Unveiling the perception of blogs as marketing tools: a randomized controlled experiment

Francesca Bandini

2014

Supervisor Paolo Pin
Assistant supervisor Nicola Dimitri
To those who consider me better than I am.

A chi mi considera migliore rispetto a quello che sono.
# Table of contents

Abstract .......................................................................................................................... 4
Introduction .................................................................................................................... 5

1. Search Engine Optimization and quality results ....................................................... 9
   1.1. Search engines and Google .............................................................................. 10
   1.2. Information retrieval and algorithms ............................................................... 14
   1.3. PageRank and Google .................................................................................... 16
   1.4. Search Engine Optimization (SEO) techniques ............................................... 19
   1.5. Speculating about quality .............................................................................. 24

2. Blogs .......................................................................................................................... 27
   2.1. Definition of blog ........................................................................................... 28
   2.2. A single word, several purposes .................................................................... 30
   2.3. State of the blogosphere ................................................................................. 34
   2.4. Literature review on blogs’ influence ............................................................... 37

3. The role of the blogs in online marketing companies: an experience ..................... 44
   3.1. Blogs and link exchange ............................................................................... 46
   3.2. A conceptual framework for supporter blogs ............................................... 48
   3.3. The development of the contents ................................................................... 51
   3.4. Public Relations and blogs ............................................................................ 55
   3.5. Banners and effectiveness ............................................................................. 59

4. Assessing blogs’ impact: a randomized controlled experiment .................................. 65
   4.1. The randomized controlled experiment ......................................................... 65
   4.2. Hypotheses .................................................................................................... 71
   4.3. Statistical analysis fundamentals .................................................................... 73
   4.4. Results ............................................................................................................ 80

Conclusions ................................................................................................................... 86

Appendix a ..................................................................................................................... 89

Appendix b .................................................................................................................... 94

References ..................................................................................................................... 96

Francesca Bandini
Abstract

The increasing weight of the Internet is shaping our knowledge and ideas, since we strongly rely on this medium to find information. Therefore the positioning of a Website within the results retrieved by the most popular search engine, Google, acquires importance. Particularly harmful is the high rank of blogs pretending to have informational purposes, but which act as marketing tools to expose advertising, often banners.

The research aims at investigating the effect of banners contained in blogs, which have not a clearly promotional objectives, on the Italian population. A positive linear relationship between banners and perceived usefulness of the Web contents was found by implementing a randomized controlled experiment.
Introduction

The idea of the thesis came out during a period of internship at *Suerte*, an online marketing company based in Southern Spain, where the role of blogs is crucial. By working there, it was acknowledged that the depletion of blogs is widespread, above all for marketing purposes. Thence, it appeared interesting to detect the perception of blogs, used as marketing tools, from the user viewpoint and a randomized controlled experiment turned out to be the most suitable way.

Legitimately, blogs may embody marketing tools, the problem arises when people are not aware of their purpose and consider them pure sources of information, at the same level of newspapers or encyclopedia. This ignorance is due to the difficulty to trace the presence of a blog, to distinguish it from other typology of Websites and to the scarce level of literacy in the World Wide Web domain.

The issue acquires incredible importance because of the growing number of Internet users, the increase availability of wireless connections and last but not least, the surge of one search engine over the competitors, namely Google. Furthermore, search engines are becoming our major source of information, in fact, it has been argued that for 40% of the Italians, the Internet represents the substitute of the paper based news (La Repubblica, 2013): almost all the users (96.2%) are able to use search engines (ISTAT, 2013). Indeed, the thesis investigates also Google's working and the connected harmfulness of blogs, since their high visibility.

Technically Web search engines are software aimed at information retrieval: they explore their databases and display the more relevant results in harmony with the words typed by the users.

Search engines represent the gateway to the Internet, they allow to consult a massive quantity of data, coming from around the world and facilitate the access to the
information; by helping us to deal with this flow, they are at the heart of our understanding of the reality (Seymour et al., 2011). In addition, they are in line with the frenetic rhythm of our society, in fact, they have been increasingly faster: today they are able to give back instantaneous results and may be used by everybody without the need of expertise.

However, the tendency to rely on the latter, almost exclusively, is affecting the quality and quantity of information at disposal. Indeed, nobody is questioning the accuracy of the search engine results and the common attitude is to stick on the first page: scrutinizing more pages would lead to a more time consuming activity, so that the advantages of search engine would be nullified.

The stickiness on the first page of results has given rise to the notion of *democratic censorship* (Ramonet, 2001), because there is not a real censor on the knowledge, nowadays, but some pieces of information are confined in less accessible places. This restricted access needs major emphasis, since the quasi monopolistic position occupied by Google.

Google is the most used software and its success is mostly connected to the relevance of the results retrieved. Relevance can be defined as the degree of coherence between the page returned and the user's request (Smith, 2003). The problem connected to the relevance is that the parameters defining relevance can be learned, so that ranking higher does not reflect necessarily more authority or credibility. Actually, something very different could be true, because the more money you can spend to optimize your Website, the more the popularity gained. In fact, one of the main points of the enquiry is whether the highly ranked pages deserve trustworthiness or not.

The potential damages which might be caused by Google are sharp, since the increasing influence of the Internet on our everyday life (Wiggins, 2003): people are strongly relying on SERP to make decision, for instance 32% of the Italians consult social media to obtain information, before a purchase and almost the totality of people with the Internet access, 84 %, search for information online and then buy at the store: the phenomenon rate is increasing (Nielsen, 2011).
The intrusiveness of the Internet is also due to the availability of the connections and devices allowing them, such as mobile phones and tablets. An apocalyptical scenario may be forecasted owing to the advent of the ubiquitous Internet (Kenny & Marshall, 2000), which will be everywhere and incorporated in always more items, such as e-wallets. The consequences will influence primarily the business sector, because companies will be able to approach customers in more occasions and a shift from the importance of the content to that of the context will be implied. Therefore, on one hand it appears necessary to have search engines, to arrange the huge amount of information, on the other hand the problem inherent to their trustworthiness should be addressed.

The other focal topic of the study is connected to the theme of blogs, which is strictly related to the Internet pervasiveness, because the first page of results retrieved by search engines is populated by them: blogs reveal to be quite powerful in the link citation, a good parameter to be ranked high, so that bloggers have learned how to influence the results of the search engines (Wiggins, 2003). Google is trying to find a way to condemn such behavior, classifying it as SPAM, but the measures implemented are still weak to deal with the scale of the events.

The increasing number of blogs raises concerns about the quality of information (Weerkamp & de Rijke, 2012), because usually blogs represent a marketing tool, therefore presumably their prominent rank is given by their compliance with search engine's working and not by their quality. The issue becomes far more important, when marketing objectives of blogs are hidden and people are led to consider them as a credible source of information.

Under the scope of the presented analysis credibility refers to authenticity and integrity, so that a source is judged credible if it clearly declares the reasons of its existence within the World Wide Web: for example an e-commerce Website, such as Amazon or E-Bay, is considered credible because its nature and purpose appear straightforwardly. Contrary, a Website offering cooking recipes for example, but

Francesca Bandini
holding masked promotional purposes, is not a credible source, because the user is not aware of the promotional aim.

The impact of those blogs containing advertisings needs more investigations: there is not a large body of literature enquiring this topic; above all, no studies inherent to Italy were found. Indeed, the cornerstone of the thesis is represented by an experiment to probe the effects of advertisements on blogs' readers. The randomized controlled experiment design was chosen because it assures a causal relationship, in fact, it is in our interest to transmit relevant and valuable results; this goal is plenty achieved because the main findings reveal that the presence of banners increases the perceived authoritativeness of a Website. The outcome is not only statistically significant, but also internal valid: the variables which might render the causal relationship false have been excluded. Moreover, the result has been checked for external validity.

The presented work departs from the analysis of search engines and their functioning, stressing the aspects of Google, then it considers the blogs' phenomenon and the related literature, subsequently the experiential information acquired, during the internship, are submitted and in the final part, the randomized controlled experiment is explained.
1. Search Engine Optimization and quality results

Search Engine Optimization (SEO) is an online marketing technique, dealing with visibility within the Web search engine results: the visibility refers to the striking presence of a Website in the results of the organic or natural search. The sites are ranked according to some criteria, given by the algorithm of the engine, but people cannot pay to obtain a better positioning. Instead, when we discuss about the paid search, the related technique is the Search Engine Marketing (SEM), which aims at placing Websites in the first places at the top, on the right or at the bottom of the Search Engine Results Page (SERP), as it can be seen from Figure 1, the paid search results provided by Google are under an “Ads” label; even in this case there is an onward competition, reflecting the increasing importance of the Internet for marketers. In principle, the distinction between the two kinds of results should be clear to users, but the boarders are increasingly overlapping.

![Snippet of a search on Google Web search engine.](image)

Figure 1. Snippet of a search on Google Web search engine.
SEO and SEM pertain to the sphere of the online marketing, as well as, e-mail marketing, consisting in the application of traditional mails sent by post to the Internet. Or the social media marketing, which attempts to take advantage of the social media platforms, or strictly connected to the argument, there is the viral marketing, consistent with the knock on effect derived from a specific advertisement. A further option is the paid advertising concerning sponsored advertisement in the Websites, such as banners. Before proceeding to examine the topic of SEO and quality, it is necessary to outline details about search engines.

1.1. Search engines and Google

Firstly, it is fundamental to highlight the main functions of a Web search engine, whose main objective is to give back to the user Websites which are in line with his informational needs.

A search engine stores data of the different Web pages and thanks to the html code, the data are retrieved through Web crawlers, also known as spiders: this stage of retrieval assumes the name of Web crawling (Seymour et al., 2011). Web crawling corresponds to the searching activity, from the hardware viewpoint.

Another important action is linked to the indexing, where the contents of the pages stored are analyzed and categorized, according to particular criteria, into an Index server, afterwards the information will be matched to the query (the request of the user), with the purpose to make the retrieval as quick as possible.

Secondly, the mechanical work behind a SERP (Fig. 2) might be explained in the following essential way: all the process starts with the user, who types a certain query, then, the Web server takes into account the request sent. The Web server is a sort of link between the machine and the human being, in fact, it transmits the query to the Index server, which, in turn, analyzes the data stored and chooses which pages contain the information desired. After that, the Index passes the query to the Doc server, which physically retrieves the relevant documents (Seymour et al., 2011).
Thirdly, it is interesting to briefly review the history of the search engines, in order to underline that Google has not always been the most powerful.

One of the oldest is Archie (Seymour et al., 2011), created in 1990: it was a program responsible for finding documents on the Internet, its structure was quite basic, because there was no need of indexing the results, since their limited number.

In the 1990s, another important search engine was Jump Station (1993), which was using the three major characteristics of actual search engines: searching, indexing and crawling. However, the results given were still poor, since the search was restricted to the title. The progress in this respect arrived soon, thanks to WebCrawler, launched in 1994, which was the first software allowing the full text research.
Another step forward was accomplished by Alta Vista, created in 1995, indeed, it entailed also a natural language search, so that the user could type questions as "how many years Clinton is" without incurring in nonsense results for "how" and "many". Thanks to the quality provided, it became the most consulted search engine of that time. Other important features were its rapidity and its capacity to deal with millions of documents.

Things changed in 1998, when Google was launched by Larry Page and Sergey Brin with the aim to create an honest and trustworthy search engine, since AltaVista was found guilty of selling the first positions of the SERP. Google's revolutionary difference, which made it the winner, consists of its algorithm's parameter, named PageRank (Seymour et al., 2011).

PageRank enables to sort results according to the relevance and represents a development of another parameter previously used, the link popularity: evaluating a Website by means of link popularity leads to attribute more importance to sites being cited (through a link) within other sites. Google refined the link popularity mechanism, by considering also the quality and the authority from which links are received, for instance a link from a governmental institution counts more than one from a blog.

Neither PageRank nor Link popularity are new system for weighting the information, in fact, they derive from the method used to retrieve academic journal (Paragraph 1.2).

Despite the undoubted value of its PageRank, the growing popularity of Google should be tied to its capacity to introduce continuously innovations. Moreover, Google became a well established brand, known all over the world, also because of the importance given to the user, for instance it provides with the correction of the grammar through a suggested research or the autotype of the query (O’Leary, 2010). In addition, it allows to easily find information on the Internet, for the first time the user could retrieve the information without knowing exactly what he is looking for (Hazan, 2013). As a demonstration of Google's importance, the name of the company has become synonym of searching the Internet (to Google). Furthermore, it is offering additional services, such as the e-mail service, the drive, the social network (Google
Plus), software products to analyze Websites or to improve a blog and so forth and so on, completely for free: Google is more than a simple information retrieval instrument (Bar-Ilan, 2007).

To sum up, the striking characteristics of Google which made it the preferred search engines, are its speed, its perceived reliability, the amount of materials indexed, its deep research inside the text (full text research), its usefulness mostly due to the easier access to information provided (Dawson & Hamilton, 2006). Then Google is always on the side of people by helping them whenever they need: it has several guides for its services, courses to train for particular applications and releases information about the company. The progressive weight of Google can be ascribed to its user centered philosophy and to the completeness of the solutions offered.

Figure 3. Google (2014a).
In view of all that has been mentioned so far, one might suppose that Google embodies the perfect option to deal with the huge flow of information produced by our society, instead it holds some not negligible drawbacks, that in recent years have been drawing the attention of many studies: in the next sections some specific problems inherent to Google's algorithms are underlined.

1.2. Information retrieval and algorithms

Before the advent of the World Wide Web, there was a concern regarding the quality of information retrieval among scholars. Indeed, the problem of gauging papers and publication's authority has a long history and many systems have been invented to cope with it. Relevant to this essay is PageRank's ancestor, the so called impact factor proposed by Garfield in 1975. The method, whose aim was to measure the impact of scientific journals, establishes the importance of a journal, within a certain year, according to the number of citations it received in a period of two years. The impact factor appears fallen, because it attributes the same importance to all the citations.

An improvement of the system was accomplished thanks to Pinski and Narin who pointed out that a journal enjoys a greater impact if it has been cited by other important journals, so that the recursive issue was introduced (referring to a potential number of citations which do not loop).

In the late '90s, Kleinberg refreshed the concept introducing the notion of hubs within the Internet. This represents one of the striking examples of the attempts to establish an algorithm capable of retrieving quality results. In fact, the model built is a viable option to Google's algorithm.

Kleinberg clearly differentiates the term authority from relevance (Chakrabarti et al., 1999): relevance corresponds to the adherence of the results to the query typed, in order to make higher quality information appear on the top of the results.

Whilst the concept of authority ads a valuable parameter to the algorithm, it is established by seizing the hyperlink structure of the World Wide Web, wherein the
WebPages are interconnected: if a page refers to another the cited page is considered authoritative, because it is supposed that behind the reference there is a motivation of quality. The structure of the WWW made up of interconnected links constitutes the point of departure for understanding and using it effectively, according to Kleinberg (1998). In fact, all the hyperlinks form a network which cannot be ignored, this structure contains several information created by humans, which build the authority of a source.

