide drugs.

Factor 2 - This factor has a factor loading of 3.70 and accounts for 5.52 % of the total variation in the data. We propose that this factor represents an **R&D organization or contractor** strategy. The following keywords are correlated with this factor (correlation coefficient in parentheses):

1. R&D AND Testing (0.93)
2. Contract AND Research (0.81)
3. Technology Development (0.72)
4. Contract OR Subcontract (0.49)
5. Consulting (0.59)
6. Research AND testing (0.55)
7. Machines OR machining (0.45)

R&D Organization or Contractors are typically small firms that employ a small, very highly qualified workforce that specializes in a limited number of scientific and technical disciplines. This type of firm is fully engaged in **technology development**, in essence prototype development from scratch that is usually patented. Furthermore, these R&D contractors provide **consulting** services on very technical subjects and areas. The core of their business model is the performance of **contract research** or **subcontract** research. R&D contractors often have unique and specialized **machining** capabilities that small and even large production firms often lack, enabling them to offer unique rapid prototyping services.

Here is an excerpt from an R&D contractor's website:

> *Founded in 1982 the firm consisted of a relatively small group of individuals and lots of innovative ideas. Today, almost 200 engineers, technicians, and machinists work in our electronics and software engineering labs, machine shop, and on CAD stations. Our facilities have been designed to promote constant interaction between and within the engineering groups. Our on-site machine shop and molding facility are central to the success of our projects; ideas are prototyped and tested in record time. . . . Our mission, first and foremost, is to foster innovation. It is a company where the questioning of conventional thinking is encouraged and practiced by everyone—engineers and non-engineers alike—because open minds are more likely to arrive at workable solutions. This has been our formula for success since we began, and it will continue to drive our success in the future.*

Factor 3 - This factor has a factor loading of 3.28 and accounts for 4.89 % of the total variation in the data. We propose that this factor represents a **product solutions provider** strategy. The following keywords are correlated with this factor (correlation coefficient in parentheses):

1. Capability (0.51)
2. Product AND Performance (0.55)
3. Solution (0.57)
    a. (System AND Integration) OR (System AND Solution) (0.65)
    b. Total AND Solution (0.70)
4. Cost AND Effective (0. 54)

A product solutions provider is typically a small firm that identifies and addresses business problems by providing an innovative and **cost-effective total solution** to the problem. The problems identified are naturally complex (because they would have already been solved if they were simple), so the product offered by the product solution provider is systemic and architectural in nature, which means

that it consists of several subsystems or modules that can be integrated into the core product (the skeleton, no-frills standard product). The product solutions provider therefore needs to possess strong internal **system integration** skills and **capabilities**. Solutions built from a modular product architecture are flexible and can be adapted to specific customer needs. Because of the complexity of the product offering, the production solutions provider needs to have superior customer service capabilities such as technical support, repair, and maintenance to ensure sustained **product performance**. These products are typically made in relatively high volumes and so can compete on **cost effectiveness**.

Here is an excerpt from the website of a product solutions provider:

> *The firm combines microfluidics, liquid handling and laboratory automation to deliver unique research tools for today's Drug Discovery and Development and Genomics and Proteomics laboratories. Our products address critical applications within the life sciences industry; each of these applications is on the critical path to improving human health.*

Factor 4 - This factor has a factor loading of 3.06 and accounts for 4.58 % of the total variation in the data. We propose that this factor represents a **service solutions provider** (Technical consulting) strategy. The following keywords are correlated with this factor (correlation coefficient in parentheses):

1. Integration (0.61)
2. Solution (0.49)
3. Enabling (0.64)
4. Consulting (0.63)
5. Module OR Modular (0.62)
6. Government (0.42)
7. Provider (0.56)

Service solutions providers or technical consultants are small firms with superior technical competencies that tailor a service to the needs and requirements of large customers. They often provide technical **consulting** services and training (e.g. change management) that entail business **integration** processes or **integration** of tailor-made software modules in a larger system that serves as an **enabler** to conduct operations over an electronic platform (chip design services, avionics software in airline or jet fighter cockpits, etc.). This often requires close collaboration with the customer and with other contractors in order to provide a 'total **solution'** to the customer. The service is provided around **modular** products. SAP for instance is built of an accounting/finance module, an operations module, a logistics module, etc. that need to be configured and tailored to the customer's needs. These service solutions **providers** often view the **government** as a major market for business process reengineering and informatization/automatization. Well-known service solutions providers are the major consulting companies.

