CONTROLLED EXPERIMENT ON SEARCH ENGINE KNOWLEDGE EXTRACTION CAPABILITIES

Pasquale Ardimento, Danilo Caivano, Teresa Baldassarre, Marta Cimitile and Giuseppe Visaggio
Department of Informatics, University of Bari, Via Orabona 4, Bari, Italy

Keywords: Search Engine, Knowledge Transferring, Knowledge Acquisition, Knowledge Acquisition Process.

Abstract: Continuous pressure on behalf of enterprises leads to a constant need for innovation. This involves exchanging results of knowledge and innovation among research groups and enterprises in accordance to the Open Innovation paradigm. The technologies that seem to be apparently attractive for exchanging knowledge are Internet and its search engines. Literature provides many discordant opinions on their efficacy, and to our best knowledge, no empirical evidence on the topic. This work starts from the definition of a Knowledge Acquisition Process, and presents a rigorous empirical investigation that evaluates the efficacy of the previous technologies within the Exploratory Search of Knowledge and of Relevant Knowledge according to specific requirements. The investigation has pointed out that these technologies are not effective for Explorative Search. The paper concludes with a brief analysis of other technologies to develop and analyze in order to overcome the weaknesses that this investigation has pointed out within the Knowledge Acquisition Process.

1 INTRODUCTION

In the last century Internet has represented the largest communication and knowledge trasferring media. Moreover data, information, knowledge, experiences contained in the Web increase every day. This phenomena encourages researchers and developers to study and to use all the Internet related aspects (Tonchia, 2003) (Hee-Dong Yang, 1998). In particular, research of knowledge resources through a search engine is an issue of great interest for both research and practitioner communities (Marchionini, 2006) (Gersh, 2006) (Ryen, 2006). Indeed, even if Search Engines have contributed to knowledge research and diffusion, we are aware that search engines have many limitations (Andrews, 2003) (Aswath, 2005). In this sense, (Andrews, 2003) states that 40 percent of companies rate the available search tools as “not very useful” or “only somewhat useful”; other studies emphasize that much time is needed for extracting the searched knowledge (Grandal, 2001). An explanation and description of these limits from the technological point of view are reported in (Papagelis, 2007). In this work we analyse search engine data collection, quality search, and updating of data characteristics. Another accredited analysis of this limit confirms that general queries produce a large amount of documents and that there is not a natural language interface of the search engine (Aswath, 2005). The analysis of Search Engine issues is usually characterized by alternative Search Engine solutions that overcome these issues (Papagelis, 2007), (Moldovan, 2000) (Joachims, 2007) (Al-Nazer, 2007). In (Papagelis, 2007), collaborative search engines that can be adopted between traditional search engines and web catalogues is proposed; while in (Joachims, 2007) a search engine that provides accurate training data towards learning techniques is proposed.

Moreover several new approaches in search engines are beginning to adopt intelligent techniques for improving search precision (Choi, 1998), (Zhang, 2004), (Mingxia, 2005).

Finally, we can’t avoid considering the Experience Base and Experience Factory approach that allow to store, select and search specialized Knowledge and Experience (Basili, 1994).

In this work we do not introduce our own approach to knowledge searching and transferring, which is described in previous papers of the same authors (Ardimento, 2007A), (Ardimento, 2007B). Aim of this work is to investigate the available Search Engine limitations from the user point of view in

Ardimento P., Caivano D., Baldassarre T., Cimitile M. and Visaggio G. (2008).
CONTROLLED EXPERIMENT ON SEARCH ENGINE KNOWLEDGE EXTRACTION CAPABILITIES.
In *Proceedings of the Third International Conference on Software and Data Technologies - PL/DPS/KE*, pages 388-395
DOI: 10.5220/0001894603880395

order to extract some lessons learned and some useful suggestions for searchers and developers that are working in these areas of interest. In this sense we consider our observations of interest for knowledge searching independently from the proposed approach. In fact the new search engine approaches and tools need to start from an accurate analysis of the limitations related to existing approaches and tools. We observed that sometimes the search engine limits are mentioned but not rigorously investigated to overcome them.