Indeed, the main characteristic of Kleinberg's algorithm, HITS (Hypertext Induced Topic Selection), is its focus on hubs (Kleinberg, 1998), which refers to a page leading to other pages, as a sort of landing page, while the authority is transmitted by being referred from different hubs.

Hubs and authorities exhibit what could be called a mutually reinforcing relationship: a good hub is a page that points to many good authorities; a good authority is a page that is pointed to by many good hubs (Kleinberg, 1998, pg. 8).

This recursive system needs to be broken, in order to distinguish between authorities and hubs. The introduction of the concept of hubs (Chakrabarti et al., 1999) brings to the partition of links in categories, for instance within the topic of televisions there are several hubs, which means that there are several groups of highly dense links populated region, each region may assume a specific subcategory named according to the subject: televisions on sale, new television models etc. Even in this case, the problem of the reliability of links appeared, in fact, it was foreseen that not all the links can be considered as a synonym of quality. Then, it was hypothesized that a link connected to another page within the same Website might have the purpose of enhancing usability, or that it is not always true that most authoritative pages contain the keywords typed in the query. To obviate to this problem, some solutions were adopted, such as the exclusion of links from the same domain name.

HITS was also updated to improve the performance, for instance Clever is a subsequent version, but the main differences with PageRank by Brin and Page (1998)
remained stable. In fact, Google's algorithm transfers the authority from a page to another without considering the hubs and the process of retrieving is considerably different. Indeed, Google gives back documents query-related by scanning the indexed pages, while HITS performs the text search at first and afterwards weights the results in a small retrieved group (Kleinberg, 1998). The approach proposed by Kleinberg goes beyond the debate on the authority, in fact, it stresses the relationship between hubs and authorities, representing an important deepening of the quality of the SERP.
To conclude, HITS' aims at ranking results according to the query, while PageRank strives to provide a rank for all the WebPages, independently from the query.

1.3. PageRank and Google

PageRank is the main component of Google’s algorithm, at least until few years ago: it has been defined as a system judging the probability according to which a potential user may incur in a certain Webpage, adjusted to the damping factor, e. g. the user changes page. The pages are considered randomly, but it is interesting that the damping factor changes according to the page, in order to discourage algorithm's manipulation.
The major characteristics of PageRanks described by the founders at the early development stages of the algorithm are listed below (Brin et al., 1999).

♥ The more the back links the higher the position on the SERP, but these links are counted according not only to the quantity, but also to the quality: this means that potentially a Webpage with fewer number of in links might come first if it has more quality links: this system is a recursive one.
♥ PageRank does not consider the content, but the structure of the Web and how the pages are allocated.
Problem of the rank sink: when two pages have links to one another, but not to other pages, the algorithm can be trapped. That is why the damping factor (or E vector) was introduced.

"Random walker" the algorithm has been compared to a person randomly clicking to links on the Web, and when he is snared, he feels bored and clicks on something else: the vector E refers to this system, which technically consists of the probability of other pages to be found by the anonymous user. The pages encountered by the vector E represent a sort of neutral indexing, but this approach shows the problem of increasing highly the value of some pages, with an elevated number of in links.

Basically, PageRank works by recursively propagating weights through the link structure of the web (Brin & Page, 1998). PageRank algorithm developed from Black Rhub, created by Page himself, whose goal was to retrieve information by giving importance to the citations. In fact, Page wanted to find out which are the links pointing to a certain page, abstracting the concept of importance of citations from bibliometrics, the science dealing with informational retrieval, in the sphere of academic documents. Not surprisingly, the importance of citations was already in the mind of researchers, as pointed out earlier.

However, the further intuition of Page was related to the different weight attribution to back links (incoming links of a page). Thanks to the help of Brin, the math mind of the team, Page achieved his desire to find a system to consider not only the number of the links, but also their sources, and establishing penalties for low quality Websites.

Pagerank idea existed already in 1997, but it came into force effectively in 2000. This represents absolutely the first attempt to better off search engines: it might seem odd, but in 1997 there was no intention of improving the search on the Internet (Lastowka, 2007), due to the advertising purposes of the latter. Indeed, at that time, it was thought that search engines would have given access to contents, specifically information and entertainment, by reaching agreement with other media and gaining money from the
advertisements (Lastowka, 2007). Therefore, the main goal was to attract and to make people exploring the pages and not to provide quality contents.

On the contrary, Bring and Page did not sustain this revenues focused goal because reluctant to advertising. As a demonstration of this, the austere Google's homepage was chosen.

Subsequently, the approach changed, because of the need of financial resources, they created the distinct search engine Go.To.com: a search engine just providing advertisements, but predictably it was not a great success. Therefore, the system was incorporated on Google itself, by distinguishing between paid and organic search.

It seems opportune to underline the problem arising from the Google's algorithm and its ever-increasing popularity.

First of all, in order to check the accuracy and the quality of the information, a mathematic formula is applied, which is a fallen system per se. Thus, the manipulation of the algorithm is a viable menace: in principle, anyone can be highly ranked on Google if aware of the importance of some factors. Despite the secretiveness of Google's algorithm to the public, some principles have been disclosed either by Google itself or by powerful webmasters, so that the potential harmfulness of the mean must be considered.

However, Google is attempting to discourage poor Websites to rank higher by pointing out the punishments for those who will infringe the conduct rules; it has also introduced spam filters, which still are far from being perfect. Therefore, to accomplish this objective Google has also created a Webmasters' guidelines to foster the disclosure of relevant, authentic and quality information. The opened democracy proposed by Google represents a double edged sword and the biggest weaknesses to overcome: by giving the possibility to be visible to everybody, the unfair behavior is strenuous to retain.

Another limitation of Google is that it retrieves just what it has indexed (Wiggins, 2003), relying on just what Google suggests represents a foreclosure of the information.
Furthermore, the monopoly issue must be stressed. In fact, Google has been accused of antitrust behavior several times and investigated by the Federal Trade Commission (Hazan, 2013) for anticompetitive conduct, because it seemed that Google was favoring its own services on the SERP, granting them a better position. This behavior is deemed to violate the antitrust law, because the search engine in question, occupies a position of monopoly, from which it might hinder the mechanism of the free market by restricting the competition. If Google had not been in such position of superior power, in respect with its rivals, its practices would have been considered completely lawful.

Table 1. Search engine global market's share (Clicky, 2014).

1.4. Search Engine Optimization (SEO) techniques

SEO is a discipline dealing with the visibility of Websites, within the results provided by the search engines, in order to increase the number of visitors and the popularity, the so called authority of a specific Website or page (Quoniam, 2012). In the last
years, there have been a growing number of companies defying themselves as providers of optimization services, which is an interesting factor, because of the alleged secretiveness of the algorithms used by search engines to rank the data. Notwithstanding, this controversial issue, those companies are pursuing legal activities, even though sometimes they have been accused of unethical practices.

For discerning between lawful and unlawful conducts, the distinction between White hat and Black hat SEO has been adopted: White hat SEO conducts are solely complying with the search engine recommendations. Websites ranking higher are often considered to be involved in bad practices, but experiments falsify this belief (e.g. Quoniam, 2012), in fact, research show that it is possible to reach the first page of results through the before mentioned recommendation. On the contrary, Black hat SEO refers to the deception of search engine algorithm, one of the bad practices, for example the so called Google bombing phenomena. Studies reveal that Google is still weak and improvements are needed, in fact, it has been argued (Bar-Ilan, 2007) that even though Google's algorithm is secret, the ways to be ranked high has been discovered. Indeed, there have been numerous attempts to explain how a search engine processes and classifies the data in the SERP. Since the secrecy and the updating of the algorithm, in order to prevent manipulation, it is not possible to state an ultimate formula.

However, a large and growing body of literature has investigated the functioning of algorithms to understand whether is possible to influence the results, as claimed by many Webmasters, or not. The confirmation of this hypothesis would justify the increasing number of online marketing companies, dealing also with Search Engine Optimization techniques.

Collectively, these studies outline the critical role of ranking factors, because it has been suggested that it is possible to influence ranking according to some tactics. Therefore, the most powerful Websites become the most visible; in the end just the companies which can afford to pay for SEO will be prominent (Bar-Ilan, 2007).

In spite of the difficulty of guaranteeing the effectiveness of these techniques, it is possible to analyze ranking factors, also without looking into millions of search results.
Unveiling the perception of blogs as marketing tools: a randomized controlled experiment

(Evans, 2007). In fact, what we know about algorithms is largely based upon empirical studies that investigate how to ascend the list of results.

By observing the literature, the ranking factors can be divided into two broad types, one related to the query indicators, such as density and frequency of keywords, the other comprising factors which are independent from the query, such as the links within the Website or the external links leading to the Website (Evans, 2007).

The evidences presented in this section suggest that influencing the ranking system is probable, by applying more than one technique. Turning now to the experimental results, the most salient factors are presented.

- **Number of pages of the Website**: in principle, the bigger the Website, the more the likelihood to be ranked first. Evidences do not confirm this factor, the quality is preferred over the quantity (Evans, 2007).

- **In links and PageRank**: the number of in-links, incoming links, is deemed to increase the visibility of a Website (Evans, 2007). PageRank, as previously stated, is a particular parameter of the algorithm introduced by Google, with the aim of improving the quality of the search results.

  In fact, according to the PageRank system, a Webpage appears on the top of the results if it is cited in quality Websites, so the authority-popularity of a Website depends not only on the amount of the in links, but also on the quality of the latter (Bar-Ilan, 2007). These studies taken together suggest that there is an association between visibility and incoming links, for example the enquiry of Evans (2007) identifies that the popularity of a Website is directly proportional to the number of in links, but points out that other elements can nullify the positive correlation.

- **Age of the domain name**: the more a Website is older, the more the probability to be ranked higher, this factor would harm newly constituted sites. In spite of the claims of the importance of the age of the domain name, some findings
(Evans, 2007) raise doubts about its truthfulness, owing to the difficulty to measure it. Actually, more than the age of the domain name, the authority of the name is valued, such as domain names of an accredited university, a government institution or a public foundation would be considered more relevant (Dawson & Hamilton, 2006).

- **Being listed in DMOZ.org**: DMOZ.org is an Open Directory created by people, whose information are indexed by other browsers, e.g. Yahoo! and Google. Studies claim the importance of this factor in influencing SERP, but being included in the directory takes from 6 months to one year (Evans, 2007).

- **Appearing in the social bookmarking De.li.cious**: it has been argued that being in a social bookmarking would increase Website popularity, but contrary evidences were found (Evans, 2007).

- **Anchor text**: this refers to the text of the hyperlink, which is clicked by the user. For example, the user within a Web page content can click the words **ricciarelli Siena**, which is connected to the page

  <a href="http://www.fornoitalia.com">ricciarelli-siena</a>

  Briefly, the anchor text is the clickable part of a text connected to a specific URL, so that by clicking on it the user is led to another page. This element may represents an incentive to be ranked higher, in fact, it has been recorded (Bar-Ilan, 2007) that is a good practice to create the anchor text according to the query supposed to be typed by the user.

- **Meta tags**: Meta tags refer to meta data, literally "data about data", which are information about the content or the structure of the Website. In some cases, Meta tags are not visible to the users, as in the case of the instructions referring to the layout of the page, but contribute to the quality of Websites perceived by the search engine. Meta tags should contain the keywords of the Web page, it
has been reported that quality Meta tags allow a high position in the SERP (Dawson & Hamilton, 2006).

A fundamental Meta tags is the title, which deserves peculiar attention. In fact, it is advisable to include in the Meta title, the relevant keywords or the name of the company and it should not exceed the 70 characters, in the case of Google results, because this is the maximum numbers of characters displayed. The relevance of the title has been well documented (Dawson & Hamilton, 2006), furthermore it is the first thing that the user sees, so that it represents a sort of calling card and the default name in the bookmarks.

Figure 4. **Example of the visible title in the SERP of Google.**

- **Simple design**: the easier the design of the Website, the higher the chances to be intercepted by the spiders of the search engine. In fact, crawlers interpret frames, animations, log in requests, as an obstacle, so that it has been highlighted the importance of minimizing them (Dawson & Hamilton, 2006).

- **Cascade Style Sheet (CSS)**: this language works with the HTML and it has been created in order to understand clearly which attributes of the Webpage refer to the content and which to the structure. The simplification of the information obtained through this kind of language has been found in line with a mounting visibility (Dawson & Hamilton, 2006).
Words and nouns: it has been stressed (Dawson & Hamilton, 2006) that words and nouns, over other part of the text, are more important to boost popularity.

Keywords: using good keywords and repeating them has been considered as a good way to gain popularity. Moreover the right keywords present the advantage of enhancing user's usability.

1.5. Speculating about quality

The doubts about the quality of the results retrieved by search engines arise because potentially all kind of contents can be highly ranked. Despite the importance of the issue, an unambiguous framework for evaluating search engines' quality cannot be established, indeed the debate is opened. Defining the quality is not a simple matter, because it depends on the perspective from whom quality is concerned, for instance in the academic realm, quality is associated with credibility, which in turn refers to peer review sources (O'Leary, 2010). However, it is interesting to present the perspectives according to which quality has been evaluated in the past literature.

Quality might be seen as equivalent of credible, credibility has been evaluated by relying on the correspondence of the query, typed by the user, and the SERP (Weerkamp & de Rijke, 2012). So, if the user finds the information he was looking for, the Website is positively judged. Relevant studies have been conducted in this respect, even because of the simplicity to carry out this kind of experiments.

A step forward is exemplified by the work undertaken by Lewandowski (2008), who considers the relationship between the query and the accuracy of the Meta descriptions, from the users' perspective, in fact, this study, unlike the previous, highlights the experience of the users. Indeed, opinions add value to the assessment of quality, although this leads to a major bias on the users' preference. The study reveals that the growing importance of Google as search engine may be due to its better accuracy in the description tag (Fig. 4) compared to others search engines.
There have been several investigations in the perception of quality from the user viewpoint, but the results are not encouraging: people know a limited number of search engines and they tend to use always the same, so that the popularity of one search engine could lead to its preference (Lewandowski, 2008).

Quality may be associated to the **relevancy**, in the meaning proposed by Google, but it has been questioned that relevancy is not a synonym of authenticity and honesty (Wiggins, 2003). On the contrary, users perceive the first results as more trustworthy, in respect to the last. As regards the honesty concept, this might be explored more in depth, because not only the honesty of Web contents may be questioned, but also that of search engines. Indeed, in the case of Google, it has been asserted (Hazan, 2013) that the algorithm is just one side of the ranking system, the other side is made up of editorial decisions, leading to biased searches, because compromised by the human intervention. In this regard, the potential harmful position of monopoly of Google reappears.

