

Apr 24th, 12:00 AM

HUMINT or WEBINT? Concept study on possible routes for improving knowledge discovery within organizations

Nader Naghshineh

Information Studies Laboratory [ISL], Faculty of Psychology and Education, University of Tehran

Nader Naghshineh, "HUMINT or WEBINT? Concept study on possible routes for improving knowledge discovery within organizations." *Proceedings of the IATUL Conferences*. Paper 26.
<https://docs.lib.purdue.edu/iatul/2008/papers/26>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries.
Please contact epubs@purdue.edu for additional information.

HUMINT OR WEBINT? Concept Study on possible routes for improving knowledge discovery within organizations

By: Nader Naghshineh,
Information Studies Laboratory [ISL],
Faculty of Psychology and Education,
University of Tehran
nnaghsh@ut.ac.ir

Abstract:

The history of science is replete with instances of constant sociological and economic challenges posed by technical breakthroughs. On the other hand one could equally argue that it had been the sociological factors fomenting such breakthroughs. Both interpretations lead to the same effect, the increased complexity of the information sphere confronting us everyday. A number of researchers such as Shannon have argued an entropic effect for information. Some psychologist have suggested that our history from the standpoint of information dexterity could be interpreted by the degree of interaction of individual intelligence with collective intelligence as represented by institutions –such as universities- or constructs – such as web. Social Network Tools could be seen as just another challenge in dealing with complexity of our information sphere. The primary question is whether there is a criticality of information mass that leads to discovery and whether it is still a process driven by individuals or communities? Human Intelligence (HUMINT) approach would mean designing modalities for information worker to serve as an agent for such discovery. The Web Intelligence (WEBINT) approach would mean designing modalities for communities in which a complex system is constructed to ensure an expected, and therefore, quantifiable level of discovery. Borrowing heavily from the concept of Multiple Intelligences, the paper indicates that perhaps new pedagogical models need to be defined to address the issue as well as mixed approach to information handling to facilitate such discovery within organizations, especially the technical universities.

Keywords: Web Intelligence, Multiple Intelligences, HUMINT, Multi-Agent Systems, Pedagogical Systems, Knowledge Discovery, Post Modernist Knowledge Structures

We live in a moment of history where change is so speeded up that we begin to see the present only when it is already disappearing.

R.D.Laing

ENVIRONMENT SCAN

On Christmas day, 1990, Tim Berners-Lee switched on the world's first website server. Even though he had invented it, the event wasn't particularly important for him, mainly because his wife was about to have their first baby. But Berners-Lee had a pretty insightful idea it had the potential to be sweeping. For one thing, he had considered forming a start-up company and patenting the technology used to create what would become the greatest software innovation of the last decades of the 20th century. But he decided against doing so. Berners-Lee wanted to make the world a better-off place, not line his own pockets (Data Conversion Laboratory, 2002). On that day, as some argue, cyberspace was brought from the realm of fiction into reality (Berdayes & Murphy, 2000). It gave rise to incredible concepts such as Digital library, Internet intelligence, web-mind, Netscape as well as avatar-centric applications (Goertzel, 2002). It also brought about the dark forebodings of a technopoly voiced by many scholars and ordinary people alike (Postman, 1993). Yet the growth rate of internet usage in the middle-east is almost four times higher than that of Europe (see table 1. At the same time, while there are 550 billion documents on the web, mostly on the invisible or deep web (Bergman, 2001). The size of publicly indexable web, however, is about 11.5 billion pages (Gulli & Signorini, 2005). The ideology, or software of cultures, was being supplanted by that of a technology that could chip away against the traditional power bases of the governments. As many as 13 countries, including china reacted by instituting content monitoring/denial protocols restricting their citizen's access to the internet (Espiner, 2006)). On the other hand, through its de facto control of Domain Names System, the US Government may seek to deny internet to its enemies (Coulburn, 2007)

