An Empirical Study and Evaluation on the Cost-Effectiveness of Pay-Per-Click e-Marketing

Liming Chen

University of Ulster
Shore Road, Newtownabbey
Northern Ireland, BT37 0QB
Tel: 0044 28 90368837
l.chen@ulster.ac.uk

Connor Scullion

InvestNI Company
Bedford Square, Bedford Street
Belfast, BT1 7ES
Tel: 0044 28 90239090
connor.scullion@investni.com

Abstract—**Pay-Per-Click (PPC) e-marketing has recently attracted growing attention. It is particularly critical for the success of Small and Medium size Enterprises (SME) because they usually do not have extensive national, regional or global marketing networks. Given the cost, however, the question is how advertisers are rewarded for paid online advertising and to what extent. In this paper we present an empirical study aiming to evaluate whether or not active participation in PPC advertising programs would improve the indexing and ranking of a website in the organic search results of the major search engines. In order to do this we propose a systematic methodology for experiment design, data collection and performance analysis. We build comparable benchmarking websites and submit them to different search engines using both paid and unpaid methods. We monitor and collect the crawling activity of the search engine robots and the ranking placement of each of the sites in the search results of the respective search engines. The captured data are then compared and analysed from which findings are drawn and discussed. Finally we outline conclusions and point out future work.**

Keywords—**e-marketing, pay-per-click, ranking, search engine, empirical evaluation**

I. INTRODUCTION

If a business wants to be taken seriously in their chosen market today there are a number of things that customers will expect it to have; for example, a contact telephone number during business hours, an email address and a business website. However, when it comes to marketing their products or services online via a business website some small business owners find that they have little understanding of the technology involved. In fact, international benchmarking studies indicate that whilst website adoption within SMEs is widespread, e-commerce adoption is actually declining or static [1]. This seems to indicate that some businesses do not appreciate the potential of the Internet as an additional marketing communication tool or sales channel. As a result they may under-resource their web development projects because they think that the potential return on investment is insufficient or they simply don't have the budget to employ a reputable web design firm. A major consequence of this is that there are far too many business websites, e.g. in Northern Ireland, that have been developed either by owner/managers themselves after learning the basics of HTML and web design or by well-meaning amateur webmasters keen to attract an additional source of income.

In order to have an effective online presence SMEs need to understand more than just website design. They need to understand how to develop an online marketing and/or trading strategy that will successfully attract potential customers to their site. To achieve this they will need to learn more about the tools that a customer will use to find their website and this includes Internet search engines. Internet technologies such as search engines are increasingly driving one-to-one marketing relationships. On the Web, the first chain in customer relationship building begins not with a service's website, but in ensuring that customers and other desired demographics are referred to the site [2].

Generally speaking, there are three main ways that people can arrive at a website - direct navigation, web referrals and search engines [3]. Direct navigation is where a user finds a website by using the exact web address for example, by entering it into the web browser's address bar. Web referrals are links to a website from other sites such as online business directories. The third option - search engines, allow web users to find web pages on topics that they are interested in based on keyword searches. One of the most popular examples of an Internet search engine is Google, no other search engine updates so many sites as consistently fast as Google [4].

Search engines like Google 'crawl' the World Wide Web and harvest information on web page content and metadata and store it in a database. When a user searches for a particular term a list of relevant web pages is returned by the search engine. Gaining a listing in a search engine in this way is called an 'organic' listing. It is made possible by placing metadata in the

form of Meta-tags in the HTML code of web pages to describe what the page is about. Metadata provides an effective mechanism for describing and locating data that is relevant to a particular interest [5]. However, since too many webmasters have used Meta-tags for spamming, like repeating keywords to give pages a higher ranking, some search engines have stopped using them entirely [6]. For example, rather than rely solely on Meta-tags Google uses more than 100 criteria by which to rank sites and constantly refines them to stay ahead of the unscrupulous optimizers [7].