The literature reports, also, some possible shields to defeat against the influence of the SERP or to avoid to be obfuscated by Google's power. One of the solutions proposed is aimed at fostering user's exploration (Singh et al., 2013): it attempts to provide new ways to display information on the results page, such as information grouped in clusters or page previews, it holds the worthy advantage of engaging users in the exploration of the results, instead of sticking on the ready available information. Notwithstanding, the appreciable effort, the proposed displaying method could lead to other kind of biases and influences.

Another option is to use more than one search engine, in fact, not only the number of information stored in the search engine's database is important, but also the algorithm sustained brings different results (Smith, 2003).

A further alternative proposed is to rely only on certain authoritative sources of information, such as bibliographies or to analyze the intention of the author, his objectives and the timeliness (O'Leary, 2010), but a considerable amount of time is required: this method does not suit the need of rapid information which people are
experiencing today. Another viable solution, for the improvement of the search on the Internet, is embodied by the Web search engine Sweet Search [www.sweetsearch.com], whose index derives from Finding Dulcinea [www.findingdulcinea.com], the so called Librarian of the Internet. The results given by Sweet Search are composed of texts which have been judged by experts (O'Leary, 2010). Even this reliable search engine presents its drawbacks, in fact, the number of contents indexed is limited.

To conclude, it looms valuable to attempt to detect the best way to explore the Internet, but the path towards a finite solution is still long.
2. Blogs

The web search results of Google, as well as those of other search engines, are populated of thousands of blogs whatever the query inserted, and they are increasingly occupying the first positions of the SERP. Therefore their prominence should be contemplated: people are so exposed to blogs, that they do not even realize their presence. Indeed, nowadays it has become common looking for information in blogs that nobody notices the difference between a Website, which is supposed to be more authoritative, and a blog. Considering that people strongly rely on blogs, because of their placements, the discussion on blogs' influence and impact acquires importance, in fact, the faithfully of the latter is controversial. Often blogs have been established with promotional purposes, which, in turn, are not clearly manifested, because this relatively new marketing tool does not intend to be invasive and perceived as advertising.

The novelty of this tool is relative, because, actually, their importance was already predicted in the early 2000s, when some teachers started a blog, believing in their high potential to spread the knowledge.

Today, blogs have been seen as a modern form of Word of mouth, following the principle that consumers are led to value more people's opinions than the advertised announcements of companies: indeed, advertisements show only the positive side of products, while consumers' reviews present positive and negative elements (Chih-Chien Wang & Hung-Yu Chien, 2012). As mentioned in the introduction, consumers refer also to blogs before buying something or going to a restaurant and this tendency is upward.

The chapter is structured as follows firstly a definition of blog is reported, secondly its roles are portrayed, thirdly a state of the blogosphere is depicted and in the end a literature review on blogs' influence is conveyed.
2.1. Definition of blog

In the past, the blog was referred also as Web blog, but today the term is almost disappeared. The blog is part of the Social media (Svatosová, 2012), even though in the last years the notion has been associated exclusively with social networks due to their popularity. In contrast, Social media represents a broader category, including for example bookmarks or videos sharing platforms, they refer to all the online instruments allowing interactions between people. Those are contributing to the evolution of our society and are testifying a social change by permitting to create networks and identities very easily and to penetrate the lives of people (Wilcox-Ugurlu, 2011). Subsequently, companies by not to subscribe to them would lose a bulk opportunity of growth (Svatosová, 2012). Social media are also noticeable because they are the symbol of a different paradigm, in which the consumer is empowered with the control of the communication (Hoffman & Fodor, 2010).

Turning now to the conceptual definition, the blog is a sort of online personal journey, where a person, or a group of people or a company, write public messages, called posts, addressed to an audience, who can leave comments and interact with the blogger/s. The contents respect a chronological order and often are divided into categories to simplify their exploration.

Blogs are increasingly similar to generic Websites, so that it is important to stress their differences. The main one is related to the Web 2.0eness, indeed, the blog is a typical expression of the Web 2.0, by embodying a two way communication: readers can share their opinions, and the distinction between author and reader disappears. A blog needs, also, to be updated more often, this attribute has been referred as timeliness (Gill, 2004), and it is easier to implement, no technical competences are required. In addition, the language style adopted by a blog is informal and usually subjective (Filimon, 2010; Jeanne, 2013), while a Website requires a certain degree of formality. Moreover, a blog contains links pointing to sources or to other blogs, while this feature is not binding for other Websites (Gill, 2004).
The first blog dates back to 1992, but blogs’ boom started around 1994 and their prorogation grew rapidly in the early 2000s (Wyld, 2008). Their success is deemed to be in the perfect compliance with the actual society, which is permeated by the need of participation: the urgency to be visible and recognizable which has been defined as the "you phenomenon" (Wyld, 2008, pg. 448). Furthermore, the last years have seen the rise of service providers allowing the possibility to start a blog for free and without any specific ability.

A blog exists and thrives if it receives attention from the public, in fact, its main goal is to generate buzz and gain visibility. Several studies have attempted to determine the features to create a successful blog and the relative techniques to boost its popularity: a resume is presented.

The importance of the quality contents has been stressed (Filimon, 2010; Collier, 2011), "content is the king", the famous Bill Gates' sentence fits perfectly. Moreover, attention should be paid to the way according to which posts are written, because they should not be created just to comply with search engine and to increase the ranking, but they should result interesting and relevant to potential readers (Collier, 2011). At the same time, search engine's algorithm should be kept in mind (Jeanne, 2013), in fact, optimization criteria are valid, such as keywords, metadata... Optimizing the contents of the blog according to the relevant keywords, by including them in the Website's title and in the Meta description reveals fundamental (Filimon, 2010), also including within the posts hyperlinks connected to highly authoritative Websites, such as governmental bodies or encyclopedias is important. In addition, it has been observed (Jeanne, 2013) that the number of in links affects popularity. Indeed, creating posts on similar blogs and referring the URL of its own blog is a practice assumed to bring brilliant outcomes. Despite, Google's condemnation of link exchange practice through penalties, inserting links from other Websites hoping to have a link in turn has been suggested (Filimon, 2010). Moreover, a blog can be widespread thanks to the videos to be posted in relative platforms (e.g. Youtube) or through social networks (e.g. Facebook, Twitter).
A blog should be perceived **valuable** from the readers’ side, so that some resources should be added, such as e-books, papers (Filimon, 2010). The blogger should also transmit honesty of intentions (Collier, 2011). A blog needs to be **updated** and so it is highly recommended not to create a blog if there is no time at disposal (Collier, 2011; Jeanne, 2013). To conclude, a blog should be also improved by using analysis tools, such as Google analytics, which provides information about the traffic and the actions of users, necessary to understand its possible lacks (Filimon, 2010).

As regards the effectiveness of blogs as marketing tool, there have been many attempts to determine their return on investment (ROI). According to Hoffman & Fodor (2010) and Chaffey (2009), simply calculating the ROI (a proportion between costs and revenues) would diminish the value of a blog and only the short term gains would have been taken into account. On the contrary, other non monetary values should be stressed, such as brand awareness. In fact, in order to estimate the value of a blog as marketing tool, consumers' perspective should be considered, which might be accessed through the **4c's** (Hoffman & Fodor, 2010): connections, creation, control and consumption of a blog.

### 2.2. A single word, several purposes

The word blog might refer to completely different platforms, according to their purposes two large categories may be distinguished (Fig. 5).

- **Blog as a social tool**: the original function of the blog was related to its capacity to gather people and make them interact with each others. Indeed, the blog was born as a social platform; nowadays this pure typology is almost untraceable. Within this category some major sub-groups are recognizable (Wyld, 2008): a blogger can write for self therapy objectives, because writing help to express feelings or he can write for popularity's sake or for opinions' sharing; in

Francesca Bandini
addition, a blog can be used within a firm to connect, to motivate or influence employees or to spread the knowledge. Another group refers to the phenomenon called *participatory journalism*, studies (e.g. Gill, 2004) identify the use of blog as a substitute of traditional information Websites or newspaper.

- **Blog as a marketing tool**: the blog might serve as instrument to promote the services or products of a company; usually it is referred as corporate blog, meaning a company establishing it within its Website. Another purpose is related to the study of the responses of consumers to products, in fact, often in a blog people make comments and probe for questions, from the emotions and information emerged a company can adjust its strategy (Consoli, 2009) Sometimes blogs are used to boost product's popularity or providing product reviews.

![Figure 5. Classification of blogs according to the objectives.](image)
Despite the attempted classification, the borders between the categories are not completely definite. It might happen that, even though a blog was not created with the intention to earn money through advertisement, a change in the policy occur. Indeed, it is increasingly tempting and easy to sponsor products or renting a space within a Website: Google (namely Google AdSense) and Amazon are also providing this kind of opportunity by placing automatically (under consensus) banners promoting something in line with the contents (Filimon, 2010), incredibly also open source software to create blogs or Websites are providing with this possibility; the phenomenon belong to the contextual advertising and is revised in the next chapter.

Since the interest of this work on detecting negative blogs' influence, the category of blogs as marketing tools will be taken into account, in fact, the potential damages for consumers and the society as a whole, are represented by those blogs with not declared commercial aims. Before analyzing this category, it is important to stress some notions about corporate blogs.

Several studies have stressed the different motivations behind a blog in business (e.g. Filimon, 2010 and Wei-Li & Yi-Chih Lee Wu, 2012). Primarily, blogs may be seen by professionals as an inexpensive tool to reach customers, in fact, it has been claimed that is less expensive than being on the yellow pages (Jeanne, 2013). The corporate blogging phenomenon may hold different purposes within the strategy of a company; for example it might constitute the less formal side of a corporate Website, throughout which the firm acquires a certain degree of humanity and enter people's lives. Furthermore, by catching the reactions and feedbacks of customers, it gives insights for further development; other benefits are inherent to the simplification of the communication (Filimon, 2010) and the prompted engagement with customers (Wyld, 2008).

Another interesting aspect is related to the role connected to public relations that a blog may serve, in fact, it might be used by the company to launch news about its ongoing projects (Filimon, 2010) and chief executives might communicate directly with customers (Wyld, 2008). It might be depleted, also, for SEO purposes, in fact, to
be ranked high, it is advisable to have a blog with quality contents, because the blog can be updated very often and easily, these elements matter for Google ranking. Secondly, blogs have been seen as an opportunity to trigger word of mouth, prior the diffusion of other well known social networks platforms. In fact, the effect of word of mouth is widely known; indeed it outlines the most effective marketing tool. These promotional blogs may acquire two different aspects, in fact, in one case they are blogs directly linkable with the company, proposing the marketing activity, in the other cases, blogs are not associable with a company, but they contain advertisements. The latter fulfils the role of "disguised word of mouth" (Magnini, 2011, pg. 243), harmful for consumers, who, by not being aware of the hidden intents, may interpret the contents as genuine word of mouth. The issue has received considerable critical attention, but the studies are still limited in number. The main forms of disguised word of mouth are the following.

♦ The company leaves comments on others' blogs, recommending the use of its own products or services and the origin of the comments is not directly associable to the company itself. The people posting comments are paid and have been referred as brand pushers (Magnini, 2011), this tool has been documented also on traditional word of mouth.

♦ The company pays celebrities to act as secreted testimonials and the audience is not aware of their roles.

♦ Blog advertising refers to the promotion of specific products or services throughout a blog, in this case, the blogger clearly promotes the specific brand, for instance by posting pictures with visible labels. This is an affiliate marketing technique, where affiliate indicates someone who has a commission based agreement with the company to perform a certain promotional activity (Chaffey, 2009).

♦ The company constitutes a supporter blog on purpose to indirectly suggest the use of its products or services, trying to influence potential consumers' decisions on a subtle way.
For clarity's sake, it should be stressed that the latter form of disguised word of mouth is the core of the empirical analyses reported in the fourth chapter, but no evidences of this kind of tool has not been found in the literature; in fact all the notions connected to the concept derive from the internship period at Suerte.

2.3. State of the blogosphere

Since blogs might really condition people's lives by giving information and given the potential magnitude of the audience reached, their influence should be assessed (Gill, 2004). Many different companies are trying to acknowledge the influence of media through market research, in order to choose the better investment in terms of advertising. It is therefore interesting to report some findings to testify the importance of the blog issue. According to Technorati (2013), an authority in the field of media influence, worldwide consumer's media consumption (Table 2) is above all concentrated on YouTube and Facebook, although the position occupied by blogs is relevant.
In respect to the online purchase, data show that consumers look at retail Websites in prevalence (56%), the brand site is in the second place with 31% of preferences and the blogs are in the third place with 34% (Technorati, 2013). Therefore, while purchasing consumers do not attribute a significant major authority to corporate sites in comparison with blogs. This assumption is confirmed by the figure on the influence, where blogs result to have weight for 31.1% of the audience, this percentage is quite akin to that of corporate Website (34%), however the most influent platform is constituted by retail sites (56%). As regard to trust, blogs occupies a similar position.

The results of the surveys conducted in the Italian market seem to be in line with the aforementioned trends, in fact, on a sample of 1000 Italians (fullPlan, 2013), 49% of the people relies on Social media (e.g. blogs, forums, social networks) before the
purchase, but it is interesting to notice that at the same time 55% refer to the corporate Websites.

Furthermore, it is also worth noting the survey conducted by Imageware (2013) on 125 of most influent bloggers in Italy. Concerns over the quality and trustworthiness of the contents seem to find a justification, in fact, 64% of the bloggers report to take the information from the Internet, without specifying the source, 18.4 % admit also to find the information on other blogs. The major part (84%) retrieve information from the personal experience or ideas, only 40.8% look actively for information on authoritative newspaper or reports, while 53% wait for news which are sent by the press or companies. As regards the role fulfilled by the blogs expressed by the owners (Tab. 3) only 16% disclose to use the blog as a marketing tool and within the sample 17.6% are corporate blogs, so that it might be assumed that the same people consider the blog as a marketing tool.

![Motivations behind blogs](image)

Table 3. **Motivations behind blogs (Imageware, 2013, pg. 9).**

The primary reason expressed (Tab. 3) is connected to the desire of self expression (52%). The credibility of this motivation is hampered by another question, in which 30% of the bloggers claim openly to be interested in earning money and 75% wish to transform blogging into a profession.
Furthermore, 51.2% of the bloggers say to gain revenues from the activity. As regards the composition of those, the major part comes from advertising (65%), but a relevant part is represented by the links granted to other Websites (31.7%) confirming the existence of such practice. Interesting, bloggers believe to enjoy major or equal authority in respect with traditional means of communication.

Major evidences of the undeniable presence of blogs in Italians lives are found, by looking at the most visited social portals (Nielsen, 2014): blogs provided by open source software appear in the first positions. The following data refer to the unique visitors received by those social media in a month period (February 2014):

I. Facebook: 18,801 visitors
II. Blogger: 9,814 visitors
III. Wordpress.com: 4,917 visitors

As it can be spotted the first position is occupied by Facebook, but the second and the third most visited social portals are Blogger and Wordpress.com, which are free services allowing the possibility to create a blog.