Moreover, in spite of the large amount of works (Scoville, 1996), (Leighton, 1997), (Ding, 1996), (Leighton, 1996), (Chu, 1996), (Clarke, 1997) that have evaluated the efficacy of different Search Engines, to the authors' knowledge, no replicable empirical investigations have been carried out concerning the capability of these instruments in an Exploratory Search concerning Knowledge Acquisition. For clearness, Exploratory Search is the set of activities for extracting existing knowledge and analyzing it in order to verify that its relevance allows to learn new results or technologies within a specific knowledge domain (Marchionini, 2006).

As so, this work intends carrying out an empirical investigation that answers the following research question: Are the Search Engines available on Internet effective for Exploratory Search? The investigation analyzes the cause-effect relation among use of the technologies and their effectiveness in an Exploratory Search. It is rigorously described so that other researchers can replicate it to confirm or deny the results. Replication of an empirical investigation allows overcoming contrasting opinions in literature and, at the same time, collecting a set of lessons learned on the current Search Engines.

The rest of the paper is organized as follows: the controlled investigation is described in section 2; section 3 illustrates the measurement model used; results of the study including statistical analysis are presented and discussed in section 4; finally conclusions are made in section 5.

2 CONTROLLED EXPERIMENT

In order to assure experiment replications (Baldassarre, 2008), we have introduced an Exploratory Search Process (ESP) representing the sequence of actions that experimental subjects (Searcher) have to carry out. The process is shown in figure 1. The Searcher specifies a query to a search engine which expresses its need of knowledge (Step 1); the engine provides a set of Extracted Resources (ER). The Searcher selects the resources, among the Extracted Resources, that can be classified as knowledge (Step 2). This set represents the Selected Resources (SR). The Searcher further reviews the SR and selects the ones that contain the requested knowledge (RK). The RK is a more specific search question than the topic specified in the Query. The selected resources make up the Relevant Resources (RR) (step 3).

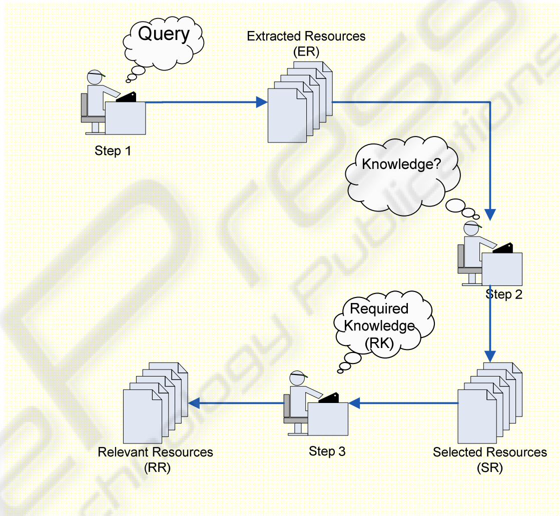

Figure 1: Exploratory Search Process.

2.1 Research Goal

According to the process described above, the research goal is formalized as follows:

Analyze the Search Engine Tools *in order to* evaluate them *with respect to* Effectiveness *from the* Knowledge Searcher *point of view, in the context of* a controlled experiment.

The following research hypotheses have been made:

H_0: The available Search Engines are effective for extracting relevant knowledge.

H_1: The available Search Engines are ineffective for extracting relevant knowledge.

The effectiveness of a Search Engine for extracting relevant knowledge in this work is defined as the ability of a Search Engine to extract useful knowledge and experience according to a defined search request.

2.2 Experiment Variables

The efficacy of a search engine is measured through a factor (Leighton, 1997) that we will call Relevance. Relevance represents the dependent variable of our study, and is defined as the evaluation of in what terms the knowledge contained in the selected resources are relevant for the searcher. Evaluation of Relevance depends from variables that represent the independent variables of our experiment. The independent variables are described as follows:

Searcher (S_k)*:* the experimental subjects are 4 software engineering researchers with a similar experience on research projects carried out in the SERLAB laboratory of the Department of Informatics at the University of Bari, Italy. They are able to evaluate whether the knowledge contained in a selected resource following a search, is relevant from a software engineer perspective.