Here is an excerpt from the website of a service solutions provider:

> *The firm was founded as a circuit IP, design and fabless manufacturing company specializing in the non-volatile memory and embedded segments. Our mission is to provide superior technology, simpler process and reduced manufacturing costs for the NVM market with our IP portfolio and high-quality design services. We have has filed over 60 U.S. patents (39 granted, 22 pending) that showcase our emerging technology and breakthroughs in the embedded NVM IP market.*
>
> *We serve as a one-stop NVM shop of design, testing, production and wafer foundry support and turnkey solutions for our customers. We can secure your IP protection and can lead the design and manufacturing process every step of the way - from working with you to define your embedded spec to performing design and layout. We also provide testing and production flow, and can connect you to our multiple foundry sources for cost savings as well as manufacturing*

capacity and flexibility. What does this mean for our customers? A lower total cost of ownership, and timely developed products that are smaller, faster and more powerful.

Factor 5 - This factor has a factor loading of 2.97 and accounts for 4.43% of the total variation in the data sample. We propose that this factor is a proxy variable for a **highly specialized component/ module/consumable or raw materials supplier** category in the serial innovator taxonomy.

The following keywords are correlated with this factor (correlation coefficient in parentheses):

1. Assembly (0.41)
2. Accessories(0.41)
3. Contract OR subcontract (0.61)
4. Improve (0.54)
5. OEM (0.59)
6. Unmatched (0.41)
7. Maintenance (0.66)

Like the other firm types discussed above, this sort of highly specialized component/module/consumable or raw materials supplier operates across industrial markets, and market themselves to large Original Equipment Manufacturers (**OEM**) for whom they serve as a **contractor** or **subcontractor**. These firms emphasize **unmatched** manufacturing skills where process innovation is of crucial importance in order to **improve** production processes, and provide cost-effective products that perform in a superior fashion. These firms are often co-located at the manufacturing sites of the OEM (or vicinity) and provide technical services such as engineering, design for manufacture, and customer service like repair and **maintenance**. For these highly specialized component/module/consumable or raw materials suppliers, it is paramount to fully exploit economies of scale and therefore have very cost-competitive products of superior quality and performance that can be supplied to large Original Equipment Manufacturers or their contract manufacturers. These firms also provide **accessories** and often **assemble** modules for OEM customers.

Here is an excerpt from the website of a highly specialized component supplier:

We are the leading supplier of high-voltage analog integrated circuits (ICs) used in power conversion. Power Integrations ICs enable compact, lightweight power supplies that are simpler to design and manufacture, more reliable, and more energy-efficient than those made with competing technologies. Our ICs power a vast range of consumer and industrial electronics - computers, DVD players, TV set-top boxes, cell phone chargers, home appliances, telecom networking equipment, and many others.

Factor 6 - This factor has a factor loading of 2.60 and accounts for 3.88% of the total variation in the data. We propose that this factor represents a **specialized subcontractor** strategy. The following keywords are correlated with this factor (correlation coefficient in parentheses):

1. Integrated (0.65)
2. Custom (0.64)
3. Efficiency (0.58)
4. Component (0.54)

Specialized subcontractor firms typically develop a range of very specialized technical skills and competences and bid for highly complex projects in which their (often proprietary) products need to be **integrated** into a much larger product system. They **customize** their product offering and integrate it into the larger system. These firms provide tangible **customized** products (they can be **components**) as well as the integration capabilities to fit their offering into the larger, complex product systems. In a sense, this firm is a hybrid product/service firm with extremely specialized and rare competences. External integration capabilities and skills are of utmost importance to ensure the smooth and **efficient** operation of the larger,

complex product (an aircraft carrier, air traffic control system, intelligent buildings, etc.). The three keywords that describe such a firm are competences, **customization** and **integration** because these terms succinctly describe the business model for these type of firms. The specialized subcontractor operates exclusively in industrial markets and for public organizations because the complexity, the organization, and the product cost are of such a magnitude that is beyond the reach and needs of regular consumers.