WORLD INTERNET USAGE AND POPULATION STATISTICS						
World Regions	Population (2007 Est.)	Population % of World	Internet Usage, Latest Data	% Population (Penetration)	Usage % of World	Usage Growth 2000-2007
<u>Africa</u>	933,448,292	14.2 %	43,995,700	4.7 %	3.5 %	874.6 %
<u>Asia</u>	3,712,527,624	56.5 %	459,476,825	12.4 %	36.9 %	302.0 %
<u>Europe</u>	809,624,686	12.3 %	337,878,613	41.7 %	27.2%	221.5 %
<u>Middle East</u>	193,452,727	2.9 %	33,510,500	17.3 %	2.7 %	920.2 %
<u>North America</u>	334,538,018	5.1 %	234,788,864	70.2 %	18.9%	117.2 %
<u>Latin America/Caribbean</u>	556,606,627	8.5 %	115,759,709	20.8 %	9.3 %	540.7 %
<u>Oceania / Australia</u>	34,468,443	0.5 %	19,039,390	55.2 %	1.5 %	149.9 %
WORLD TOTAL	6,574,666,417	100.0 %	1,244,449,601	18.9 %	100.0 %	244.7 %

NOTES: (1) Internet Usage and World Population Statistics are for September 30, 2007. (2) Demographic (Population) numbers are based on data contained in the [world-gazetteer](#) website. (3) Internet usage information comes from data published by [Nielsen//NetRatings](#), by the [International Telecommunications Union](#), by local NICs, and other reliable sources. (4) For definitions, disclaimers, and navigation help, see the [Site Surfing Guide](#). (www.internetworldstats.com. Copyright © 2007, Miniwatts Marketing Group. All rights reserved worldwide.

Table 1 World Internet Usage and Population Statistics

While by the end of the twentieth century it was common wisdom to see knowledge to be transformed substantially by technological power shifts, it seems not to be so. Does this mean that the scientific revolution is not so much a result of economic and technological advances as it was the culmination of the religious and ideological upheavals of the sixteenth and early seventeenth century? Does this imply that instead of a scientific revolution we are faced with a series of knowledge revolutions characterized by a new chain of command or hierarchy for authorization of knowledge? (Raschke, 2003). And in case we inevitably subscribe to such concept, how do we benefit from it given the degree of specialization in various sciences that an organization relies on for its operation?

An unfortunate consequence of specialization in the sciences is poor communication across research domains – which can hamper the knowledge discovery process. Research findings in one area may be pertinent to another, researchers may be unaware of relevant work by others that could be integrated into their work and important findings just outside a researcher's focus can be overlooked. Compounding this problem is the difficulty of keeping current with new research findings that continue to grow at an exponential rate. (Fuller, Revere, Bugni, & Martin, 2004). In essence there might be pockets or islands of relevant knowledge within the public domain that remain undiscovered due to the sheer volume of information one has to sift through.

Technology is an absolutely wonderful thing. It has unquestionably made our lives better in many, many ways. But it is not the answer to every problem, and blind faith in technology can create as many problems as it solves. Porter Goss, the new head of the CIA, has made it known that he wants less of a focus on technological wizardry, and more on "**humint**" (human intelligence). Unfortunately, there are a lot of people out there that think that every problem can be solved by a new piece of hardware or software (McGovern, 2004).

Repercussions

Technical Universities, the Libraries that serve them as well as their clientele (comprising of the faculty, students, government and other stake holders) essentially exist within a commonality of social networks with various degrees of formality. These networks could be as informal as student blogs to special community of practitioners that form a kind of knowledge web. Thus the role of libraries as organizational repository/depository of knowledge becomes somewhat blurred due to the fact that it has to address different perceptions governing knowledge queries. If such library is to be considered as merely a technical intelligence unit, then library servicescape should be mapped at four levels (Abushar & Hirata, 2002). These are as follows:

Competitive Intelligence Gathering
Industry and Competitor Analysis
Competitive Intelligence for Strategic Decision Making
Competitive Intelligence as a Core Capability

Stage one is primarily a library function. The primary skill here is oriented towards the ability to *find* the information. Companies such as *Fuld & company* and *FIND/SVP* are primary examples of such orientation. The traditional issues of information supply notwithstanding, this would mean that the libraries would need to seriously repackage the content of their information assets. This may require forming the patrons into study circles, staging of search webinars, incorporation of automated indexing software and even adopting a program similar to About.com. Even create a blog of issues and topics of interest for various interests within the faculty (Borgman & Furner, 2002).