Search engines (SE_l)*:* the search engines have been selected according to the data on network traffic concerning searches on Internet (http://www.onestat.com/):

SE_1: Google (http://www.google.com/);
SE_2: Msn (http://www.msn.com);
SE_3: Yahoo (http://www.yahoo.com);
SE_4: Altavista (http://www.altavista.com/);

Query and Query DetailLevels (Q_{ij})*:* 4 different queries, each with 3 levels of detail have been selected. Qij is the i-th Query with a j-th level of detail:

Query 1
- Q_{11}: <“Software Engineering Quality”>
- Q_{12}: <“Software Engineering Quality” “Process Quality” >
- Q_{13}:<“Software Engineering Quality” “Process Quality” “Process Performance”>

Query 2
- Q_{21}: <“Software Engineering Process”>
- Q_{22}: <“Software Engineering Process” “Process Model” >
- Q_{23}: <“Software Engineering Process” “Process Model” “Quality Metric” >

Query 3
- Q_{31}: <“Software Engineering Best Practices” >
- Q_{32}: <“Software Engineering Best Practices” “Process Best Practices”>
- Q_{33}: <“Software Engineering Best Practices” “Process Best Practices” ”Software Development Process” >

Query 4
- Q_{41}: <“Software Engineering Development”>
- Q_{42}: <“Software Engineering Development ” “ Product Development”>
- Q_{43}:<“Software Engineering Best Practices” “Product Development” “Quality Metrics”>

2.3 Experiment Description

The experiment was organized in 4 experimental runs, one for each Search Engine. In each run 4 Queries, with the three levels of detail, were assigned to each Searcher. Searchers used the same Search Engine.
Each run was divided into two phases: first, each searcher was assigned to the lowest level Query. Step1 of the ESP was then executed; the Search Engine produced the Extracted Resources. According to the results, the searchers carried out the selections at Step2 of the ESP, within 30 minutes. In Step2 the i-th searcher produced a set of Selected Resources SR_{i1}. After 30 minutes the intermediate level Query was given to each searcher, the process was iterated and led to SR_{i2}. Finally, after 30 minutes the highest level Query was given to the searchers, and the process iterated, producing SR_{i3}. At the end of this first phase the Searcher was informed of the Required Knowledge (RK). At that point, each Searcher extracted, among the SR_{ij}, the resources containing knowledge corresponding to the RK. A set of Relevant Resources, RR_{ij}, corresponding to the SR_{ij}, were produced.

The RK, corresponding to the queries are:

RK_1: Quality models to evaluate process performances in Software Engineering, described so that they can be transferred without help of their producers.
RK_2: Metrics for evaluating the quality of the process models, described so that they can be transferred without help of their producers.
RK_3: Best Practices on the Software Engineering development processes, described so that they can be transferred without help of their producers.
RK_4: Quality metrics of Software Engineering for product development, described so that they can be transferred without help of their producers.

Each searcher used their own self-defined process for selecting the detailed resources, according to their own experience in the knowledge domain. This procedure remained tacit, in that it was out of the scope of the investigation. Selection was carried out within a time limit of 45 minutes, in particular 20 min. for SR_{i1}, 15 min. for SR_{i2}, and 10 min. for SR_{i3}. The time available for the Searchers was less than the time estimated for evaluating the Extracted Resources and the Selected Resources. This restriction was necessary to be sure that the Searchers dedicated the same amount of time to their tasks and were not influenced by secondary effects that could have biased the results.

Given the previous considerations, each RUN lasted 135 min, other than the time that each Search Engine implied for producing the Extracted Resources. This time was considered non relevant. The experimental design is reported in Table 1.

Table 1: Experimental Design.

Experimental Subject	**RUN_1**	**RUN_2**	**RUN_3**	**RUN_4**
Searcher $_1$	SE_1,Q_{1j}	SE_2,Q_{2j}	SE_3,Q_{3j}	SE_4,Q_{4j}
Searcher $_2$	SE_1,Q_{2j}	SE_2,Q_{4j}	SE_3,Q_{1j}	SE_4,Q_{3j}
Searcher $_3$	SE_1,Q_{3j}	SE_2,Q_{1j}	SE_3,Q_{4j}	SE_4,Q_{2j}
Searcher $_4$	SE_1,Q_{4j}	SE_2,Q_{3j}	SE_3,Q_{2j}	SE_4,Q_{1j}

2.4 Metric Model

The research question, related to the goal of the study, that we have tried to answer is the following:

- What is the relevance of the search engines in Internet?

We will consider a search engine relevant if it allows the user to extract useful knowledge according to an assigned search scope. The search scope is assigned with refer to the queries and RK assignment.

In order to answer to the proposed search question we have introduced the following metrics, which are named and described in Table2:

Table 2: Metric Model.