2.4 Literature review on blogs’ influence

The purpose of this paragraph is to report the literature on blogs’ influence. It begins by explaining the different parameters used to evaluate the influence, followed by the variables measured and the methods involved; in the end the conclusions are drawn. Overall, there seems to be agreement on the assumption that blogs do influence people. As a rebuttal to this point, it might be argued that sometimes their power has been exaggerated (Murphy, 2007), due to their high rate of mortality, especially nowadays that is easy to start one. It has been claimed that just few bloggers have a real impact on people's opinions, but this does not diminish the importance of the argument, because even though just a few have the capacity to condition, they might occupy the first positions of the SERP. Another critical issue is related to the
connection of blogs to other influential Websites which implies the augmentation of their weight, despite having a small amount of readers.

The influence of bloggers on people has been investigated according to several angles. Some studies have examined the role of blogs in the political sector, which are seen as the major source of information from political affiliates (Johnson et al., 2007). Within political blogs, Kien-Weng Tani et al. (2011) consider the influence as a synonym of referrals, in fact, a person is judged influential if she reports the ideas read, so there is a valuable action which might suggest the bias.

Other studies, such as that of Wei-Li & Yi-Chih Lee Wu (2012), analyze the impact of bloggers on consumers' decision making process. Similarly, Wilcox-Ugurlu (2011) provides an in-depth analysis on the sociality values which affects consumers' decision making process within the social media.

Another thread focuses on corporate blogs and their effects on customer engagement (Ahuja & Medury, 2010). Chih-Chien Wang & Hung-Yu Chien (2012) assess the various consumers' attitude towards product reviews on blogs, by measuring the skepticism belonging to the presence of advertisements.

A further research (Fuertes-callén et al., 2010) explores the revenues of an e-commerce shop, in order to determine whether blogs' traffic has a positive impact on sales.

A small scale of studies outline the importance of on line word of mouth (WOM), for instance Ramaprasad, (2009) observes the influence of blogs on consumption in the music industry. Along the same path, Finin (2006) reports the topics' spread in blogs together with the influence of sblogs (spam blogs) on certain ranking system, interestingly a model to analyze the knock on effect triggered is submitted.

Other investigations relate to the journalistic side of blogs, Gill (2004) attempts to detect the influence within and outside the blogosphere and Murphy (2007) rises concerns about the superior authority of bloggers over journalists, admitting the difficulty of evaluating blogs influence on news.
As may be noticed from this introductory part, not only the word *influence* adopts different meanings, but it can impact on diverse things, so that the issue is rather intricate.

It is interesting to mention, also, the variables measured and the point of views assumed, in order to depict the complexity of the issue. As regards the blogs with political purposes, the measurement of Kien-Weng Tani et al. (2011) refers to the prorogation of ideas, through a content analysis aimed at detecting emotions and common ideas. Therefore, the influence shows off if people demonstrate agreement with bloggers’ ideas. Whereas Johnson et al. (2007) focus on the alleged credibility of blogs from users and non users’ perspectives, in relation with the traditional means of information, the study provides a comprehensive framework by considering the motivations, the frequency of use, the credibility judged by the users, the level of political involvement, the demographic characteristics of participants.

In respect to more commercial blogs, Chih-Chien Wang & Hung-Yu Chien (2012) evaluate consumers’ skepticism connected to both advertisements and the blog’s product reviews. The concept of skepticism is seen on the light of message source, product involvement and previous product experience of the respondents. Those variables were supposed to show an impact on product attitude, purchase intention, and recommendations. Even Chin-Lung Hsu et al. (2013) investigate the influence of blogs through advice on shopping online to see if bloggers recommendations are trusted. Similarly to product reviews, Wilcox-Ugurlu (2011) bases his study on the consumption experience of clients and its impact on the business: consumption experience refers to the sharing of the experience through social media, a sort of *potentiated* product's review. The adjective *potentiated* has been used because of the psychological implication of the latter, in fact, the experience shapes the self-construction through Social media, an example of this phenomenon is the so called *food pornography*, which entails the sharing of pictures of food on social media. Only slightly different is the approach of Wei-Li & Yi-Chih Lee Wu (2012), who emphasize three dimensions, blog’s trustworthiness, product attitude and blog involvement, on purchase intention. Likewise, Ramaprasad (2009) argues that blog's influence is

Francesca Bandini
evident from purchases and interactions. A more scientific attitude is undertaken by Fuertes-callén et al. (2010), who connect the Website traffic to the augmentation of sales, by looking at blogs' popularity, link popularity and position on the SERP.

It is also worth noting the variety of methods used to track the influence before proceedings in the analysis of the results. The official system is based on weighting the links received by a blog through a mathematical formula, this has been strongly criticized by Gill (2004), because older blogs would be considered more influential thanks to the more links. However, in the literature quantitative methods have been applied along with qualitative ones. As regards the qualitative methods, Finin (2006) employs the PageRank algorithm to understand the path of influencing topic. The study carried out by Fuertes-callén et al. (2010), illustrates a statistical analysis of the financial statement, merely sales and profit, to determine the connections among blogs, Website's visitors and sales. Generally speaking, since the complexity of the issue qualitative methods have been preferred, Johnson et al. (2007) use an online survey filled by a sample of people with interest on the political scenario. Alike, Wei-Li & Yi-Chih Lee Wu (2012) and Chih-Chien Wang & Hung-Yu Chien (2012) investigate the motivations, reactions and behavior of consumers by means of online questionnaires. Sometimes mix methods are applied, in fact, Ahuja & Medury (2010) not only perform observations, but analyze also Website's traffic. In the same way, Kien-Weng Tani et al. (2011) check the supposed impact through software for the quantitative blog features (by considering links, authority ranks, number of copied sentences) and observations in respect to the sentiments sphere. Also Wilcox-Ugurlu (2011) uses several methods: non-participant observations, participant observations and interviews.

Together these studies bring great insight into the issue of blogs' influence and in some cases suggestions for companies, which cannot ignore the phenomenon, also because of the switch in the communication paradigm. In fact, the communication is
increasingly users' generated, a shift from companies to consumers have been reported (Wilcox-Ugurlu, 2011).

The conclusions of all these studies suggest that companies should be aware of certain factors when they establish an e-commerce, because blog's popularity is correlated to the traffic of a Website which translates into more sales, also search engine relevance has influence on the number of visitors and link popularity is directly proportionate to the traffic (Fuertes-callén et al., 2010), as intuitively can be foreseen. In fact, Fuertes-callén et al. (2010) and Ramaprasad (2009) found a positive correlation between blogs and sales. Moreover, in order to be effective on customers' engagement a blog should be updated frequently, indeed, the number of posts is positively linked to the influence on customers (Ahuja & Medury, 2010). As underpinned by the experiment of Chih-Chien Wang & Hung-Yu Chien (2012), consumers' skepticism is higher in the case of advertisements than in that of blogs, regardless the authority of bloggers. Although, the level of skepticism towards advertisement was lower in the case of highly involved consumers. Surprisingly, consumers are not statistically significant more influenced by expert bloggers than by non-expert bloggers.

Wilcox-Ugurlu (2011) infers the tendency of people to show only the positive sides of things, when involved in online conversations, this happens in order to receive confirmation and please the audience. Therefore, companies should not be too concerned about the possible negative comments, in fact, the behavior in Social media is different from that of the real life: the more people are involved with the media, the more they are led to the inversion of the reality, which refers to the exaggerated importance given to WWW generated reality than to real life itself. These results are not applicable to occasional users.

As regards credibility, Johnson et al. (2007) argue that blogs are not considered fair within the political realm and the perceived credibility depends on people reliance on the media. In fact, blog readers consider them more credible, while non users judge them differently. The same is valid for motivation, people who are using blogs for informational purposes see them more credible in respect with those using the media for entertainment. Furthermore, the study stresses the attitude of users to address to
blogs especially for in dept information, when the other sources are providing scarce
details. People commonly relying on blogs, judge them more credible also over
traditional and online means of communication.
Mixed results are reported by Wei-Li & Yi-Chih Lee Wu (2012), indeed, the study
sustains that the trustworthiness of the blogs is not proportionate to the purchase
intention, but the latter is affected by the involvement, because consumers are brought
to probe for information. Also the interactions between variables have been explored
revealing that low level trustworthiness combined with high level involvement prompt
the purchase intention, in addition the involvement might change consumers' attitude
towards the product, so it holds effects on the intentions.
The finding of Chin-Lung Hsu et al. (2013) reveals that the influence of a blogger
depends on his reputation, who enjoys a high reputation impacts also the attitude and
the intention to shop online. When the reputation is high, the recommendations are
seen automatically more useful, and in this case the trust affects directly attitude and
intention. In the case of people attributing less authority to the blogger the trust affects
the alleged usefulness, but does not influence attitude and intention.

All the authors infer the difficulty of studying blogs' influence, but their results might
suggest that often the strength of the impact is related to the emotional side of the
brain. In fact, the power of emotions and its effects on consumers have been
highlighted along the history, there is even a part of the marketing defined emotional
marketing, which demonstrates that the emotions influence all the decision making
process. Emotional marketing aims at stirring up emotions and understanding the
psychological needs of consumers: the value of products dwells more with the
associations, the symbols evoked, than to the product itself, this is evident from the
importance given to the 5 senses marketing by luxury brands. The 5 senses marketing
tries to stimulate all the senses of customers to deliver a unique experience, for
instance Tiffany & Co. has its own fragrance to spray in the shops.
Emotions may be defined as a particular state arising because of a certain stimulus,
which often distorts the cognitive consciousness, creating disorder, distraction...

Francesca Bandini
Emotions are acquiring incremental importance due to the saturation of the market, the consumers do not see just the characteristics of the product, but they look for other intangible assets: dreams, feelings... (Consoli, 2009). In fact, the trend is towards the satisfaction of both functional and emotional needs. The weight of emotions and feelings has been enquired for a long time by the psychology, well known the experiments conducted by Harlow with monkeys showing the preference of those animals to emotions over primary needs.

Since the strong bias of emotions, companies appeal to them for advertising purposes, also because attacking the rational part of people seems to be more difficult than ever. Indeed, people are bombarded with millions of stimuli from promotional campaigns and through time they have developed a sort of shield to defend against their influences, so that the advertisements seem to be ignored, at least from the rational side of the human brain.
3. The role of the blogs in online marketing companies: an experience

The present chapter is devoted to report an attestation of the role of blogs at Suerte, an online marketing company, based in Spain. In fact, the contents of the chapter derive from the personal experience of the internship in the aforementioned company, where the role of blogs is fundamental. The evidences illustrated here corroborate the concerns about blogs' fairness, transparency and genuineness.

At Suerte, the blogs are obviously seen as a marketing tool, specifically they may assume two different versions according to their aims (Fig.6).

![Diagram of different blog types](image)

**Figure 6. Different blogs.**

Therefore, for each client two blogs are built: one corporate and the other acting as a supporter blog. Before proceeding in the analysis of their functions, it must be stressed that the corporate Websites refer always to electronic commerce shops, this means that their final goal is to convert visitors into customers. Furthermore, in the case analysed in this thesis, the firms are newly constituted, so that they do not enjoy authority and need to be ranked in Google, and even the products sold are new in the market contemplated.

Francesca Bandini
In addition, the budget issue is relevant, because all the techniques adopted to prompt the popularity of the Website do not involve a monetary expenditure (except for the Google AdWords paid search). This factor is interesting because it demonstrates that potentially everybody could initiate an advertising campaign online.

As regards the corporate blog, it serves for Search Engine Optimization reasons, this means that the attempt to be ranked higher on Google is accomplished, also, through a blog. In fact, by maintaining a blog the chance to occupy more prominent positions is considerable, due to the frequent updating, to more contents and to the potential bigger audience. Secondly, that blog provides more information, not only about the products sold, but also about the unique values in respect with competitors, so that it could be seen as an authoritative source of information. Even though the content is optimized for SEO the quality remains a pillar, above all at Suerte. In fact, the philosophy of Suerte, is based on quality contents and even the texts of blogs must be written according to trustworthy sources as long as this is possible. This feature is unusual for marketing companies, particularly for blogs contents, in fact, by surfing different blogs on a specific topic, exactly the same sentences can be found, as a result of copy and paste actions. Indeed, it is common to copy sentences from bloggers, who enjoy a higher level of credibility. This is detrimental firstly for people, who rely on blogs and consider them as remarkable sources of information, secondly for the company itself, in fact, by copying contents the authority of the Website decreases. Turning back to the goals of the corporate blog, another important role is related to customers' engagement, in fact, the relation with customers, or potential, would be smoother, the dialogue more direct and informal. However, this goal has been taken into account only marginally by Suerte, because the e-commerce is still at the beginning: usually customers start to interact when the brand is already well established, otherwise they prefer to reverse the attention to community, forums, blogs to look for recommendations.
The supporter blog is not directly associable with the company, so that readers are not aware that the blog has been written for commercial purposes. Indeed, it appears neutral, as if it were written by an independent private individual willing to dispense advice. The aims of the supporter blog are included in the following list.

- **Source of information:** similarly, to the other typology, the support should be perceived as a credible and useful source of information, so that even here, the quality of contents matters. The same is valid for the Search Engine Optimization principles applied.

- **Banners (advertising):** the only thing which might make people presuppose a relation with the e-commerce Website is the presence of banners. In fact, a marginal role of the supporter blog is to display banner advertisements of the company's Website. This objective is linked to the *contextual advertising* (Paragraph 3.2), in fact, the themes of the Website are more or less consistent with the products displayed.

- **Link building:** the major purpose is to provide a basis for link exchanges. As previously said, this practice is punished by Google, but it has actually been applied by almost all the Websites to boost their SERP positioning. The link building strategy involved with the blog is quite structured. In fact, it refers to the *triangular link exchange*, in contrast with the reciprocal one (Paragraph 3.1).

### 3.1. Blogs and link exchange

Notoriously, blogs have been performing link exchanges, in order to improve their positions in organic search results. A link exchange refers to the importance of citations, the referrals, attributed by Google's algorithm, in an attempt to have a link (*outbound link*) in other Website, with the purpose to be ranked higher. The exchange
of links is performed upon an agreement between bloggers, but sometimes instead of an exchange a blogger may ask for money for the link placement. Nowadays, bloggers are increasingly careful in using this tactic, because of the fear to be penalized by Google. In spite of the risk, the number and the quality of links pointing to a site are deemed to be one of the most important factors to ascend the SERP, therefore, link exchange is a system still used, but with some caution. For instance, it is believed that by inserting more than one link in a text Web crawlers are less likely to note the spam link, but not all the Webmasters agree on that.

Link building might be performed in different ways. The basic method is known as reciprocal link exchange (Fig. 7) and is judged quite unsafe; in this case the Website A have a link to Website B, and B has a link pointing to A.

Figure 7. Graphic representation of the reciprocal link exchange.