Stage two essentially deals is more infoactive since it would require the library to tap into faculty knowledgebase. The patterns of search, projects in progress, document purchasing patterns, peer statements could be used to involve the faculty. The faculty would be serving as your node points for various information analysis that you could market to the local businesses or enterprises. In fact, the library would shift the force of disintermediation toward a synergy one often hopes for within virtual communities of professionals. There is a myth that small businesses cannot benefit from CI because of the cost component being prohibitive. As an information node for the local technical/SME community, the technical library could dispell this

myth by catering to this market. In the case of Sharif University of technology, the local businesses in fact support their “adopted libraries”. The libraries in turn not only serve as a recruiting ground for these businesses but also would serve as their information analysis point.. Their reports could be simple limited access blog, SDI or even executive summary of the trends of interest. Information visualization tools could be incorporated for this purpose.

As an individual organization, the library cannot evade the explosive rise in data. It could lead to a great deal of frustration (Eltemasi & Naghshineh, ~2008). It would also provide a great opportunity. It has been demonstrated that using various mining techniques that could be automated, it is possible to reveal the invisible connections, trends and patterns that exist in data. Various techniques such as Data mining, Web-usage mining, tech mining and bibliomining can transform massive volumes of data into invaluable knowledge (Naghshineh, 2007). In fact, when coupled with visualization systems to translate complexities, such techniques could provide unique tools for individual information handling capability. In fact the usual library exercises in bibliometric analysis such as counting of scientific publications or patents, the number of times those articles are cited in subsequent articles and co-citations or co-words used from articles to articles are a kind of measurement and signature intelligence (MASINT). One should take this allegory one step further and consider a librarian or information professional as a Human Intelligence Agent or HUMINT.

Essentially the traditional role of a technical library becomes increasingly supplanted by the requirement for actionable knowledge, or intelligence. Figure 2 demonstrates a typical intelligence cycle.