Relevance	
Metric Name	**Metric Description**
Extracted $Resources_{klij}$	Number of resources extracted by the k_{th} Search Engine by the l_{th} searcher using the i_{th} query with j_{th} level of detail
Relevant $Resources_{klij}$	Number of extracted resources by the k_{th} Search Engine that are selected by the l_{th} Searcher, because considered relevant knowledge to answer the RK, using the i_{th} query with j_{th} level of detail.
$Relevance_{klij}$	$\frac{RelevantResources_{klij}}{ExtractedResources_{klij}}$

3 EXPERIMENTAL RESULTS

The data collected during the investigation have been synthesized through descriptive statistics in order to represent them graphically, identify possible outliers and decide if it they must be eliminated from the sample. Finally, data have been analyzed through hypothesis testing, where observations of statistical analysis were statistically validated with respect to a significance level.

3.1 Descriptive Statistics

Figure 2 reports the Line-Plot of the Mean Relevance for each Searcher, figure 3 reports the Line Plot of the Mean Relevance for each Search Engine and finally, figure 4 illustrates the Box-Plot of the Relevance distribution with respect to a level of detail of the Query. $RDetail_j$ is the distribution of $Relevance_{klij}$ as k, l, and i vary.

According to the Relevance definition given in Table 2, we obtained that the Relevance value is very smaller than 100%. It means that the number of extracted resources that can be considered relevant according to a given RK is very small.

This Relevance value is conditioned by the introduced independent variables as reported in the shown graphs.

In particular, according to figure 2, it is possible to observe that there are no significant differences in Mean Relevance between the results of the Searchers, even if there are some minor differences caused by the ad hoc selection procedures carried out by each Searcher. These differences are reflected in the results.

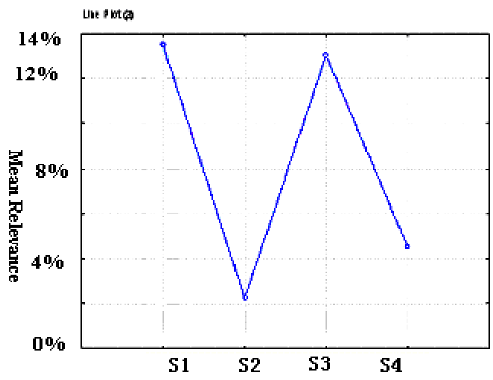

Figure 2: Line Plot Relevance/Searcher.

According to figure 3 the Mean Relevance achieved by each Search Engines is small. Some non significant differences could be linked to their different navigation techniques and/or different selection algorithms.

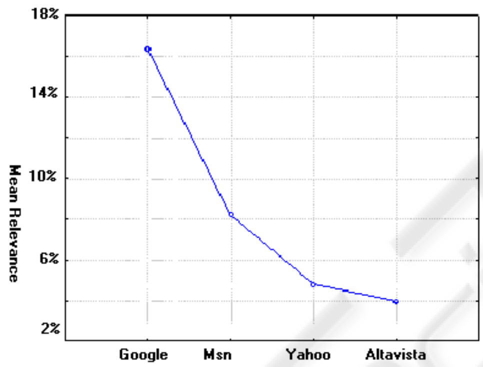

Figure 3: Line Plot Relevance/Search Engine.

Finally, in figure 4, box plots describe the relevance results range for each query detail level. In the box plot, median values are indicated. We can observe that there is a consistent difference in results among the three query levels of detail. In the lowest level of detail the distribution of relevance values are concentrated around 0%. Increasing the detail level, relevance assumes values around a larger range. The Relevance values result being small also in the case of a higher detail level. Given the previous considerations, Figures 2, 3, and 4 confirm small relevance values.