Here is an excerpt from the website of a specialized subcontractor firm:

> *The firm produces information-based process improvement solutions for non-contact measurement and inspection applications. Automotive and manufacturing companies throughout the world rely on our process management solutions to help improve quality, shorten product launch times and reduce overall manufacturing costs. . . Our products are traditionally installed during the production cycle, but our scanning systems can be integrated with a number of mechanisms including CMM machines, articulated arms, robots and fixed mounts… giving our customers the flexibility to install our systems at any point in the manufacturing process.. . . By consistently utilizing the same methodology, we remove measurement variables, and ultimately provides more reliable results.*

Note that a weaker factor, number 8, appears to be related to firms that develop and market physical health products such as diagnostic equipment, healthcare products and medical devices.

The interpretation of the factors described above is summarized in Table 3 which lists six (plus one) types of firms, that is six types of technology commercialization strategy used by highly innovative small firms.[4] The factor number is given in the first column, followed by the name of the category, a description of the type of firm and the typical competitive advantages offered by each strategy.

[4] A seventh category, consumer goods, was noted in manual examination of the sample of firms but was not associated with common term use and so was not identified in the factor analysis.

Table 3 – Taxonomy of innovative small firms

Factor	Type of firm	Description	Potential sources of competitive advantage
2	R&D organization or contractor	Typically, a very small firm with a highly qualified workforce (PhD level scientists, MDs, and/or Master degree holders) that has developed a deep understanding of scientific phenomena in a limited number of research areas in which they conduct basic/applied research with a commercial orientation. Outputs are working prototypes, patents, or novel production processes and tacit know-how.	• explicit and tacit knowledge base developed over time and often inherited from previous employers (govt. labs, universities, large corporate R&D labs) • special formal or informal ties with previous employers/clients/institutions • highly specialized areas for which there is often only monopolistic (e.g. DoD, DoE) or oligopolistic demand • strong, reputable R&D team
1	Science- based product/service firms	Typically, a small firm that develops applications based on findings made in basic research laboratories. Often spinouts from the science base or from large established firms. Output typically are FDA-approved drugs, medical devices or diagnostics, novel electronic devices.	• patented library of molecules with therapeutic potential that was identified in a university lab/govt. lab or in-house • patented medical device initially developed in a university lab/by a medical practitioner • linkages to the science base • strong R&D team and alliance/partnership skills
5	Highly specialized component supplier (high volume production)	Typically, a small firm that develops discrete patented products that are often used in larger systemic products of medium to high complexity (e.g. car, TV set, scientific instruments etc). Output is typically discrete components manufactured in volume.	• internal R&D, co-design with OEM • efficient manufacturing processes • low-cost but skilled labor force • reputation • flexibility to follow large customers (OEM, ODM) overseas • long-term contracting relationships
6	Specialized subcontractor firm (one-offs or very low volume production)	Typically, small firms that excel in very specialized technologies for which applications need to be customized and integrated in often highly complex integrated products (aircraft carriers, mass transport system, intelligent buildings, manufacturing systems, etc.)	• customization and external integration capabilities • flexibility • track record, reputation • bidding and positioning skills, customer relationship capabilities • product boasts many unique features
3	Product solutions provider	Typically, a small firm that identifies a market need that needs to be addressed holistically, and provides a turnkey solution for that problem.	• strong internal system integration skills • high degree of vertical integration • strong customer service skills
4	Service solutions provider (technical consultants)	Typically, a firm that tailors a service to the needs of large customers (chip design services for wafer foundries, software development that addresses a particular business need).	• close relationship with client • strong customer service skills • strong external system integration skills • strong engineering and implementation skills, customer relationship capabilities • long-term supply and maintenance contracts
NA	*Consumer goods supplier (not captured in factor analysis)*	*Typically, a small firm that develops and often manufactures niche consumer products, e.g. barbeque sets, sports goods, coffeemakers, etc.*	• *targets niche markets with high-quality products* • *strong brand reputation* • *flexible, efficient manufacturing system*

5. Conclusions

In this exploratory study we focus on classifying the business models of serial innovator firms, that is, highly innovative small firms. We built a taxonomy of six (plus one) technology commercialization strategies used by high tech small firms. We identified the categories in a factor analysis based on the frequency of occurrence of a set of keywords related to value chain strategies on firm websites. We searched for 67 keywords on the websites of 407 highly innovative firms. Factor analysis detected patterns in keyword use, and a good factor analysis solution was obtained. Six strong factors emerged in the analysis based on 43 of the 67 keywords, and we interpreted these factors as reflecting six categories in a taxonomy. The same analysis run on a matched control group of firms failed to converge on a solution, suggesting that the strategies identified are unique to innovative small firms.