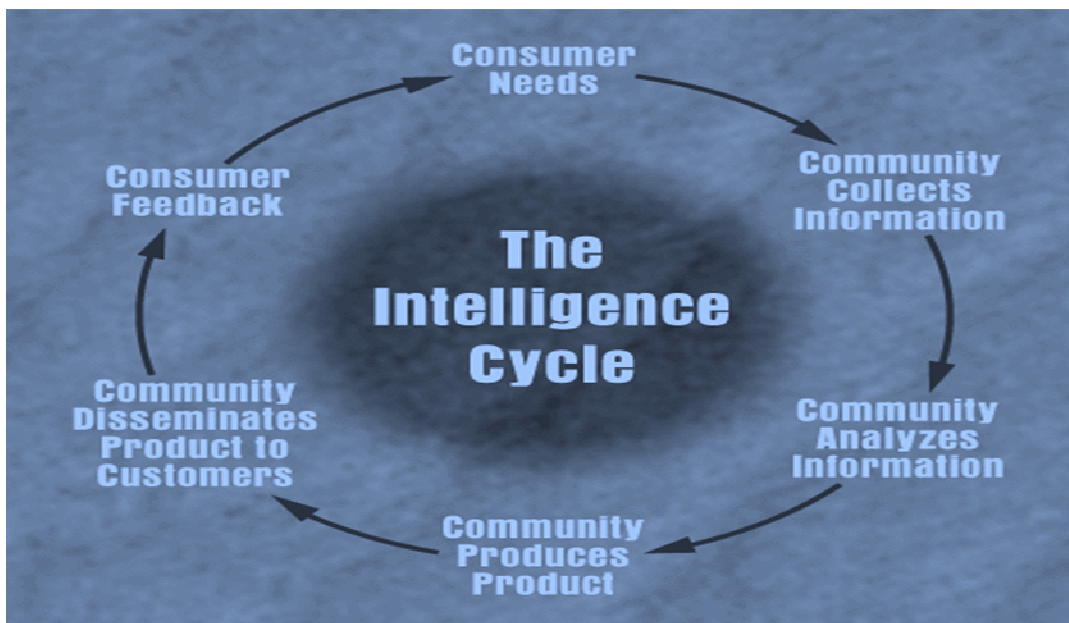


Figure 2 Typical Intelligence Cycle
(Courtesy www.FAS.org)

However given both the reach and scope of web-based data, we are confronted with yet another challenge area which we call web intelligence or WEBINT. Web Intelligence (WI) exploits AI and advanced information technology on the Web and Internet. It is the key and the most urgent research field of IT for business intelligence. It provides a fertile ground for data mining. However, searching, comprehending, and using the semi-structured information stored on the Web poses a significant challenge because this data is more sophisticated and dynamic than the information that commercial database systems store (Han & Chen-Chuan Chang, 2002). However, this offers great promise in such areas as using weblogs for social network mapping. Figure 3, provides an overall view of Web Intelligence Approach.

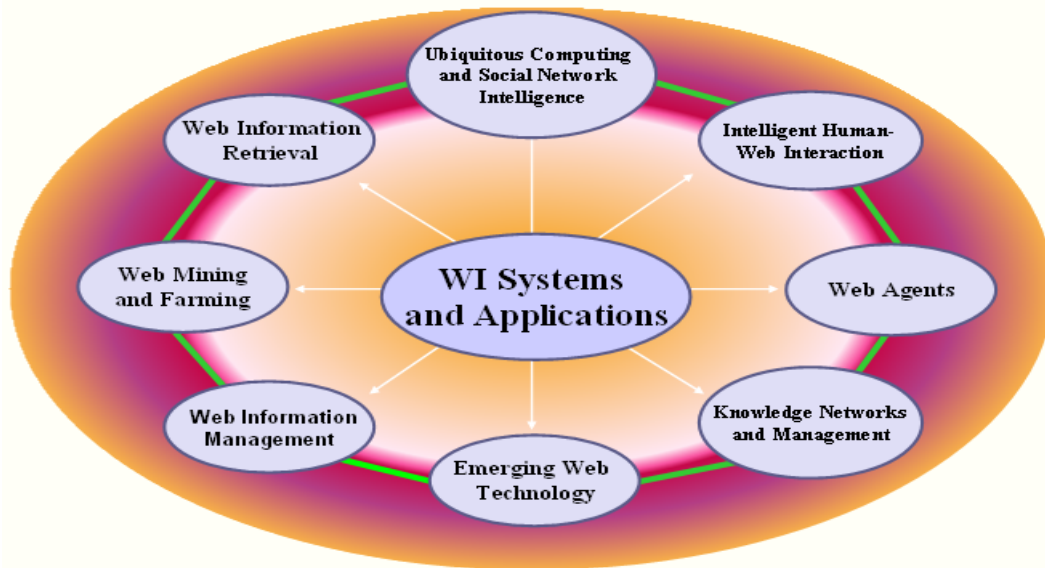


Figure 3 Web Intelligence (courtesy of Web Intelligence Consortium)

However, one should note that in spite of the predominant position of English as the De facto, lingua franca of Web, the contextual logic behind these web pages is far from being linguistically uniform. And there are other sources of information that are not in English. Regardless of the state of the art in artificial intelligence and multi-agent search technology, in order to arrive at novel information, the human interaction and intervention is necessary. The overall process of HUMINT is organized into two major loops of activities: (1) a foraging loop (Figure 1) that involves processes aimed at seeking information, searching and filtering it, and reading and extracting information, and (2) a sense making loop that involves iterative development of a mental model (a conceptualization) that best fits the evidence. Information processing can be driven by bottom-up processes (from data to theory) or top-down (from theory to data in an opportunistic mix) (Pirolli, 2006). Figure 4 provides a notational model for information analysis.