In the case of triangular link exchange (Fig. 8), also known as three ways link exchange, the Websites involved are three: A points to B and B refers to C in turn. This method is more sophisticated and is applicable when a person owns two different Websites (A and C).
3.2. A conceptual framework for supporter blogs

Before moving on to the practical experience, it appears reasonable to illustrate a conceptual framework for supporter blogs, since their existence has not been found explicitly stated in the previous literature.

In an attempt to depict its conceptual basis, two sub disciplines of marketing
shall be considered, one is that related to the content and the other to the context.

**Content Marketing** refers to the editing and publishing of substantial and compelling contents to comply with target's needs.

*Content marketing is a marketing technique of creating and distributing valuable, relevant and consistent content to attract and acquire a clearly defined audience with the objective of driving profitable customer action* (Content Marketing Institute, 2014).

This process involves the planning, the creation and the distribution of contents. Moreover, it does not pertain only to the Internet realm, but to all possible channels which can reach customers and it can be deployed in different sectors, for example in the health care industry to get people used to certain practices.

The main aim of content marketing is to engage people, this objective is attained through the valorisation of the contribution of the audience: storytelling techniques, a pleasant visual aspect of the text, the qualitative and usefulness of contents delivered. In fact, as previously said contents shall not be written for SEO reasons, but to foster public's attention (Davis, 2013).

Companies are increasingly using content marketing (Pulizzi, 2012), owing to a series of transformations intervened in recent years: the possibility to publish quickly and for free, the availability of workforce to perform this typology of job; in fact writers are always more interested in the obtainable economic benefits. Furthermore, people trust and accept contents without considering their sources. Another motivation, connected to the widespread of content marketing, is the role played by social media, indeed a company shall create appealing contents, in order to successfully integrate within those means. Moreover, the last updates of Google's algorithm, namely *Penguin* and *Panda*, attribute an enormous weight to WebPages referring to other credible contents.
Sometimes contents marketing is confused with one shot campaigns, such as those coherent with viral marketing, in reality it holds peculiar and different characteristics. Contents shall be delivered at a constant rate, because the final goal is to attract or retain customers throughout all the process: they should be “engaged in a dialogue” (Pulizzi, 2013).

Since the importance of the continuity, supporter blogs might be referred to content marketing, in fact, the posts published strive to provide a certain degree of utility, are published at regular interval and in further stages of development will be addressed to customers' interactions.

The other marketing realm which may be associated with the supporter blog is the contextual advertising, which exploits the capacity of the Internet to deliver customized messages, people's needs are matched without efforts. Contextual advertising is known also as context match (Broder et al., 2007) and is correlated to the positioning of advertising in a coherent context, specifically, in a Web contents in harmony with the sponsored product or service. The rationale behind the contextual advertising is the assumption that the target is reached easily and therefore the ads are more effective (Broder et al., 2007). This practice is deemed to be an evolution of Sponsored Search or Paid Search and reveals particularly useful for companies which have not a Website or a specific Internet strategy, because they can gain benefits from the Internet.

Thanks to the contextual marketing the consumers are not prompt to visit a specific Website, but actually the message is delivered to them (Kenny & Marshall, 2000), and in this sense there is a shift from the usual dynamics. Nevertheless, it plays a central role for businesses which do not dispose of a Website, contextual marketing fulfils a significant role also for well established companies, because consumers consult more than one platform. For instance, before buying a Philips bulb, a person visits not only the corporate Philips Website but also resellers or price comparing sites, that it is why contextual advertising...
marketing is essential for every kind of company.
Contextual advertising does not refer only to advertisements placed on a certain Webpage because of the congruity between the sponsor and the contents (given by keywords match), but also to the target. In fact, it can be enabled some settings which allows to display a certain advertisement on an uncorrelated Webpage, because the audience is likely to be interested in the text and in the product: for instance a banner promoting gym shoes in a Website for trekking addicted.
Apart from the more efficient targeting, the contextual advertising allows to smoothly measure the impact by looking at the click through rate. Those clicks might refer to banners, links or anchor texts within the content itself, in fact the format of the advertisements may be various.
This system works thanks to the cooperation of different operators: the publisher, the ad-network, the advertiser and the users. The ad-network (Broder et al., 2007) is a third party, between the content writers and the advertiser, in charge of matching the contents and ads; as previously reported, the role is embodied by specific service providers, such as Google AdSense.
Nevertheless, the supporter blogs differ from the last mentioned feature, in fact, the banners are manually placed in line with the contents of the post (actually, the contents of the post depend on the banner to be displayed) and there is not the intervention of a third party.

3.3. The development of the contents

In this section, the path towards the creation of blogs' posts is examined. For the sake of clarity, it must be underlined that only the development of the supporter blogs is considered, because the interest of our analysis is in detecting the influence of those not alleged commercial blogs. In spite of this, all the techniques referring to SEO are used for building also corporate' blogs contents.
In contrast, to set up a supporter blogs the presence of a programmer is not necessary and open source software is used: the publishing process on corporate blogs is longer, because first the content writer creates the texts and secondly those are checked and transferred into HTML language by an expert.

At Suerte, Wordpress is the software exploited, which gives also indications for SEO, so that the tool can be easily handled by everybody. Indeed, the presented explanations derive partly from Wordpress itself and partly from the recommendations within the company. The steps to create an optimized post are portrayed in the chronological order.

1. **Keyword analysis phase.** The step involves singling out the keywords, according to which the post has to be written. In order to perform this task, the keyword planner tool provided by the Google AdWords has been used. Before examined this tool, it is worthwhile to stress that Google AdWords is a service run by Google to set up campaigns for the paid Search results. Even the paid search spaces, which can occupy the primary positions, or the right or the bottom of the SERP, are experimenting fierce competition for certain keywords. Indeed, Google has established a specific algorithm to deal with the issue, which works similarly to PageRank: the winners are those who have established a higher budget and show a higher quality. Google reports that in principle, a firm with a lower budget might be ranked higher within the paid search results, if its quality is major than competitors: the budget refers to the maximum amount of money that a person is willing to pay each day for a click on the ads or for a conversion. The options for budget might be customized; therefore a company which sets the advertising campaign has full control over the expenses.

Coming back to the explanation, through Google AdWords it is possible to have information about keywords.
The **relevant keywords** should be selected: *relevant* from the point of view of the content and from the users' viewpoint. Indeed, the words, which the users are likely to type for finding a response to their doubts, should be considered. Once having selected few different keywords, they should be tested in the keywords planner tool. Then, by typing the latter in the search bar (Fig. 9) the software will give back information about the words inserted.

As it can be seen from the snippet, the number of searches per month is displayed, the data indicates the average times a word has been looked for by users monthly: the higher the number the more popular the search. The competition is inherent to the paid search results in an alleged campaign, so it refers to that within companies trying to conquer the space of the paid search. Along the same line, there is the suggested bid, associable to the estimated cost per click in a pay per click campaign, usually the higher the competition the higher the cost per click and vice versa.

Subsequently, the words must be chosen according to the criteria selected. In fact, at *Suerte*, for corporate Websites' the most searched words are picked up, while for blog posts the focus is on long tail keywords. Those refer to a more complete sentence which has a small number of searches, indicatively it should have maximum of 100 searches per month, but an inferior number is preferred. The rationale behind this...
strategy relies on the difficulty to be ranked high in a competitive environment, so that the hope is to manage to be more popular attacking a *niche search*. In addition, long tail keywords pertain to the domain of natural search, which means that the research is performed according to the more natural human language, e. g. sentences, instead of a single keyword without context.

During this phase, it is possible to get hints for different keywords by looking at those suggested by Google itself. Alternatively, another tool that might be used is Ubersuggest [www.ubersuggest.org], which retrieves several ideas for keywords by typing the topic in a devoted search bar.

2. **Editing phase.** Throughout this stage, the optimized content is obtained. The optimization pertains to the congruence with search engine’s recommendations and users' expectations.

**Keywords:** the keywords selected for the blog post must be contextualized; in fact, they should be in harmony with the argument and result natural. The keywords should appear at least in the title, in the Meta description, in the first paragraph, because the purpose of the blog should be immediately clear. In addition, it is suggested to comprise them in some subheadings and in the alt tags of images. It should be avoided to use the same keywords in more than one post. The URL should contain the keywords, in the case of the long tail, each word is separated by a hyphen [-]. As regards to the keyword density, Wordpress automatically gives instructions. For instance, the repetition of keywords for eight times in a text of 340 words is already considered a good density.

**Length:** according to Wordpress each post should contain at least 300 words. At *Suerte*, the rule is a minimum of 500 words. The length of the title recommended is 40 characters minimum and 70 maximum; the Meta description should not exceed 160 characters.
♥ **Special features:** the content should be readable, so that short sentences and paragraphs (at least 3) are preferred, as well as, words in italic and in bold. Obviously, underlined words should be banned because they might be misinterpreted as links.

♥ **Outbound links:** they should be included and must point to high quality Websites: official associations, governmental institutions...

♥ **Media:** the posts should contain images or videos, but attention to the copyright's issue must be paid. In order to avoid these problems, the files can be taken from the database of pictures royalty free (e.g. flickr, morgueFile) or the URL of the source may be inserted, instead of saving them in the computer.

3. **Gaining visitors:** once the blog has been set up, visitors should be attracted to the site. To acquire traffic, the contents should be kept updated, at least one post per week. The popularity within the SERP should be augmented thanks to the presence of citations (links) in other Websites and this part is deeply revised in the next section.

3.4. **Public Relations and blogs**

Caring about the Public Relations (PR) of a blog means to bring it into life, so that the PR fulfill a crucial role. Indeed, on one hand a blog thrives thanks to visitors and interactions and on the other hand these conspicuous visits augment the visibility on the SERP. In addition, visits are important also to instrumentally gain links: the situation becomes a sort of vicious circle, because the more you are visible the more the links you get, but for a newly created blog, it is hard to find a way to be noted and to grow.

At **Suerte**, since supporter blogs are functional for triangular link exchange, the aim of the blog is to obtain links pointing to the e-commerce Website (the corporate one), so
that the mission is even harder because, generally, high quality sites are not willing to be associated with commercial purposes sites. A major constraint is in finding quality sites wishing to include a link without paying money, in fact, there is not budget available for this kind of activity at the time of the launch of the products.

However, the first step is to find blogs treating themes in line with the topic of the blog. Secondly, detecting the contacts and thirdly sending an e-mail to request the link exchange.

As regards the search for suitable blogs, even though it seems a silly duty, because of their large quantity, it is actually quite complicated, because most of them do not have a visible email address, and others are clearly sponsored by big online marketing companies, which surely will charge a fee. Therefore, the targeted Websites are those placing themselves in the middle, not with a wide audience, but with quality contents. A tool which can be used to assess the authority of a Website is Majestic SEO, which provides information about the authority of the domain name and the number of links. Furthermore, this instrument (Fig. 10) allows to see the back links, (the sources of incoming links), so that the same Websites might be contacted to gain a link.

![Site Explorer Results](http://www.curacorpo.com/)

**Figure 10.** Source: Majestic SEO [Online]. Available under subscription.
The software is useful, because, as previously stated, is not that easy to find blogs willing to grant links, but disappointingly, the major part of links of the Websites analyzed comes from marketing companies. This confirms that half of the worldwide bloggers derive their incomes from sponsored articles/posts (Technorati, 2013).

The activity of writing on someone else's blogs is known as guest posting.

The editing of guest posts, as well as to the other link building techniques, which are described subsequently, an important rule must be reminded: do not insert an anchor text containing the keywords of the Website, because this is one of the most common ways to detect a fraudulent link exchange by Google's crawlers.

In order to have an idea about the responses obtained concerning link exchange in Italy, some data deriving from Suerte are provided. The numbers reported are based on a sample of 180 Websites’ contacted by e-mails, which were in prevalence blogs, just a small amount of online newspapers.

First of all, it must be stressed that just the 22.22% of Websites responded, so the rate is quite low. This can be due not to the unwillingness to grant links, but to their interests in selling links. In fact, in the emails sent we do not ask directly for advertising purposes, because of the limited budget. This shows that without having a budget for links it is quite hard to obtain them, but it is still possible, in fact, the overall rate of positive response, which means, sites allowing free publishing of links is 9.99%: within this rate almost all of them do not ask for a link exchange, but they just permit to publish a content (containing a link).

It is interesting to have a closer look to the compositions of the responses received (Tab. 4): 37% required a payment, 22% do not allow links in their policies and 41% grant links for free.
Furthermore, it is important to observe the composition of the Websites allowing links for free, almost half are blogs in line with the topic of the link (e-commerce) and more than half corresponds to online newspapers: the phenomenon is relevant considering that just a few of newspapers were contacted. Indeed, apart from asking for links exchange to bloggers, another way to gain links, without granting anything in turn, is through articles. Commonly, online newspapers have sections not devoted to the news, such as health, society, curiosity, when people are allowed to propose pieces. Obviously, newspapers admitting publications of public contents are not the most famous and authoritative at national level. In spite of this, a newspaper embodies a status, people are led to trust them, without considering its popularity, so that the potential influence on readers is ample. Surprisingly, from the experience of Suerte, it can be seen that a great quantity of links belong to newspapers, this implies that people, while reading news, might incur in a link, click on it and be directed to an e-commerce shop. The issue should be addressed and not undervalued; nevertheless the articles before being published are checked, in the case of Suerte, they have always being accepted, probably because of the apparent quality contents.
The presented scenario might enact the need for more people in the sector of journalism, more reporters who transfer what they see through their eyes, instead of pursuing a collage of news taken from the Web.

Furthermore, another way to obtain a link is represented by forums, which are sites gathering people for discussion finalities, constituting some sort of real communities. They represent one of the oldest forms of social media and still some of them are active. It might seem easy to insert a link in those sites, but actually the anti spam rules are quite strict, so a great percentage of them does not allow link publication. Another problem is related to the time consuming matter, in fact, to enter a forum the user needs to be registered, and afterwards a discussion correlated to the link must be found or otherwise a new discussion should be started, but the commercial purpose should not be evident: people joining forums are aware of promotional messages, they are suspicious and this because in the past several platforms have been closed due to the big amount of spamming messages.

A further option for link building is represented by directories, for years they have been considered an easy way to obtain a link, nowadays things have changed, in fact, it is worth to place the link only in high quality directories with traffic. Moreover the topic of the directory should be related to the Website, the so called niche directories, so that it is not easy to find the right one.

3.5. Banners and effectiveness

The difficulties in weighting the impact of blogs have already been stressed in previous chapters; in fact, some data such as the number of visitors are strenuous to interpret. Despite this hindrance, some of the standard metrics may result useful, for instance the conversion rate in respect with the objectives set. Or qualitative information, such as readers' engagement, may be traced from content consumption
(Wilcox-Ugurlu, 2011), for example the amount of time spent on a page, the bounce rate, which refers to the percentage of people leaving the page, just after having entered it, without interacting with it and the derivation of the users. Another observable number is that referring to the click through rate of banners displayed on blogs, which would be particularly interesting in the case of supporter blogs, where the contents are neutral and not contaminated with advertising purposes, in fact, it might be used to assess the blog's importance for the official e-commerce site. Before proceedings in viewing this data deriving from Suerte, it appears logical to portray part of the literature about banners and their functions, since our interest is also contingent to banners' effect.