While search engines provide an invaluable service to users, they are ultimately owned by for-profit organizations. Business models may vary among these companies but most rely on advertising to generate a significant part of their revenue [8]. The advertising mechanism preferred by Google, Yahoo and MSN(Live) is Pay-Per-Click (PPC) advertising. PPC is an advertising model used on search engines, advertising networks, and content websites/blogs, where advertisers only pay when a user actually clicks on an ad to visit the advertiser's website. Advertisers bid on keywords they believe their target market would type in the search bar when they are looking for a product or service [9].

e-marketing is now big business for the likes of Google. In fact, spending on e-marketing surged significantly, for example, UK e-marketing spending overtook newspaper ads in 2006 for the first time [10]. However, the costs and unknown performance of PPC advertising presents a dilemma for most small businesses that have a website and they need to attract potential customers to. Should they pay for advertising using PPC and hope that the return on investment is substantial enough to derive a profit from or, should they rely on the 'organic' listing to help drive visitors to their site?

This is a major headache for SMEs with little financial resources for marketing. What make the situation worse are many conspiracy theories on how the participation in paid listings can affect the position of a website in the organic search engine results in a positive or negative way and some contradictory statements from relevant stakeholders. For example, we have personally heard on-the-street anecdotal evidence suggesting a link between the two. At search engine marketing seminars, speakers have actually stated that any web links used as part of a paid listing campaign will automatically be added to the search engine index more quickly and that the link will be crawled more frequently. However, Google firmly rejects this claim and states that there is absolutely no connection between being an AdWords advertiser and having the advertiser's site appear in the unpaid search results. One does not affect the other in any way. To put it another way, being an AdWords advertisers will neither help nor harm the advertiser's chances of appearing on the 'organic' search engine [22]. But, contributors to forums on the topic of search engine marketing are less convinced by Google's explanation of the issue and point to what they perceive to be more than

coincidental changes to their organic listings based on behaviour of their paid listings campaigns.

To help clarify the vague situation and evaluate the performance of the "paid" and "organic" website listing, we carry out an empirical study aiming to determine whether or not active participation in Pay Per Click (PPC) advertising programs would improve the indexing and ranking of a website in the organic search results of the major search engines. This paper presents the research, which is organized as follows: Section 2 presents an overview of the methods and mechanisms for search engine based website listing. Section 3 describes a systematic methodology for the empirical study and its implementation. We collate and analyse the experiments' data in Section 4 and discuss findings in Section 5. Section 6 concludes the paper and points out future work.

II. SEARCH ENGINE BASED WEBSITE LISTING REVIEW

Though a large portion of web sites exhibit some form of online advertising, search engines are no doubts the popular means for it and all major search engines provide such services. A search engine is an information retrieval system designed to help find information stored on a computer system, such as on the World Wide Web, inside a corporate or proprietary network, or in a personal computer. The search engine allows one to ask for content meeting specific criteria (typically those containing a given word or phrase) and retrieves a list of items that match those criteria. This list is often sorted with respect to some measure of relevance of the results. Search engines use regularly updated indexes to operate quickly and efficiently [11]. Imagine not having a phone book, no Yellow Pages, no directory assistance. It would be very difficult to use a telephone. That's what the Web would be like without search engines. Users would be aware of only a small fraction and could find new ones only through word of mouth [12].

In 2006, 256 million people used a search engine (81% of the internet population) creating 27 billion search views [13]. As can be seen from TABLE 1 the top five search engines have more than 90% of the entire search engine market. Google currently dominates the search engine industry. The most popular search engines utilised by web users are Google and Yahoo [14].

TABLE1.   INTERNET SEARCH ENGINE UK USAGE SHARE

Google	49.6 %
Yahoo	23.9 %
MSN(Live)	8.8 %
AOL	6.2 %
Ask	2.8%

Source: Nielsen/Net Rating (October 2006)

However, getting a good ranking in a search engine i.e., seeing a website listed in the first two pages of results is a tall order with so much competition. The first obstacle to overcome is the fact that a search engine will need to know the existence of a website in order for it to show the website in its results pages. The process of making a search engine aware of a website is called search engine submission.