What individuals are willing to accept as a normal part of reality is a complex result of personality, beliefs, cultural environment, and life history. Even people we interact with on a daily basis at home or at work may be operating in a perceived reality greatly divergent from our own (Barnett, 2004). This would seriously impact on the HUMINT value of the library service. It affects not only the analyst (e.g. the librarian, patron or stakeholder), but also the recipient of such analysis. Ever wondered why people continue to make bad decisions while the pertinent information is there at their finger tip?

The primary question is whether there is a criticality of information mass that leads to discovery and whether it is still a process driven by individuals or communities?

Multiple Intelligences and More

Our history from the standpoint of information dexterity could be interpreted by the degree of interaction of individual intelligence with collective intelligence as represented by institutions – such as universities- or constructs – such as web. But exactly what is intelligence? Is it intelligence a property of mind that encompasses many related mental abilities, such as the capacities to reason, plan, solve problems, think abstractly, comprehend ideas and language, and learn. The ability to learn, reason and problem solve. Debate revolves around the nature of intelligence as to whether it is an innate quality or something that is developed as a result of interacting with the environment. Many researchers believe that it is a combination of the two. A casual search on the web shows that many competing definitions exist for one of the most controversial concepts in psychology. The most influential in the assessment of intelligence in workplace settings is 'the innate ability to perceive relationships and identify co-relationships'.