Banners derive from the display advertising and might be defined as small billboards on a Website respectively placed on the right, left, upper or lower side of a page: they entail interactivity and are functional to promotional purposes, in fact the user may click on it and being directed to another site (usually a corporate one). The first banner appeared in 1994 in a Web magazine, Hot Wired (Evans, 2009); in subsequent years this typology has assumed various formats in respect to the size, the position and the aspect. A fundamental change intervened is the development of networks, as that established by Google, to provide available banners to Website's owner, in exchange of revenues. The phenomenon has been referred as contextual advertising, as previously said: the banners are not randomly positioned, but they are inserted in harmony with the Website's contents, this is obtained through the match of keywords, which allows targeting the audiences more cost-effectively. This capacity along with the ability to receive almost immediate feedbacks are the major advantages of online advertising. The ability of performing a good targeting is possible thanks to the constant collection of data users while they are navigating (Evans, 2009). In fact, there have been a growing number of software permitting to spy visitors, the IP address is just one of the numerous details which can be retrieved, such as interactions, time length of the visit: the depletion of cookies (responsible for the traceability) raises privacy concerns, because several information about the users are easily stored, it is
true that people can enable some settings, in order to prevent this practice, but not all are aware of their existence.

Moreover, banners not only increase the traffic to a Website, potentially they have a cognitive impact, leading to brand awareness, indeed also its exposure effect should be taken into account.

Banners are increasingly widespread, as confirmed by the data about European Online advertising diffusion (iaB.Europe, Allison Fennah, 2012), which show that almost 50% of the budget is spent on display advertising, followed by the 32% in paid search and 19% is inherent directories and other activities.

Furthermore, Technorati (2013) reports that 61% of bloggers make money thanks to banners. Companies around the world are spending approximately 75% of their advertising online budget in display advertising and the trend is up growing.

As regards to the compensation models to reward external parties hosting banners in their sites, it evolved through time: at the beginning, the exposition of banners was paid according to impressions, which means that according to how many viewers see the ads the company has to pay the space's provider. Nowadays, instead the payments are performed on the basis of the click rate, in this case affiliate revenues are concerned and from the side of the firm, each click is considered a cost per acquisition (Chaffey, 2009), but Google AdSense allows flexibility also on this point.

Turning now to the assessment of banners effects, the literature presents contradicting finding, in fact its generalisability is problematic, so that different approaches and angles are illustrated.

Kuneinen (2013) stresses the effectiveness of banners' shapes by combining impressions (times that an ad has been seen), conversion rate (related to the purchase of a product) and click through rate. The results show that the static format performs better over the dynamic one, the square shape is preferable over the rectangle, the
horizontal and the vertical ones: he concludes affirming that the ideal banner would be square (or rectangular) and static.

Another thread concentrates on the exposure effects, such as Nihel (2013), who reports that the memorization and the click rate is enhanced when the banner is at the top of the page, big and the colors are bright. Contrary to his hypotheses, the memory and the click rate are not facilitated by animated banners and the image contained; in addition the more user experience does not correspond to more memorization. As regards to the clicks, the study infers that they do not depend on the length of the session performed.

In contrast, Li & Bukovac (1999) argue that animated banners facilitate the response of users and the impact on memory is greater in respect with static banners. They suggests that large banners, not only enhance comprehensibility, but represent also a major assurance for consumers, because the company is supposed to spend more, therefore to provide a higher quality.

A further research (Martín-Santana & Beerli-Palacio, 2012) explores whether textual advertising is more effective of banners, surprisingly the results reveal that displaying formats (especially rectangular) are superior to in-text ads. This superiority is related to cognitive, affective and conative (users' interaction) spheres; moreover, people more involved with the product itself and with the Website show more intention to purchase, a more favorable attitude to the product and the contents are judged more informative. Along the same path is the study of Manchanda et al. (2006), which exacerbates the effect of banners on the actual purchase, finding out that the increased exposure to banners is positively correlated to purchase probability, especially on repeated purchase. But the more the variety of the ads, in term of content of the message, the decreased purchase.

A different goal is pursued by Newman et al. (2004) investigating whether the presence of banner decreases the credibility of the Website or not. The main assumption, retrieved by psychology is that the more the people's usage of the Internet, the more banners are favorably perceived, indeed the conclusions confirmed this; although the results are unclear in respect to the purchase intention. In relation with the
brand's image the enquiry sustains that incongruent banners lower brand images, so that in order to obtain a reverse effect, the banners should be in line with the contents.

Now, the attention is driven towards the measurement of supporter blogs' effectiveness at Suerte, which is computed by traffic rate to the corporate Website, since the main aim of those is to be a platform for link exchanges. Fig. 11 depicts the percentage of visitors of a supporter blog, who subsequently have visited the corporate Website in a month period.

![Visit Sponsor](Visitors who completed the goal.)

Figure 11 Conversion rate of a supporter blog [www.clicky.com].

The percentage 10.47% needs to be disjointed in order to well interpret the actions undertaken by those visitors in order to be directed to the corporate Website: in fact a person could have clicked on one of the two banners, presented in each page of the blog, or could have clicked in a link presents in just one specific page, as part of the contextual advertising. Indeed, even though the supporter blog is not directly associable with the official company, it contains only one link within the text of a Webpage, which brings to the e-commerce shop. By splitting the index it can be observed that barely all the visitors, 9%, have clicked on the contextual link, prompting the argument in favor of contextual advertising; the rest, only 1%, have clicked on banners. This percentage might seem insignificant, but it shall be compared to a benchmark to grab its relevance. In this case, the benchmark is represented by the click through rate of banners, related to e-commerce, of the same market of the
supporter blog, in the same period. Indeed, Fig. 12 illustrates that the average CTR (click through rate) of the contingent sector is 0.16%.

![CTR Diagram](image)

Figure 12. Google (2014b).

On the light of the comparison it can be seen that the supporter blog is performing well also in respect with the banners, by having click through rate twofold higher than the industry benchmark (0.16%).

From the data presented it might be supposed that the supporter blog is successful in driving readers to the Website, although without prompting the achievement of this objective, because primarily constituted for link exchange practices.
4. Assessing blogs' impact: a randomized controlled experiment

Since the wide diffusion of blogs, the argument of the submitted research is related to the people awareness of blogs' features and the effect of banners on the perception of the contents. In order to verify these premises it is necessary to endeavor an exploratory research, in fact, a randomized controlled experiment was set up by means of a questionnaire. The questionnaire appears to be the most appropriate technique to examine the perception of blogs; because of its large use in previous studies. Furthermore, it allows a deepen insights to detect ideas and attitude, indeed questionnaires represents one of the most effective method when qualitative data are required.

The randomized controlled experiment brings solid and accurate results, because an ultimate causal relationship may be stated, statistical tools may be used for the interpretation of the results and the biases given by the socio demographic factors of the sample are eliminated.

In the subsequent sections the questionnaire's design, the hypothesis to be verified, some statistical concepts and the results are illustrated.

4.1. The randomized controlled experiment

The randomized controlled experiment was established to assess the impact of the banners within a blog post, the method employed is an online questionnaire, whose format does not reflect a standard questionnaire: participants were asked to read the content of a blog post, which referred to the benefits of yogurt. After having read the Webpage, respondents could start to reply to the questions.

An online questionnaire has been delivered for two reasons. First of all, it constitutes the optimal way to reproduce the Webpage and secondly the targeted sample was
people who are familiar with the Internet, otherwise it would be impossible to be influenced by it.

The questionnaire was created via the open source software *kwik surveys* [www.kwiksurveys.com]: the Internet offers different services for research purposes, but the free subscription entails always some limitations, for instance with regards to the number of responses obtainable. In contrast, *kwik surveys* allows unlimited number of answers, the possibility to set more than one questionnaire contemporary and gives back reports.

In order to set out the experiment, two questionnaires were distributed: in version 1 the blog post did not contain banners, while version 2 contained banners. This fashion reflects the aim to create an experimental design, wherein responses of the version 1 represent the control group and those of the version 2 embody the treatment or experimental group. The version 2 contains two banners, which are animated and in a small format. Since the open debate on banners' features effectiveness, animated ones have been chosen only in order to increase their weight within the post, because they should be visible. In addition, they sponsored two brands of the same product (yogurt), but one of them is well known, so more easily remembered, whilst the other is not existent.

The software used for the creation of banners is *banner snack* [www.bannersnack.com], which grants the service for free and provides pre standardized banners.

![Figure 13. The two brands shown in banners: the first is the popular brand, the other is the unknown.](image-url)
Importantly, the questions were the same in both versions, in order to enhance comparability; there were just some extra questions for people exposed to banners (Appendix a).

As regards to the assignment of participants to the groups, it was random, indeed, participants were asked to fill out survey version 1 or 2, without considering any difference: potentially all the participants hold the same chance to be allocated in one or in another group.

In relation to the blog posts, their publication has been possible, because an entire blog was set up, called Il benessere in cucina. The program used to develop it, is AlterVista [www.it.altervista.org], which is similar to others open source such as Wordpress.

The texts of the posts of the two questionnaires' versions are exactly the same, none of them contained advertising or suggestions for certain brands. This choice reflects the willingness to create a post complying with the philosophy of the supporter blogs, developed at Suerte. Furthermore, all the rules for contents' optimization were respected, such as keywords density and positioning, the presence of quality hyperlinks etc.

The structuring of the questionnaire requires a peculiar attention because from the disposition and the wording of the questions depend the results themselves. The rationale followed to accomplished this objective is that proposed by FAO (1997), the different steps will be singled out.

1. **Decide the information required.** The main points to be enquired:
   
   attitude towards *banners* and related influence

   knowledge of the term *blog*

   socio-demographic variables

2. **Define the target respondents.** The target, as mentioned before, is constituted by Internet users: a wide target has been chosen mainly because of the increase likelihood
to have a high rate of responses. Nevertheless, the answers related to the socio-demographic variables give insights about the compositions of the audience, for an eventual segmentation of the results and permit to test the comparability of the two respondents' groups.

3. **Choose the method of submitting**: the survey have been spread through different channels, this was necessary not only to have more answers, but also to avoid that the same people would have done the two versions of the questionnaire, as well as to avoid selection bias.

**Facebook**: the surveys were published not only on Facebook profiles, but also in different groups based on different interests.

**Forums**: forums involving different topics which allow to publish links were randomly selected.

**Online newspapers**: articles for online newspapers were written, collectively around eight different articles, but it must be stressed that the argument proposed was inherent to the Internet in general; there were not connections with the hypotheses to be tested and this in order not to risk the influence of the audience.

**E-mails**: the surveys were sent through email to people working for the university and former students.

4. **Decide on questions' content**. The contents were established firmly according to the objectives, since the necessity to shorten the number of questions, in fact, respondents are likely to be more reliable if they do not get bored by filling the survey.

5. **Develop the question wording**. The question wording refers to how the questions are written; here several pitfalls shall be avoided, because directly linked to the reliability of the survey (Eiselen & Uys, 2005). The questions were simple and
concentrated in few lines; they used a clear and structured language, to adapt to the different members of the audience. They attempted to be neutral, not emotional and not leading, in order to preserve the influence on the respondents. They were straightforward and avoided double negation. The options were mutually exclusive, for instance in the case of the age, each interval contained diverse ages, they did not overlap (e.g. Option a: 36-45. Option B: 46-55).

6. **Put questions into a meaningful order and format.** Even the order impacts on the reliability of the answers, for instance it has been decided to insert the socio-demographic questions in the end, to elicit the interest of respondents in the beginning. Furthermore, the order was logical and the demographic questions pertained to the same block.

7. **Check the length of the questionnaire.** The length was judged essential to evaluate the hypothesis.

8. **Pre-test the questionnaire:** the pre-test was performed by two people; this led to adjust a question. This phase was not taken into account sufficiently, in fact, when the surveys were already been lunched, it was found out that two mistakes in two different questions were committed, so that those responses were deleted. This reflects the scarce accuracy in the conduction of the pre-test, whose importance was undervalued.

9. **Develop the final survey.** As soon as, all the mistakes were sorted, the survey was launched in the different platforms before mentioned.

The surveys presented structured questions with nominal answers, no binary answers were provided, because even the question holding just the options *yes* or *no*, permits a third answer *I don't know*. This factor is important to evade an error of Question
Wording (QW), in fact, not all the people have an opinion about a certain subject and by obliging them to answer yes or no, the results are falsified. Researchers should be aware of QW errors, which refer to the possibility to influence the answers through the formulation of the questions. The likelihood of this type of bias is limited by providing linear and simple questions, without mixing the preference of one or more items. For instance, it should not be asked: "Do you like apple or banana?" Yes or No. People might find difficulties in answering and the responses are not reliable.

Closed questions were selected for several motivations. First of all, the tool kwik surveys does not permit opened answers for basic members, secondly closed answers present advantages for the respondents and for the researchers; indeed, they do not require a particular thinking activity, whilst they are easy to be interpreted, since the results might be counted in an effortless way.

Therefore, all the questions were close-ended: sometimes they required to choose among different items, usually only one answer must be selected. Another typology of answer was ordered scales, wherein respondents replied according to their opinions, in fact, they are specific to measure the intensity (Taylor-Powell, 1998). Ordered answers were built according to the Likert scale (Fig. 14), which consists of a set of choices to report thoughts in an accurate way, in fact, the answers may be chosen within several options corresponding to the level of agreement or disagreement. When a participant expresses agreement, the answer is on the left side of the scale, on the contrary sentiments of disagreement refer to the right side, and the neutral attitude is exactly in the middle.

![Figure 14. Likert scale.](image)

The Likert scale was selected because of its advantages, indeed, it is possible to calculate the results by attributing to each option a certain weight expressed in
numbers: it was established to attribute values from 1 to 5, where 1 represents a strong agreement and 5 a strong disagreement. The scale is also a valuable way to avoid the response bias, in fact, instead of asking directly if a person agree or disagree, which might lead to assertiveness, an opinion is demanded. The response bias relates to the attitude of respondents to comply with the supposed expectations of the researcher. The response bias might be related also to the tendency of people to maintain the same attitude, a sort of internal consistency, with the answers given, to discourage this behaviour questions inherent to the yogurt (which was the topic of the text to be read) were physically distanced in the survey.

Even a filter question was introduced in version 2, asking people whether they remember the brands shown or not: it entails a yes or no response because it is not significant in the results' count, in fact the purpose of question was to make the questionnaire more logical.