We derived a six-category taxonomy that encompasses the strategies these firms use for commercializing their innovative technology:

1. R&D Organization or Contractor – *Factor 2*
2. Science-based product/service firms – *Factor 1*
3. Highly specialized component supplier (high volume production) – *Factor 5*
4. Specialized subcontractor firm (one-offs or very low volume production) – *Factor 6*
5. Product solutions provider – *Factor 3*
6. Service Solutions provider (technical consultants) – *Factor 4*

We consider this a successful outcome to the study. We believe our newly developed web content mining methodology can be extended and applied to many issues. For instance, we have devised a way to return Google "snippets" describing the context of the term search; analyses of such data could yield better terms and phrases to get at underlying firm technology and marketing strategies.

Given the limited scope of this project, our results are preliminary. However, we have established that the general approach holds promise for going beyond historical classifications to cost effectively gain insight into large populations of innovative new players in a quickly evolving economic landscape. Further empirical verification of our taxonomy of highly innovative small firms is needed, along with consideration of the relationship between our taxonomy and other scholars' classifications of innovation and small firms. The matter of classifying each firm remains open.[5] Such studies promise a greater understanding of small, highly innovative firms and their technology commercialization strategies, thus providing insights useful for policy formation in relation to the special needs of this very unusual group of small highly innovative firms

[5] A standard method, k-means classification was tried. A robust classification of science-based product/service firms (largely biotech and pharmaceutical companies) and product solutions providers (electronics firms) was found in one variation, but this did not carry through to the final solution presented here.

References

Arora A., A. Fosfuri and A. Gambardella. 2001. *Markets for Technology*. Cambridge, MA: MIT Press.

Costello, A.B. & J.W. Osborne, 2005, Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis, *Practical Assessment, Research & Evaluation*, vol. 10, no. 7.

Darlington, R.B., accessed August 8, 2006, *Factor Analysis*, http://www.psych.cornell.edu/Darlington/factor.htm

Ellinger A.E., Lynch D.F., Hansen J.D., 2003, Firm size, web site content, and financial performance in the transportation industry, *Industrial Marketing Management*, 32, 3, pp. 177-185.

Ernst, D., 2005, Limits to Modularity – Reflections on Recent Developments in Chip Design, Industry and Innovation, 12, 3.

Glazer, R. 1991, Marketing in an Information-Intensive Environment: Strategic Implications of Knowledge as an Asset, *Journal of Marketing*, pp. 1-19.

Hicks, D. 2002. *Serial Innovators: The Small Firm Contribution to Technical Change*. Report prepared for the U.S. Small Business Administration, Office of Advocacy under contract number SBAHQ-01-C-0149.

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, pp. 80-98.

NFIB Research Foundation, 2005, *National Small Business Poll – Innovation*, Volume 5, Issue 6, ISSN – 1534-8326.

Opoku, R. A. 2005, Communication of Brand Personality by Some Top Business Schools Online, Licentiate Thesis, Lulea University of Technology, ISSN: 1402-1757.

Perry, M. & C. Bodkin, 2000, Content analysis of 100 company Web sites, *Corporate Communications: An International Journal*, 5, 2, pp. 87 – 97.

Pollach, I. 2005, Corporate self-presentation on the WWW: Strategies for enhancing usability, credibility and utility, *Corporate Communications: An International Journal*, 10, 4, pp. 285-301.

Porter, M. 1980, *Competitive Strategy: Techniques for analyzing industries and competitors*, The Free Press.

Sullivan, J. 1999, What are the Functions of Corporate Home Pages?, *Journal of World Business*, 34, 2, pp. 193-210.

Tidd, J. J. Bessant & K. Pavitt, 1997, *Managing Innovation: Integrating Technological, Market and Organizational Change*, (Chichester, UK: Wiley).

Topalian, A. 2003, The development of corporate identity in the digital era, *European Journal of Marketing*, 37, 7/8, pp. 1119-1132.