A. Website Submission

There are typically two types of search engine, i.e., directories and crawler-based. Directories are search engines powered by human beings. Human editors compile all the listings that directories have. Getting listed with the web's key directories is very important because many people will see these listings. In addition, if websites are listed with directories, then crawler-based search engines are more likely to find the site and add it to their listings for free [15]. The two most important directories to get included in are the Yahoo Directory and the Open Directory also known as Dmoz.

The process for submitting a website to such directories is quite simple. Submitting to Dmoz and Yahoo Directory for non commercial websites is free but there is a cost for business websites submitting to the Yahoo Directory. To get a free listing all that a typical webmaster needs to do is to browse the directory to find the topic area that his or her website is most relevant to and submit the website's details such as the URL and a short description. The human editor of that section will then review the site and judge its relevance. If the editor agrees with the submitter that the site is in the correct category and that it is relevant then it will be included in the directory. If not it will be rejected. The emphasis in this type of search engine is on peer review as it relies on editors to be impartial and ensure the site being submitted is genuine and meets the basic protocols of what a website should be.

In contrast, crawler-based search engines like Google, automatically visit web pages to compile their listings. This means that, unlike directories, a website is likely to have several, if not many, pages listed with them. This also means that by taking care in how the website is constructed, it might rank well in crawler-produced results [16].

Webmasters can make submissions directly to crawler-based search engines in the hope that the robot will eventually crawl their sites and list them in their results. A typical example is the page on the Google website where webmasters can submit their URL for indexing. Another way to have a website crawled is to have a link to the site from a popular website that is crawled regularly by robots. The robot will follow all the links from the popular site and therefore eventually find any interlinked website, crawl it and index its content. Crawler-based search engines can be enormously complex but they each have three basic components; a software robot, a database or storage system and a relevancy algorithm, which are described below.

B. The Anatomy of a Search Engine

Robots: The robot of a search engine is essentially an automated browser. It views or 'crawls' web pages and strips out the html text content and uses it to compile a record of the words it recognises and where those words were found. A search engine can establish the significance of these words by checking if they are being given prominence on the page, for example, being used in titles or displayed in bold type, etc. This information is then passed to the database for storage and retrieval at another time.

Relevancy algorithm: The storage system, or database, holds a record of all pages viewed by the robot and when a search engine user enters a search it is this database that they are searching. The key to producing a good search engine database is to ensure that the data retrieved by the robot is accurately described and easy to access. To make for more useful results, most search engines store more than just the word and URL. An engine might store the number of times that the word appears on a page. It might assign a weight to each entry, with increasing values assigned to words as they appear near the top of the document, in sub-headings, in links, in the Meta tags or in the title of the page [17]. The precise weight given to each of these factors is the relevancy algorithm and these are a closely guarded secret for search engine vendors, for instance, Google uses more than 100 criteria by which to rank sites and constantly refines them to stay ahead of the unscrupulous optimizers [7].

Page Rank: The most popular search engine Google, measures relevancy by analysing the link structure of the Web, an approach borrowed from citation analysis. The measure, known as Page Rank, determines a page's citation importance or quality by calculating the number of web pages that link to a particular page as well as the quality of those pages (their respective Page Rank). In essence, a higher Page Rank value would thus indicate that a web page is more relevant to a query because other important web pages link to it [18].

Indexing: The information compiled by the robot and stored in the database also needs to be found quickly and for this reason an index is produced. There are quite a few ways for an index to be built, but one of the most effective ways is to build a hash table. In hashing, a formula is applied to attach a numerical value to each word. The formula is designed to evenly distribute the entries across a predetermined number of divisions. This numerical distribution is different from the distribution of words across the alphabet, and that is the key to a hash table's effectiveness [17].