4.2. Hypotheses

The hypotheses were formulated before interpreting the results, but it was not possible to test all of them according to the randomized controlled experiment praxis, which entails statistical tools. Indeed, some results turned out not to be statistical significant and in other cases, the observation of the percentage was judged more suitable.

1.a. People are so used to banners that they are not bothered by them.

Respondents are used to banners, and then they do not consider them to be too invasive or disturbing.

1.b. People's attitude towards a product might change after having read the content, especially people who saw banners.

People reveal to be more willing to buy the product after having read the information supporting the positive sides of the product, this attitude should reveal stronger in
Unveiling the perception of blogs as marketing tools: a randomized controlled experiment

people pertaining to the treatment group, because banners may evoke the purchase intention.

1.c. People remember brands showed in banners.

Despite respondents do not believe to remember the brands showed on banners, they remind them, as soon as the name of the brand appears. It is interestingly to see the rate of people remembering the unknown brand, even though it is expected to be lower in comparison with the famous one.

2.a. People do not know which kind of website are visiting

This hypothesis is one of the central motivations behind the study, in fact, it was supposed that people are not able to identify the differences between Websites, therefore they were asked to classify the site visualized. The lack of knowledge does not depend on the education level or on the frequency of the Internet use.

2.b. People are not aware of the features of a blog

Particularly, people are supposed not to know the distinguishing characteristics of blogs. Since blogs are difficult to identify because of their similarity to corporate Websites: the blog exhibited in the questionnaires presented characteristics, which make it distinguishable, for instance the date and the archive of old posts on the right side. The rationale behind this hypothesis is that if people cannot individualize a blog, they would not be able to shield against its effect.

3. People consider blogs' content useful and trust them.

Respondents are believed to perceive the contents useful and authoritative, without paying attention the source.
4. People do not see commercial purposes on the blog which does not contain advertising.

Respondents perceive the Website as informational, without any commercial purpose.

4.3. Statistical analysis fundamentals

Before analyzing the results, some premises about the design of the experiment and some statistical concepts, which are necessary for the comprehension of the next paragraph are illustrated.

The design of the survey, involved two groups, one embodying the control and the other the experimental role, in order to test the banners influence and overall blog's awareness. The control group was established to test the null hypothesis, so that what would have happened without the treatment. The groups should be representative of the population to be studied to extend the findings to the whole population. Furthermore, to make a comparison between the two groups must be identical, obviously, this condition is impossible in the real world, because two groups of people would always be different in some respects, the aim here is to obtain participants as closer as possible in relation to all the characteristics which may affect the dependent variable to be established. Therefore the problem to have a ceteris paribus condition arises, this means that the participants should be equal in regards to relevant factors: in order to achieve the ceteris paribus condition, the random assignment of participants to the groups must be performed, but it is not easy to implement such procedure. Notwithstanding this difficulty, the experimental design allows controlling for the threats that are mining external and internal validity. External validity refers to the generalisability of the results, which entails that the causal relation discovered may be applied to other groups, while internal validity refers to the reliability of the estimated causal relation between variables. To insure internal validity, so that to provide a causal relation is important to take into account all the threats which might occur: each
experiment is more likely to be contaminated by one of them. In our case the biggest threat was represented by the selection bias, which indicates that the two groups are non equivalent and non representative of the population, because some people might be more likely to be included in the sample. A particular subcategory of selection is the self selection, which refers to the spontaneous candidacy to take part in the experiment. In the case of online surveys, this bias might really account for the explanation of the causality, because some people are more exposed and easily reached in respect to others, such as teenagers and less expert users.

On one hand random assignment procedure must be incessantly pursued, on the other hand econometrics provides a solution to control for relevant variables, so that even not having equal groups is possible to observe whether the differences between them have some weight on the dependent variable. If so we are in the case of a spurious correlation, because the variable to be tested (dependent) is affected by other factors than the experiment. To check this assumption is necessary to carry out a multiple regression analysis, permitting to discern among variables: there will be the independent variable (X) represented by the control and the treatment groups, a dependent variable (Y) embodying the questions, the hypothesis to be verified and all the other independent variables which might affect Y, refers as the population's parameters (e.g. age, level of education...). In each research those variables are selected differently, according to the goals of the observer, but it is important to carefully consider all the relevant variables, because omitting those means not to obtain reliable results. In fact, the omitted variable bias should be fought against, the latter appears when a determinant variable of the dependent (Y) is not taken into account and when it is correlated to the independent variable (X). The likelihood of this bias in our experiment is low and if the bias exists depends on casualty, because a peculiar effort was made to include and control for all the relevant variables which might have had influence on the answers, namely the population parameters.

In our case the relevant factors which would have possibly affected the dependent variable were the age, the Internet use, in relation to the activities performed and the

Francesca Bandini
time length, the level of education and the usual yogurt consumption. As already underlined the statistical analysis permits to control for the population influencers which might be important for small sample, because they might account for the explanation of the causal relation. Therefore, for each set of answers of a certain question, we performed a multiple regression analysis including those parameters, this results can be seen in the column Regression coefficient with population parameters of the first table of Appendix b. We obtained coefficients determined by the presence of the advertisements and do not dictated by other factors, e.g. age, Internet use...

Indeed, in the equation (Fig. 15) the supposed influencing variable to control are represented by $\beta_nX_n$, while in the linear model they are included in the error term ($\varepsilon$), by doing this the assumption is that the different population factors are not correlated with the main independent variable ($\beta_1X_1$): all the factors affecting $Y$ are the slope parameters.

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \ldots + \beta_nX_n + \varepsilon$$

Figure 15. **Multiple regression analysis model.**

So that the equation is constituted by the independent variable to be studied $Y$, the intercept $\beta_0$, representing the zero point (when all the independent variables are equal to zero), the slope parameters $\beta_1x_1$, $\beta_2x_2$... $\beta_7x_7$ affecting $Y$.

One of the basic assumptions is the zero conditional mean which implies:

$$E(u|x_1, x_2, \ldots, x_k) = 0$$

Figure 16. **Zero conditional mean.**

Fig.16 illustrates that all the slope parameters, the error term, potentially affecting the explanatory variables, are equal to zero, so that they do not have weight in defining $Y$. This assumption fails if some important parameters are not taken into account.
Another key assumption in multiple regressions is the no perfect collinearity among variables, which means that there is no correlation among the independent variables; each of them does not hold a linear relationship with another. This is an assumption that can be checked ex-post, and is satisfied in our dataset.

Finally, since all these premises the parameters defining the variables are judged unbiased: their values have been collected through an unbiased procedure because the data derive from a random sample. It must be stressed that it is not credible that the parameters are completely unbiased since they are taken from a particular sample and not from the whole population.

Another fundamental statistical concept to disclose is the dummy variable, in fact, in the regression analysis which is illustrated in the next sections, the independent variable X pertains to this kind. X is a binary variable, or dummy, because it embodies two different values, 0 or 1, representing the answers of the two different samples:

\[ X=0 \quad \text{control group} \]
\[ X=1 \quad \text{treatment group} \]

Then, in this case it does not make any sense to talk about slope coefficients and parameters, because the regression model acquires a different meaning. Let's take the simple regression model when X is a dummy variable \( Y= \beta_0 + \beta_1 X + \mu \). The intercept \( \beta_0 \) represents X when embodying the control group (0), while \( \beta_0+\beta_1 \) are the experimental group (1), while \( \mu \) is always the error term.

In our experiment the null hypothesis to be tested implies that there is no difference in the perception of contents between the group exposed to the advertisement and the group which was not. According to our proceedings the null hypothesis is tested for each questions, by looking at the statistical significance: we supposed that the Y random variable is represented by the answers of the two random samples, and the null hypothesis is tested to see whether the presence of advertising, specifically banners,
holds an effect or not. Through the observations of the T-value or T-statistic is possible to assess the significance level of the data from a statistical point of view. When the t-value is higher or equal to 1.96, $|t| \geq 1.96$, the null hypothesis is rejected and the significance level of the data is at 5% and the related confidence interval is 95%. The significance level at 5% means that the probability to have rejected the null hypothesis incorrectly is 5%. As a consequence, the level of confidence states that our hypothesis is true or not in the 95% of the case analyzed.

The number 1.96 is retrieved from the Student’s $t$-distribution, which provides yardsticks aimed at evaluating whether to reject the null hypothesis or not. Those yardsticks are established by mathematical theories, the comparison value shall be selected by looking at the desired degree of freedom represented by $n-1$, which is the number of observations. In our case, the total number of observations is 142, so the chart (Table 5) should be observed according to the degree of freedom connected to 142 and to the chosen confidence level, which is usually fixed at 95%.- Despite all this reasoning, in our case the sample size exceeds the 100 units, therefore it does not appear relevant to establish the degree of freedom, but the general numbers of reference ($z$) should be taken into account.

<table>
<thead>
<tr>
<th>Z</th>
<th>0.000</th>
<th>0.674</th>
<th>0.842</th>
<th>1.036</th>
<th>1.282</th>
<th>1.645</th>
<th>1.960</th>
<th>2.326</th>
<th>2.576</th>
<th>3.090</th>
<th>3.291</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0%</td>
<td>50%</td>
<td>60%</td>
<td>70%</td>
<td>80%</td>
<td>90%</td>
<td>95%</td>
<td>98%</td>
<td>99%</td>
<td>99.8%</td>
<td>99.9%</td>
</tr>
</tbody>
</table>

Table 5. Student’s $t$- distribution.

It must be stressed that the application of the t-test is exact only when the data among the sample are normally distributed, which is often not plausible in the reality. This problem is overcome when the number of observations is large, as in our case, because of the central limit theorem and thanks to the inclusion of the demographic variables in the sample. The central limit theorem infers that when the number of results to be studied is big and the sample derives from a random procedure, the means of values

Francesca Bandini
are normally distributed. The population parameters allow seeing whether the hypotheses might be rejected or approved notwithstanding the differences in the sample population.

In general, the confirmation or rejections of the hypotheses does not lead automatically to the establishment of a causal relationship, indeed statistical analysis sets a statistical correlation between variables, the causality is a different and more sophisticated concept.

However, in our case we can state that a change in the dependent variable is caused by the independent variables, because we are in the specific context of an ideal randomized controlled experiment (Stock & Watson, 2011). The notion is crucial and merits further explanations.

- **Ideal** means that the subject’s selection procedure is unbiased as well as the data collection.
- **Randomized** refers to the fact that each participant has been assigned casually to the treatment or to the control group.
- **Controlled** is because there is a control group in order to test the effects of the experiment.
- **Experiment** is a necessary part of the survey and subjects cannot to choose the type of treatment.

After these due premises, now the emphasis is in the interpretation of the statistical results derived from the regression analysis performed. The data considered to verify the hypothesis are related to the coefficient estimation of the dummy variable and the significance level given by the t-value.

The **coefficients** tell us the differences in the responses between treatment and control group and the **t-value** specifies the statistical significance of the data. It seems interesting to give an example of the methodology adopted to evaluate the coefficients,
which were computed through electronic software. The following example gives an idea of the interpretation.

| Would you define the text | promotional? | Regression coefficient | -0.31571857 | T-value | -1.5681 |

Let's assume the value of the intercept equal to 0. \( \beta_0 = 0 \)

The supposed equation of the control group is \( \text{promotional} = \beta_0 \), translated into numerical value: \( \text{promotional} = 0 \)

The supposed equation of the treatment group is \( \text{promotional} = \beta_0 + \beta_1 \), into numerical value: \( \text{promotional} = 0 - 0.31 \). This means that the observed change in the treatment group results corresponds to 0.31, this number should be considered on the light of the scale used to test the answer which was

1: total agreement
2: partial agreement
3: not sure
4: partial disagreement
5: total disagreement

In order to calculate the regression analysis all the questions must have an answer, so that the intermediate answer, referring to an attitude of doubts (3: not sure), revealed useful to represents the neutral value. In fact, when the answer was blank the value of 3 was attributed.

Coming back to the coefficient interpretation, the regression coefficient -0.31 indicates that the treatment induces a move of almost 1/3 of point to the left in the Likert scale, and hence more agreement on the statement. The negative sign points out that the answer is more oriented towards the left side of the scale, the smaller value (1 and 2),

Francesca Bandini
which in our case corresponds to the agreement: people who saw the banners judge the content more promotional.
As regards the \textit{t-value} it should be compared with the before mentioned value of 1.96, if is equal or exceed this number, the data is statistical significant and the null hypothesis must be rejected.

\textbf{4.4. Results}

The results derive from the observation of the numbers which have been processed through a statistical procedure, specifically a \textit{multiple regression analysis}, whose results are available in \textit{Appendix b}. This procedure has been undertaken in order to verify their statistical relevance and to control for variables (the population parameters), which might influence the results and contaminate the supposed causal relationship. Other results are suggested by the observation of responses expressed in percentage. In this section, the confirmation of disconfirmation of each hypothesis is illustrated.

\textit{1.a. People are so used to banners that they are not bothered by them.}

This hypothesis shall be accepted because people by a large amount do not reveal to be bothered by the presence of banners, instead almost 20\% of respondents report that they make the Webpage nicer and 24\% express doubt about this.

People are not declaring overtly to be willing to click on banners, but at the same time they are not particularly reluctant to them. By observing, those numbers under regression analysis in order to reveal their statistical significance and to exclude other factors than banners influencing the answers, it can be seen the fact that the banners do not represent an annoyance is statistical significant at a rate of 95\% and the responses are not influenced by the internet usage or the age, etc.

\textit{1.b. People's attitude towards a product might change after having read the content, especially people who saw banners.}
This hypothesis needs further investigations, nevertheless people do not show a concrete linear tendency in the willingness to buy a yogurt; actually the results are almost split. Furthermore, there is not a more positive attitude to the purchase in the case of the treatment group, so that the banners do not prompt the purchase. Instead, by observing only the responses of the treatment group (Tab.6) inherent to the inspired willingness to buy given by the banners: it can be inferred that 63% show not to be interested in buying the yogurt because. The prevalence of the unwillingness to buy is statistical significant at a 90% rate.

Table 6. Treatment's group responses to "The banners make you feel like purchasing a yogurt".

In addition, it seems valuable to look at the disposition towards the purchase in the participants who declares a certain influence of the text: a question asked if respondents consider important eating yogurts thanks to the text read. The data refers to both control and treatment groups: 77% of the people who report to believe that yogurts are important because of the text read demonstrate also a tendency to buy the product, but the units are small so that a major enquiry is necessary.
1. People remember brands showed in banners.

This hypothesis should be partially rejected; in fact, it is valid only in the case of the well known brand, Danone, which was reminded by more than half of the respondents (Tab. 7).

![Bar chart showing brand recall](image)

**Table 7. Responses related to the brand name.**

The so called exposure effect on memory reveals weak in the case of the inexistent brand, Studio, in fact, only 10% of people reported it and probably this is due to a random effect. It must be taken into account also the time of exposure which was just few minutes, so that the hypothesis needs further investigations.