3.2 Hypothesis Tests Analysis

Relevance has been investigated to confirm the considerations pointed out by the descriptive analysis and avoid threats. For this reason, two types

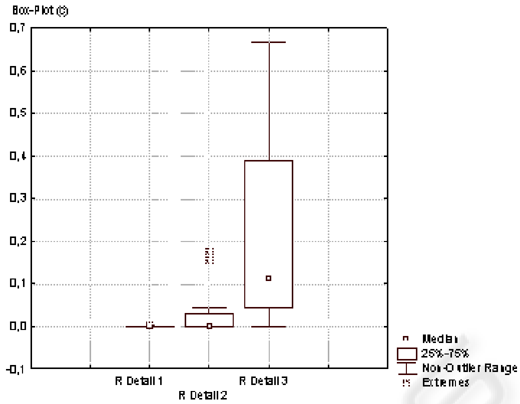

Figure 4: Box-Plot Relevance/Query Detail.

of tests have been carried out:

- Kruskal-Wallis, a non-parametric test alternative to one-way (between-groups) ANOVA. It is used to compare three or more samples, and it tests the null hypothesis that the different samples in the comparison were drawn from the same distribution or from distributions with the same median. Thus, the interpretation of the Kruskal-Wallis test is basically similar to that of the parametric one-way ANOVA, except that it is based on ranks rather than means (Siegel and Castellan, 1988). Note that all the Kruskall Wallis tests have been carried out on mean measures of Relevance with respect to the different detail levels $((\sum_j Relevance_{klij})/3)$, because the different detail levels have not been considered in this test, rather they have been analyzed through a separate one.
- Friedman ANOVA: it is a non parametric alternative to one-way repeated measures analysis of variance. In particular in the context of our analysis it is used to investigate presences of statistically significant differences in the values of Relevance collected with respect to the 3 different levels of Query details. This test assumes that the variables (levels) under consideration be measured on at least an ordinal (rank order) scale. The null hypothesis for the procedure is that the Relevance for the different levels of detail, contain samples drawn from the same population, or specifically, populations with identical medians.

3.3 Results

Tables 3, 4, 5, and 6 report test values for Relevance. According to the descriptive analysis results, Kruskal-Wallis points out that no statistically significant differences exist between Search Engine, Queries and Searcher. The significance values are the following: $p = 0.2695$; $p = 0.7924$; $p = 0.1064$.

It can be seen in table 3 that the number of *measures* considered for each Search Engine is 4, i.e. 1 for each Searcher that answered ad Query using a SE only once, for a total of 16 *measures*.

For clearness, each of the 16 values has been associated to a Rank that corresponds to its position in an increasing order. In case of n equal values in positions p1, p2, …, pn, the assigned rank is (p1+p2+…+pn)/n. The *Sum of Ranks* corresponds to the sum of the ranks related to the 4 points of the Search Engine.

The *Sum of Ranks* is displayed in the rightmost column of the spreadsheet. The Kruskal-Wallis test isn't significant ($p = 0.2695$). Thus, we can conclude that the Search Engines were not significantly different from each other with respect to Mean of Relevance.

So, we can conclude that Search Engines give different responses to the same queries although their differences in terms of knowledge relevance are not statistically significant.

In table 4, the number of *measures* considered for each Query is 4, corresponding to the 4 Searchers that have answered a query once and used a different Search Engine, for a total of 16 *measures*. The Kruskal-Wallis test isn't significant ($p = 0.7924$). Thus, we can conclude that if we consider the Mean Relevance, there no significant difference between the selected Queries.

So, we can conclude that the differences in query contents can influence the effectiveness of the ESP; however the difference is not statistically significant within the same knowledge domain.

Table 3: Dependance of Relevance from the Search Engine.

	Measures	Sum of Ranks
Google	4	49,00000
MSN	4	35,00000
Yahoo	4	27,00000
Altavista	4	25,00000

Table 4: Dependance of Relevance from the type of Query.

	Measures	Sum of Ranks
Q_1	4	32,00000
Q_2	4	29,00000
Q_3	4	33,00000
Q_4	4	42,00000

In Table 5, the number of *measures* considered with respect to each Searcher is 4, corresponding to the results that each Searcher obtained in each of the 4 Runs, for a total of 16 *measures*. The Kruskal-Wallis test isn't significant ($p = 0.1472$). Thus, we can conclude that if we consider the Mean of Relevance, there aren't significant differences among different Searchers.

So, we can conclude that searchers have analogous experiences in the search knowledge domain, although different procedures are used for selecting relevant knowledge, the differences in results are not statistically significant.

Table 5: Dependance of Relevance from the Searcher.

	Measures	Sum of Ranks
$Searcher_1$	4	49,00000
$Searcher_2$	4	22,00000
$Searcher_3$	4	42,00000
$Searcher_4$	4	23,00000

Finally, Table 6 reports the value of *Sum of Ranks*, the average Relevance, the *mean of* rank order correlation between the cases and the *standard deviation* for each of the levels of detail of the Relevance Query. For clearness, the average is intended as the average value of the ranks calculated for each of the sample data points.