In search engines like Google every word is converted into a number by using an in-memory hash table [18]. When a user enters a search term the search starts at the root of the index tree, and at every step a branch of the tree (representing many terms and related Web pages) is either followed or eliminated from consideration, reducing the time to search in an exponential fashion [19]. It is no surprise then that content really is king when it comes to the web. The main search engines gather

information about a website by crawling the content within html tags and identifying which topic area is given the greatest prominence. However, the html language has a method of providing additional information to crawlers that can describe what content is available on the page. That method is the use of Meta tags.

Meta data: The <Meta> element provides Meta-information about a web page, such as descriptions and keywords for search engines and refreshment rates [20]. Meta tags are placed in the header of web pages. They are an ideal way to give extra information to robots but they have also been abused to such an extent that only a handful are still used by search engines the rest are ignored. Metadata provides an effective mechanism for describing and locating data that is relevant to a particular interest [5]. Also in the header of the page is the title tag which is very important for search engines in that it appears as the main link in a list of search results.

There are only three Meta tags that a webmaster should use and they are; the description, keywords and robots tags. The description tag provides, as the name suggests a description of the page content. This will be used by the search engine to provide a description of the web page in a list of search results. The keywords tag allows the webmaster to list a number of keywords that should relate to the content of the page. However, too many webmasters have used Meta-tags for spamming, like repeating keywords to give pages a higher ranking so some search engines have stopped using them entirely [6]. Only a handful of the less common search engines even bother to read it. The last important Meta tag is the robots tag which tells the robot not to index the content of a particular page. This can be important if there are areas of the site that should not be made available to the public for example, a page in an administrative area.

C. Methods for Getting Websites Listed

Organic listing: A listing in the results pages of one of the top search engines can be achieved without spending any money. To do this a webmaster will need to submit a URL to the search engine itself, get a link from a popular site such as a social bookmarking site like Reddit or get listed in a search engine directory such as Dmoz.org or Yahoo! Directory for non-commercial sites. Exactly how long it will take to get listed and how successful it can be will depend on several factors that are outside of the control of the webmaster. Gaining a free listing with search engines is called an 'organic' listing.

Paid inclusion: The alternative to an organic listing is to pay for inclusion in the search engine. For example, to get a business website included in Yahoo! Directory there is a fee of $299 which means that the site will be reviewed and the Yahoo! editorial team will respond to the site webmaster within 7 business days. However, there is no guarantee that a web site will be included and if the site is accepted into the Directory for listing, there will be a $299 recurring annual fee in subsequent

years to maintain the listing [21]. But, it guarantees that the site will be crawled and listed in a relatively short space of time.

Paid listings: Paid listings are short text advertisements within search results screens, with links to the advertiser's site. Advertisers compete with one another by bidding to sponsor selected keywords. Keyword sponsorship operates on a pay-perclick basis, so advertisers only pay for those visitors who have clicked on their listing to go to their website [2].

Each of the major crawler-based search engines has their own version of a paid listings program that generates revenue, Google has Adwords, Yahoo! has Yahoo! Search Marketing and more recently Microsoft now have AdCenter. Business models may vary among these companies but most rely on advertising to generate a significant part of their revenue [8]. However, each system is relatively similar in that they manage sponsored links that appear in a prominent position on the search engine results page.

III. METHODOLOGY AND IMPLEMENTATION

In order to establish whether or not participation in paid listings programmes would improve the indexing and ranking placement of a website in the organic search results of a major search engine, a systematic methodology is developed to guide the empirical study. The methodology contains four steps. Firstly, several benchmarking websites are constructed for experiment. Secondly, each website is listed in the main search engines using three different methods - direct submission to the search engine, inward linking from social bookmarking sites and participation in paid inclusion programmes. Then, the traffic to each site by a search engine's robots is monitored and relevant data such as which robots visited them, how often they visited, are collected. At this stage searches for individual websites will be conducted using these search engines and the ranking placement of the website in the search results will be recorded. And finally, comparison and analysis of these experiments will be taken place. With this methodology we have designed an empirical experiment, which is described in details below.