2. People do not know which kind of Website are visiting

The hypothesis seems to find confirmation on the dispersion of the results among the options available; generally speaking the trend is to consider the Website a newspaper. Some typology of Website results known, such as the chat in fact, almost nobody infers to have visited this type. A relevant number of people report to have visited the site of a company selling yogurts, especially when banners were presented, 41% of respondents believe to be in this kind of site.
2. People are not aware of the features of a blog

The hypothesis might be accepted in part, in fact, the answers are quite distributed (Tab.8), basically people agree on the fact that what they are viewing is a blog, but a large amount declare himself insecure.

Table 8. People perception of the Website as a blog: results derive from the treatment and control groups.

By enquiring the results from the treatment group separately, contrary to expectations people do associate blogs with advertisement, indeed when the banners were shown people affiliate more the content to a blog: this effect is statistically significant at 80% rate.

3. People consider blogs' content useful and trust them

This hypothesis is confirmed; generally people report the usefulness of the contents and neglect its triviality. Surprisingly, people from the treatment group consider the content more useful, than the control group, this let us suppose that the presence of banners increase the authoritativeness of the Website. The statistical significance of this observation is fixed at 95% rate, when the population variables are omitted and at 90% when they are included.
4. People do not see commercial purposes on the blog which does not contain advertising.

The hypothesis should be accepted even with some reserves, in fact, by looking at the results of the two samples jointly, controlled with the population parameters; it appears that the treatment group perceives advertisements more than the control group with a high confidence interval. From the inner results of the control group, it emerges that the prevalence of respondents does not see ads, as expected. Notwithstanding, nearly half of the control group envisaged the promotional intents of the blogs. This might be due to the response bias or to the habit to be exposed to advertisements, major investigations are necessary in this domain.

Thanks to the statistical analysis is possible to confirm the internal validity of the experiment, but to assess the external one is necessary to compare the data of the sample with those pertaining to the Italian Internet users. The task is not free of obstacles in fact, it is difficult to find market research using the same parameters to calculate the percentages proposed; usually the numbers refers to the total population: we are interested in the percentages inherent only to the Internet users, because the information about non users are not relevant to the aim of the research.

The data used to assess the external validity of the results derive from two sources Istat (2013) and Eurostat (2013), but it was not possible to evaluate the eventual differences among all the population parameters considered, because of the lack of specific data.

In respect with the age parameter, the external validity revealed limited in fact, our sample is more representative of certain age groups: people aged between 36 and 45 and 26-35 years old are slightly unrepresented in comparison with the population size of the users. Despite this hindrance, the homogeneity of the whole sample seems to comply with the homogeneity of the population.
With regard to the activities performed by the Internet users, the sample is a good approximation of Italian attitudes, as it can be seen from the Table (Tab.9), the percentages weakly differs.

Table 9. **Internet activities: comparison between the sample and the Italians.**

In respect to the frequency of the Internet use, our sample is in harmony with the population tendency, in fact, the use is more common for younger people and this trend decreases as soon as the age grows and the level of education lowers.

To conclude, the external validity appears to meet confirmation, so that the results can be considered as a general rule for Italian population, but the causal relationship found is stronger for people aged 0-25.
Conclusions

The main purpose of the research was to study the effects of a blog used as a marketing tool on our perceptions and ideas, because of the potential negative influence of advertising on our behaviour and thinking. In fact, we depart from the assumption that blogs built with commercial aims do not represent a trustworthy source of information, but people are not able to discern this factor. Indeed, people do not tend to question the reliability of the information found in the Internet.

Blogs as marketing tools are usually presenting some forms of advertising, but people are not led to evaluate negatively the association of a Website with the ads, it is rather considered a norm.

The results of the study support that the influence of blogs with promotional purposes may be harmful and particularly misleading; in fact, when a Website shows advertisements is judged more useful. Indeed, the main value of the research is to have highlighted a positive correlation between advertisements and perceived usefulness of contents: a causal relationship between the presence of advertising and the alleged authoritativeness of a Website is found, owing to the high statistical significance of the results and the exclusions of other possible explanations, this causal relation must be acknowledged.

This surprising finding lets us to reflect about the issue: we consider a Website with banners more useful, probably unconsciously, but by being aware of this bias we may fight against its arousal.

In addition, this unconscious tendency is strengthen by our acceptance of banners, which is the most common advertising format, in fact, we do not disclose particular annoyance in respect to them. This evidence corroborate the outcomes of previous studies (e.g. Newman et al., 2004), which found a sort of habituation towards advertising. This quite neutral attitude might be bethought the results of years of exposure: we have developed a sort of mechanism which allows us not to consider
them, at least at a conscious level. A further demonstration of this likely relationship is confirmed by the fact that participants of the control group (who were not viewing the advertisements), envisaged a promotional purpose of the text.
Eventually, it might be supposed that we are aware of the commercial objectives that almost all Websites hold, but we are not able to include this aspect while judging their contents.

Nevertheless, the banners do not seem to prompt the purchase intention, at the same time the exposure effect on memory reveals weak, above all in the case of the unknown brand: this consequence does not comply with previous studies and the hypotheses, it might be due to the length of the exposure. So that, the consequence of banners exposure needs further investigations, above all the long term effect of repeated expositions, even though some works have falsified this relationship. Therefore, a direct influence on behavior, embodied by the purchase intention, is not envisaged, but an impact on the reported usefulness of information is detected.

As regards to the knowledge and the distinctive features of Websites, as predicted, people are not particularly keen on distinguishing the differences between blogs, corporate site and so forth, probably because they are not even thinking about the typology. In contrast, they reveal to know the features of social networks and chats, maybe because these kinds of sites are constantly grabbing media's attention. Moreover, in respect with the ignorance of the features of blogs, it can be foreseen that people are not aware of their promotional purposes; this would confirm the concerns about the quality of the information retrieved through the Internet.
To conclude, the influence of blogs as marketing tool on our knowledge and perception of the world needs further investigations, because several aspects shall be assessed, according to the kind of advertisements presented for instance.
The recommendation which spontaneously arises from the results is that people should doubt more the contents found in the Internet and not to be misled by the position
occupied by a Website within the search results. Moreover, when a banner is detected, the users should address a peculiar concern to the relationship between the Website and the ads shown, in order to intercept even the lightest form of persuasion.
Appendix a

QUESTIONNARE: Il sito web oggi

1. Leggi il testo della pagina Web Il benessere in cucina.
Potrebbe richiederti qualche minuto, ma sii paziente, è per una buona causa.

Hai finito di leggere? Allora procedi con le domande.

2. Pensi che mangiare yogurt sia importante?
   • Sì e lo pensavo anche prima
   • Sì e l’articolo me ne ha fatto rendere conto
   • No
   • Non so

3. Come definiresti il testo letto? Scegli il tuo grado di accordo con i seguenti aggettivi.

<table>
<thead>
<tr>
<th></th>
<th>Molto d'accordo</th>
<th>D'accordo</th>
<th>Indeciso</th>
<th>Disaccordo</th>
<th>Per niente d'accordo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utile</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Informativo</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Promozionale (pubblicità)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Banale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inutile</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4. Secondo te che tipo di sito Web hai appena visionato? Non c’è bisogno di tornare al sito, rispondi seguendo onestamente le tue impressioni e secondo il tuo grado di accordo.

<table>
<thead>
<tr>
<th>.</th>
<th>Molto d'accordo</th>
<th>D'accordo</th>
<th>Indeciso</th>
<th>Disaccordo</th>
<th>Per niente d'accordo</th>
</tr>
</thead>
<tbody>
<tr>
<td>. Sito web di informazione, tipo quotidiano online</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>. Sito web di un’azienda di yogurt</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>. Chat</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>. Forum</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>. Blog</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>. Sito web di medicina-salute</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

5. Leggendo il testo, ti è venuta voglia di acquistare o mangiare uno yogurt?
   - Sì
   - No
   - Non so

6. Secondo te, erano presenti pubblicità nella pagina visionata?
   - Sì
   - No
   - Non so
7. Nella pagina web, erano presenti pubblicità, sottoforma di *banners*, in questo caso si tratta di immagini animate all'interno del testo. Esprimi il tuo grado di accordo con le seguenti informazioni.

<table>
<thead>
<tr>
<th>.</th>
<th>.</th>
<th>.</th>
<th>.</th>
<th>.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mi hanno infastidito</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Rendono la pagina Web più attraente e carina</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Non li ho notati</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Mi hanno fatto venir voglia di comprare uno yogurt</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Mi è venuta voglia di cliccarci</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Ci ho cliccato</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

8. Pensi di ricordare le marche presenti nelle pubblicità visionate?
   - sì
   - no

9. Sapresti dire quali erano le marche presenti nei *banners* visti?
   - danone
   - kir
   - studio
   - parmalat
   - yomo
   - brico

---

1 The question was not included in the questionnaire of version 1 (control group).
2 *Ibidem*
3 *Ibidem*
10. Coraggio, le ultime semplici domande.
Rivela la tua età.
- 0-25
- 26-35
- 36-45
- 46-55
- 55-99

11. Qual è il tuo livello di istruzione massimo conseguito?
- Scuola media
- Scuola superiore
- Università

12. Quanto spesso usi Internet?
- Più di un'ora al giorno
- Tra un'ora e 10 minuti al giorno
- Circa 10 minuti al giorno
- 2/3 volte alla settimana
- Una volta al mese

13. Per quale motivo utilizzi maggiormente Internet? Esprimi il tuo grado di accordo con le seguenti attività in base alla frequenza delle tua attività.

<table>
<thead>
<tr>
<th></th>
<th>Molto d'accordo</th>
<th>D'accordo</th>
<th>Indeciso</th>
<th>Disaccordo</th>
<th>Per niente d'accordo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mail</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ricerca di informazioni</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acquisti</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social networks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blog (sia scrittura che lettura)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quotidiani online</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Francesca Bandini
14. Quanto spesso acquisti o mangi yogurt?

- Tutti i giorni
- Almeno una volta alla settimana
- Una volta al mese
- Mai
Appendix b

The chart reports the regression analysis' results for control and treatment groups together.
The first two columns refer to the regression coefficient and the T-value, without the population parameters, the third and the fourth column reflect the results including the latter.

<table>
<thead>
<tr>
<th>Question</th>
<th>Regression coefficient</th>
<th>T-value</th>
<th>Regression coefficient</th>
<th>T-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-0.056338028</td>
<td>-0.557928353</td>
<td>-0.015658931</td>
<td>-0.15211</td>
</tr>
<tr>
<td>3.Utile</td>
<td>-0.295774648</td>
<td>-2.185333085</td>
<td>-0.265077329</td>
<td>-1.92746</td>
</tr>
<tr>
<td>3.Informativo</td>
<td>-0.14084507</td>
<td>-1.135400563</td>
<td>-0.154996137</td>
<td>-1.16604</td>
</tr>
<tr>
<td>3.Promozionale</td>
<td>-0.267605634</td>
<td>-1.40408205</td>
<td>-0.315171857</td>
<td>-1.5681</td>
</tr>
<tr>
<td>3.Banale</td>
<td>-0.211267606</td>
<td>-1.175508984</td>
<td>-0.194722358</td>
<td>-1.03164</td>
</tr>
<tr>
<td>3.Inutile</td>
<td>-0.014084507</td>
<td>-0.088132975</td>
<td>-0.059433215</td>
<td>-0.35125</td>
</tr>
<tr>
<td>4.Quotidiano</td>
<td>0.084507042</td>
<td>0.429889437</td>
<td>0.007764349</td>
<td>0.035327</td>
</tr>
<tr>
<td>4.Azienda yogurt</td>
<td>-0.323943662</td>
<td>-1.842317514</td>
<td>-0.418071517</td>
<td>-2.18416</td>
</tr>
<tr>
<td>4.Chat</td>
<td>-0.225352113</td>
<td>-1.834636288</td>
<td>-0.281519284</td>
<td>-2.08792</td>
</tr>
<tr>
<td>4.Forum</td>
<td>-0.225352113</td>
<td>-1.174892686</td>
<td>-0.296107866</td>
<td>-1.41205</td>
</tr>
<tr>
<td>4.Blog</td>
<td>-0.352112676</td>
<td>-1.795953708</td>
<td>-0.272846625</td>
<td>-1.29584</td>
</tr>
<tr>
<td>4.Sito medicina</td>
<td>0.028169014</td>
<td>0.148964833</td>
<td>-0.029688529</td>
<td>-0.14643</td>
</tr>
<tr>
<td>5</td>
<td>0.112676056</td>
<td>0.686860678</td>
<td>0.124302804</td>
<td>0.706385</td>
</tr>
<tr>
<td>6</td>
<td>-0.535211268</td>
<td>-3.982232267</td>
<td>-0.565649102</td>
<td>-3.89281</td>
</tr>
</tbody>
</table>
Unveiling the perception of blogs as marketing tools: a randomized controlled experiment

It should be reminded that the negative sign corresponds to a major orientation towards the left side of the Likert scale, which implies a stronger agreement with the statement from the side of the treatment group.

♥ The following table shows the regression analysis of the question 7 (Appendix a), the results are presented separately in respect to the previous one, because in this case the respondents derive only from the of treatment group. In fact, the question is inherent to the banners' perception and is central for the verification of the hypothesis 1a. The dependent variable Y is represented by the question's answers, while the independent variables X are constituted by the dummy variable together with all the population parameters, which might represent the explanation for the envisaged causal relationship. The population parameters are the same taken into account for all the other calculations (age, level of education, Internet use etc.). Here, it is important to observe the t-value in order to verify the statistical significance, therefore the internal validity of the results. The points of reference to judge the t-value are retrieved from the Student's t distribution chart, depending on the number of observations, the so called degree of freedom. Since the number of participants was 72, t-value has been evaluated according to 60 degree of freedom.

Degree of freedom: points of reference for the t-value

| 60 | 0.000 | 0.679 | 0.848 | 1.045 | 1.296 | 1.671 | 2.000 | 2.390 | 2.660 | 3.232 | 3.460 |
| 80 | 0.000 | 0.678 | 0.846 | 1.043 | 1.292 | 1.664 | 1.990 | 2.374 | 2.639 | 3.195 | 3.416 |
| 0% | 50% | 60% | 70% | 80% | 90% | 95% | 98% | 99% | 99.8% | 99.9% |
| Confidence Level |

**Question 7: qual è l'effetto dei banners?**

<table>
<thead>
<tr>
<th>T-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fastidio</td>
</tr>
<tr>
<td>Pagina più carina</td>
</tr>
<tr>
<td>Non li ho notati</td>
</tr>
<tr>
<td>Voglia di comprare</td>
</tr>
<tr>
<td>Voglia di cliccarci</td>
</tr>
<tr>
<td>Ci ho cliccato</td>
</tr>
</tbody>
</table>
References


Francesca Bandini


Francesca Bandini
Unveiling the perception of blogs as marketing tools: a randomized controlled experiment


Francesca Bandini