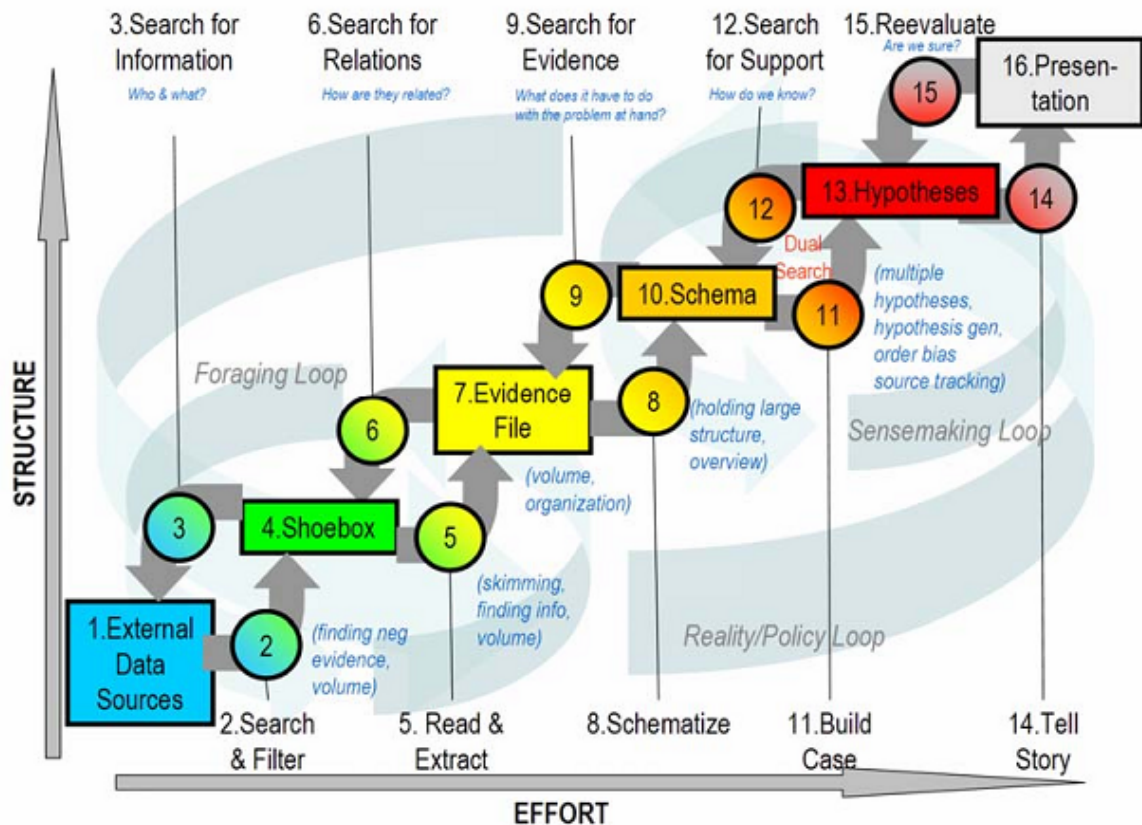


Figure 4 Notational model of Information Analysis (courtesy of FAS.ORG)

. The theory of multiple intelligences was developed in 1983 by Dr. Howard Gardner, professor of education at Harvard University. It suggests that the traditional notion of intelligence, based on I.Q. testing, is far too limited (Gardner, 1983). Instead, Dr. Gardner proposes eight different intelligences to account for a broader range of human potential in children and adults. These intelligences are:

- Linguistic intelligence ("word smart")
- Logical-mathematical intelligence ("number/reasoning smart")
- Spatial intelligence ("picture smart")
- Bodily-Kinesthetic intelligence ("body smart")
- Musical intelligence ("music smart")
- Interpersonal intelligence ("people smart")
- Intrapersonal intelligence ("self smart")
- Naturalist intelligence ("nature smart")

Our schools and culture focus most of their attention on linguistic and logical-mathematical intelligence, Gardner contends (Armstrong, 2004). It seems also that there should be ability or intelligence sought out and cultivated in our networked world. This new intelligence is a response to the cultural change brought about by digital technologies and takes into account the skills and talents possessed by the "symbol analysts" and "masters of change" (Adams, 2004). This could be called digital intelligence that is rooted in our notions of knowledge as well as integrated in our way of knowing. It is through these modes of knowing or ways of knowing that we chart the human process of knowledge internalization.

In the special case of Iran, in spite the fact that there is a web intelligence lab operating from Sharif University of Technology, the situation is heavily in favor of HUMINT portion of information business. A review of Iranian LIS literature is leaning towards new pedagogical models for training of information analysts and information professionals than the use of expert

and artificial intelligence systems. Part of this is cultural. While Iranians wish to think themselves as a nation of poets and writers and artists, its psycho-history demonstrates something quite poles apart. The need for projecting power through technical means is an overriding concern. It has been for this reason that there had been a tremendous explosion of Iranian weblogs. It is also a trait largely responsible for making the government content filtering system next to ineffective.

Human Intelligence (HUMINT) approach would mean designing modalities for information worker to serve as an agent for such discovery. This would mean that within a mediated information environment, the human agent needs to be able to reshape and develop a correct mix of intelligent styles in order to be able to better perform its primary sense-making (analytical) function. UTISL is currently developing skill imparting/enhancement processes addressing this issue. Using a battery of tests gauging different intelligences, certain individuals are recognized as "promising" because of their exceptional capacity in one or more of these **intelligences** while others are identified "less promising". A Quantitative Electro Encephalographic (QEEG) methodology has been indicated to address this issue from a cognitive psychology angle. The methodology can in theory provide a quantitative way for gauging the internalization of the skills taught as well as helping HUMINT with their performance plateau. The Web Intelligence (WEBINT) approach would mean designing modalities for communities in which a complex system is constructed to ensure an expected, and therefore, quantifiable level of discovery. This route would rely heavily on development of artificial intelligence systems especially in the areas of info-mining and information visualization. The Web Intelligence can be programmed to interface with the social network evolving around the technical library services and thus facilitate the discovery of hitherto undiscovered public knowledge. Every client of the library, technically represents a potential HUMINT asset. By providing a tailoring function in a clients interaction with the web intelligence system that supports a notational model of information analysis, these clients could not only gain proficiency in understanding different *thoughtworlds*, but could extract new interpretations of the existing data sets.