The first observation on these results is that the *Average* and *the Sum of Ranks* increases as the level of detail of the query increases. This confirms that the level of detail of the question allows for a greater relevance of the Search Engine. Also, the Friedman Anova test shows that there are highly significant differences ($p < 0.00176$) between the different Detail Levels. This difference is statistically significant.

This confirms that for relevance, a statistically significant difference exists among the results obtained with queries of different detail levels. So, a higher level of detail in the queries increases effectiveness in the relevance of the resources extracted from the Search Engine.

Table 6: Dependance of Relevance from the levels of detail of the Query.

	Average	Sum of Ranks	Mean	Std.Dev.
R_1 Detail	1,437500	23,00000	0,000390	0,001117
R_2 Detail	1,906250	30,50000	0,030081	0,055850
R_3 Detail	2,656250	42,50000	0,219395	0,215998

4 CONCLUSIONS AND FUTURE WORK

The experiment carried out has allowed us to give a preliminary answer to the research question. We can conclude that: the available Internet Search Engines are not relevant. Their capabilities in extracting relevant knowledge according to an assigned search goal are very low and they can't be used to extract reusable innovative knowledge to transfer between research organizations or enterprises.

These results are independent both from the Search Engines, from the Searchers and the searching queries. Moreover, we identified a relationship between Search Engine Relevance and detail levels of the searching Queries. As the detail level increases, the Search Engine appears more relevant, although, in all these cases, results are not satisfactory.

Given these considerations, the only difference that can be used to improve the rate of resources containing knowledge and relevant knowledge is the Query Detail Level. Note that the level of detail is not managed in the same way by all search engines, and usually depend by their parsers. So, a greater level of detail in the Query, may not necessarily assure that the Search Engine is able to satisfy the knowledge content the Searcher is interested in. Also, consider that the results are not satisfactory what ever the level of detail of the Query.

The proposed work has empirically shown some shared opinions about the low quality of the knowledge available using the internet search engines. These considerations need to be validated through the replication of the experimentation and furthermore through a family of investigations. The family of experiments will allow us to obtain a rigorous list of search engine limits from the user point of view. These lists will be used to characterize and define a new and innovative approach to knowledge searching processes and tools.

The experiment also has suggested to investigate the proposed ESP. The ESP could be encapsulated and refined in a knowledge transferring and searching approach.

Given the results of the experiment, the authors feel the need for further investigating the ESP, restricting its use to a specialized repository, in order to provide more valid solutions to knowledge transferring. For example, for Software Engineering, it may be useful to restrict the search to ACM or IEEE digital libraries, since they are repositories containing knowledge resources.

A further investigation would therefore consist in verifying the capability of selecting relevant knowledge according to specific queries. Also, a possible threat related to the knowledge evaluation process of a selected source can be overcome by adopting a rigorous process in step3 that is independent from the Searcher.

Finally, the authors intend extending their work and identifying on one hand, specific repositories for collecting formalized knowledge, and on the other tools for collecting and formalizing tacit knowledge to be stored in specific repositories. For clearness, the intention of our future work goes towards an Experience Factory (Basili, 1994), (Ardimento, 2007A), (Ardimento, 2007B).

REFERENCES

Marchionini, G., 2006. Exploratory search: From Finding to Understanding. Communications of the ACM, April, vol. 49, n. 4.

Leighton, H., Srivastava, J., 1997. Precision among WWW search services (search engines): AltaVista, Excite, HotBot, Infoseek and Lycos. Retrieved June 11, 2005 from http://www.winona.edu/library/webind2.htm/

Gersh, J., Lewis, B., Montemayor, J., Piatko, C., Turner, R, 2006. Supporting insight-based Information Exploration in Intelligence analysis. Communications of the ACM, April,vol. 49, n. 4.

Ryen WW, Kules, B., Drucker, S.M., Schraefel, M.C., 2006. Supporting Exploratory Search. Communications of the ACM, April, vol. 49, n. 4.

Tonchia, S., Turchini. F., Tramontano, A., 2003. Reti organizzative e nuove tecnologie: l'azienda estesa della conoscenza. Gestione d'impresa e innovazione, Il Sole 24 Ore.