V. Benchmarking Website Construction

We created three basic HTML websites that were identical in every way except for a single keyword. This guarantees the experiment data are comparable. The single keyword for each website is used to identify the website entries in the search engine results. The sites' code is kept as simple as possible to give the visiting robots every chance to index the sites properly and accurately. The use of Meta-data was restricted to the inclusion of Meta-tags for the title, description and keywords on each page. A robots.txt file was also included in the root folder of the live server to instruct any visiting robot to index all pages in the site without exception.

The websites were modelled on what the website of a typical small business might actually look like. We try to make websites

as close as possible to the real world scenario so that the data collected from the experiment are genuine and convincing. In this study the website is based on a local guest house, made up of five pages each. The three keywords chosen were: Aldervale, Aghacommon and Moyraverty. Each keyword produced a small number of search results (1100 results on average) making it easier to compare their performance against each other without having to trawl through huge numbers of results each time. The sites were made available on the web at www.aghacommonguesthouse.co.uk, www.Aldervaleguesthouse.co.uk and www.moyravertyguesthouse.co.uk using a third party hosting company.

VI. *Search Engine Selection and Submission*

One key decision to make was which search engines are included in this study. The most likely candidates were the most popular search engines as shown in Table 2. Worldwide, the most popular search engine is Google which accounts for almost half of all searches. Yahoo is number two but MSN(Live) is only ranked fourth with AOL and Ask as low as seventh and eighth respectively. One reason for this is the growing popularity of the Asian search engines, in particular Baidu in China (third most popular) and NHN in Korea (fifth most popular) [23]. To attempt to study all the main search engines would be impractical. Therefore, we decided to study the three most popular search engines that had UK paid listings programmes and had robots that could be easily identified. They are Google, Yahoo and MSN(Live).

TABLE 2. TOP SEARCH SITES WORLDWIDE (Billion Searches)

Google	37
Yahoo	8.5
Baidu	3.2
MSN(Live)	2.1
NHN	2
eBay	1.3
AOL	1.2
Ask	0.74
Fox	0.68
Lycos	0.44

Source: comScore (August 2007)

In order to compare the three different methods of gaining a listing in the organic search results of a search engine, each site is submitted at the same time but in a different way. The Aghacommon Guest House (Aghacommon) website was submitted directly to Google, Yahoo, and MSN(Live) using the 'Submit your site' page on each. It was also submitted to the Open Directory but not Yahoo Directory as cost was a prohibiting factor. For the Aldervale Guest House (Aldervale) website, we created links to it from other popular websites so that search engines can crawl and index them. In this instance the most accessible and popular websites available were the social bookmarking sites. We concentrate on the bookmarking sites currently endorsed by the BBC at the foot of each news story on their website namely, Delicious, Digg, Reddit, StumbleUpon and Facebook. Accounts are created on each of these through which a link back to the Aldervale website is included wherever possible.

The final method employed to gain a search engine listing was to open an account with the paid listings programme for each of the major search engines in the study and place an advert linking to the website for Moyraverty Guest House (Moyraverty). We opened an Adwords account on Google, a Yahoo Search Marketing account on Yahoo and an AdCenter account on MSN(Live). We created an advertisement on each account that contains relevant keywords for the website. We set a small budget and set a maximum cost per click for each advert. This enables us to limit the amount of money needed but still have an active paid inclusion programme for the duration of the study.

VII. *Tracking Traffic and Collecting Data*

In order to establish which method of search engine submission is the most successful and determine whether or not paid listings play a part in the indexing process, we track the activity of search engine robots on each of the sites and query the search engines themselves at regular intervals to determine the inclusion and/or placement of each site in the search engine results. The time frame for monitoring was a six week period because it is long enough to make a determination on the success of each of the submission techniques and short enough to lessen the chance of contamination i.e. robots indexing the site due to other factors beyond my control such as changes to DNS settings.