Works Cited

- Abushar, S., & Hirata, H. F. (2002). *Filtering with Intelligent SoftwareAgents*. Retrieved 2007, from University of Michigan-Dearborn, Computer Science:
<http://www.engin.umd.umich.edu/CIS/course.des/cis479/projects/FISA.html>
- Adams, N. B. (2004). Digital Intelligence Fostered by Technology. *Journal of Technology* , 93-97.
- Armstrong, T. (2004). *Multiple Intelligences*. Retrieved 2006, from Thomas Armstrong:
http://www.thomasarmstrong.com/multiple_intelligences.htm
- Barnett, J. W. (2004). *Thoughtworlds*. Retrieved 2006, from Federation of American Scientists:
<http://www.fas.org/irp/eprint/thoughtworlds.pdf>
- Berdayes, V., & Murphy, J. W. (2000). *computers, Human Interaction, and Organizations: Critical Issues*. New York: Praeger/Greenwood.
- Bergman, M. k. (2001, April). The Deep Web: Surfacing Hidden Value. *The Journal of Electronic Publishing* .
- Borgman, C. L., & Furner, J. (2002). 4. Borgman, C. L., & Furner, J. (2002). *Scholarly Communication and Bibliometrics*. In B. Cronin (Ed.), *Annual Review of Information Science and Technology* (Vol. 36). New York: Information Today.

Coulburn, T. (2007, Oct 12). *U.S. Government Seeks To Deny The Internet To Its Enemies*. Retrieved Nov 10, 2007, from Informationweek:
<http://www.informationweek.com/news/showArticle.jhtml?articleID=202402122>

Data Conversion Laboratory. (2002, 04 09). *The Next Weave of the Web*. Retrieved 04 22, 2006, from DCLAB.Com: http://www.dclab.com/semantic_web.asp

Eltemasi, M., & Naghshineh, N. (~2008). *Information Intimidation: The avoidance mechanisms*. Tehran: Information Studies Lab.

Espiner, T. (2006, Nov 8). *Rights group names 'Enemies of the Internet'*. Retrieved July 27, 2007, from C-NET News: http://www.news.com/Rights-group-names-Enemies-of-the-Internet/2100-1028_3-6133645.html

Fuller, S., Revere, D., Bugni, P. F., & Martin, G. M. (2004, sept 21). *A knowledgebase system to enhance scientific discovery: Telemakus*. Retrieved July 10, 2006, from Biomedical Digital Libraries: <http://www.bio-diglib.com/content/1/1/2>

Gardner, H. (1983). *Frames of mind: Theory of Multiple Intelligences*. New York: Basic.

Goertzel, B. (2002). *Creating Internet Intelligence: Wild Computing, Distributed Digital Consciousness, and the Emerging Global Brain*. Berlin: Springer.

Gulli, A., & Signorini, A. (2005). *The Indexable Web is more than 11.5 billion pages*. Retrieved 2007, from University of IOWA Computer Science: <http://www.cs.uiowa.edu/~assignori/web-size/>

Han, J., & Chen-Chuan Chang, K. (2002). *Data Mining for web intelligence*. Retrieved 2007, from University of Illinois at Urbana-champaign: <http://www-faculty.cs.uiuc.edu/~hanj/pdf/computer02.pdf>

McGovern, G. (2004, Dec 6). *Technology not answer to every problem*. Retrieved July 2, 2007, from New Thinking: http://www.gerrymcgvorn.com/nt/2004/nt_2004_12_06_technology.htm

Naghshineh, N. (2007). Review of Techno-commercial CI models and their implementations within special research libraries. *Ph.D. Dissertation*. Tehran, Tehran, Iran: University of Tehran.

Pirolli, P. (2006). *Assisting People to Become Independent Learners in Analysis of Intelligence*. Palo Alto: Office of Naval Research.

Postman, N. (1993). *Technopoly: The Surrender of Culture to Technology*. New York: Vintage .

Raschke, C. A. (2003). *The Digital Revolution and the Coming of the Postmodern University*. New York: Taylor and Francis.