Hee-Dong Yang, Mason, R.M., 1998. The Internet, Value Chain Visibility, and Learning. Thirty-First Annual Hawaii International Conference on System Sciences, IEEE, vol. 6, pp. 23 – 32.

Andrews, W., 2003. Visionaries Invade the 2003 Search Engine Magic Quadrant. Gartner Report, April 2003.
Aswath, D., Ahmed, S.T., James D'cunha, Davulcu, H., 2005. Boosting Item Keyword Search with Spreading Activation. Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, pp. 704 - 707.
Grandal, N., 2001. The Dynamics of Competition in the Internet Search Engine Market. International Journal of Industrial Organization 19, pp. 1103-117.
Dan I. Moldovan and Rada Mihalcea, 2000. Using WorldNet and Lexical Operators to Improve Internet Searches. IEEE Internet Computing, vol. 4, n°1, pp. 34-43.
Zhang, Y.Y., Vasconcelos, W., Sleeman, D., 2004. OntoSearch: An Ontology Search Engine. Proceedings of the Twenty-fourth SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, Cambridge, UK.
Mingxia, G., Chunnian, L., Furong, C., 2005. An Ontology Search Based on Semantic Analysis. Proceeding of the Third International Coerence on Information Technology and Applications, vol. 1, pp. 256 – 259.
Scoville, R., 1996. Special Report: Find it on the Net!.PC World, January, http://www.pcworld.com/reprints/lycos.htm/
Leighton, H., Srivastava, J., 1997. Precision among WWW search services (search engines): AltaVista, Excite, HotBot, Infoseek and Lycos. Retrieved June 11, 2005 from http://www.winona.edu/library/webind2.htm/
Ding, W., Marchionini, G., 1996. A comparative study of the Web search service performance. Proceedings of the ASIS 1996 Annual Conference, vol. 33, pp. 136-142.
Leighton, H., 1996. Performance of four WWW index services, Lycos, Infoseek, Webcrawler and WWW Worm. Retrieved June 10, 2005 from http://www.winona.edu/library/webind.htm/
Chu, H., Rosenthal, M., 1996. Search engines for the World Wide Web: a comparative study and evaluation methodology. Electronics proceedings of the ASIS 1996 Annual Conference, October, 33, 127-35.
Clarke, S., Willett, P., 1997. Estimating the recall performance of search engines. ASLIB Proceedings, vol. 49, issue 7, pp. 184-189.
Siegel, S., Castellan, Jr., 1988. Nonparametric statistics for the behavioral sciences. London: McGraw-Hill, 2nd edition.
Basili, V.R., Caldiera, G., Rombach, H.D., 1994. The Experience Factory. Encyclopedia of Software Engineering, Anonymous New York: John Wiley & Sons, pp. 470-476.
Baldassarre, M.T., Visaggio, G., 2008. Description of a Paradigm for Empirical Investigations. Technical Report: http://serlab.di.uniba.it/
Papagelis, A., Zaroliagis, C., 2007. Searchius: A Collaborative Search Engine. Proceedings of Eighth Mexican International Conference on Current Trends in Computer Science, IEEE, pp.88-98.
Al-Nazer, A., Helmy, T., 2007. A Web Searching Guide: Internet Search Engine & Automous Interface Agents Collaboration. International Conferences on Web Intelligence and Intelligent Agent Technology, IEEE, pp. 424-428.
Joachims, T., Radlinsky, F., 2007. Search Engines that Learn from Implicit Feedback. IEEE Computer, vol. 40, n° 8, pp. 34-40.
Choi, Y., Yoo, S., 1998. Multiagents learning approach to WWW information retrieval using neural networks. Proceedings of the Fourth International Conference on Intelligent User Interfaces, Redondo Beach, CA, pp.23-30.
Ardimento, P., Baldassarre, M.T., Cimitile, M., Visaggio, G., 2007A. Empirical Validation on Knowledge Packaging supporting knowledge transfer. Proceedings of 2nd International Conference on Software and Data Technologies, pp. 212-219.
Ardimento, P., Boffoli, N., Cimitile, M., Masone, A., Tammaro, A., 2007B. Empirical investigation on knowledge packaging supporting risk management in software processes. Proceedings of The IASTED International Conference on Software Engineering SE, Innsbruck (Austria), pp. 30-36.