We chose a web hosting server that contains a web traffic analysis package for the purpose of this experiment. The package included was AW Stats which is a respected industry standard web traffic analysis programme. AW Stats can determine when a particular search engine robot has visited the concerned site based upon the requests made to the web server and recorded in the server log files. It is robust, accurate and was able to identify the three robots we were looking for, namely, Googlebot, Yahoo Slurp and MSNBot. In order to track the inclusion and/or ranking placement of each website in the search engine results we conducted a manual process whereby we queried each search engine on a weekly basis using the keyword for each website i.e. Aghacommon, Aldervale and Moyraverty. We sent a query (i.e. site: www.nameofsite.co.uk) to each search engine to determine if the site was, first of all, listed in the index and if it was we used the relevant keyword to establish if it was present in the first ten pages of results. All the results were recorded and collated for later analysis.

IV. RESULTS ANALYSIS

The results in this study are divided into three parts. The first part provides details of how each of the test sites of Aghacommon, Aldervale and Moyraverty performed using their respective search engine submission method and whether or not participation in the advertising programmes of Google, Yahoo and MSN(Live) actually improved the indexing and placement of the Moyraverty site. Part two of the results compares and contrasts each the performance of each site in order to determine which method was the most effective using the number of times each site was crawled and how well the site was placed in the search results of each search engine. Finally, the raw data collected regarding the number of times each test site was crawled by the robot from each search engine will allow us to make a comparison as to how active the robots of each search engine have been during the test period.

A. Results for Aghacommon

The Aghacommon site was used to see how long it would take to get crawled and indexed by the main search engines if a business simply submits their website URL directly to the search engine via the "submit your site" page on each search engine. We had expected this method to produce little or no results given that many search engine marketers have suggested that this method of submission is now all but redundant. However, the performance was much better than expected.

The Aghacommon URL was submitted to the search engines on 25 October 2007 and it was crawled almost immediately by the Yahoo crawler Yahoo Slurp. As can be seen from TABLE 3, by the end of week one the site had been crawled and indexed by Google but it took Yahoo until week two to index the site for the first time. Despite being submitted to the Microsoft search engine MSN(Live) the Aghacommon site was never crawled nor indexed.

TABLE3. AGHACOMMON WEBSITE PLACEMENT IN SEARCH RESULTS

	Wk 1	Wk 2	Wk 3	Wk 4	Wk 5	Wk 6
Google	8	7	7	6	7	7
Yahoo	-	1	1	1	1	1
MSN(Live)	-	-	-	-	-	-

The prominent placement of Aghacommon in the search results of both Google and Yahoo was another surprise. After just one week the site was appearing in eighth place on the first page of results for the keyword search Aghacommon in Google but there were no entries in Yahoo. By week two the site was in position one and two in the Yahoo results and had climbed to seventh position in Google. The Aghacommon site was able to maintain its position in the search results relatively unchanged for the duration of the six week test period in both Google and Yahoo.

B. Results for Aldervale

The search engine submission method used for the Aldervale site was to build inward links from popular websites, which were the social bookmarking sites of Reddit, StumbleUpon, Facebook, Digg and Delicious. The site performed broadly as we had expected it to, which is shown in TABLE 4. It is no secret that building inward links from popular websites will help a site get crawled and indexed by the likes of Google and Yahoo, but the actual speed at which the site appeared in the search results on Google was something of a surprise.

The Aldervale site was crawled and indexed in the first week by Google but it wasn't available in Yahoo until week three. It took very little time for the site to start appearing in the search results of Google with an entry from the Reddit website appearing at number two in the results and two other listings from the Aldervale site itself appearing in seventh and eight positions. The entry from Reddit was an impressive return within a week from just a single inward link.

By contrast, the indexing of the Aldervale site by Yahoo did not occur until week three so the only presence the site had in the search engine results of Yahoo were links from the social bookmarking sites of Digg and Reddit. The Digg site was the most prominent in search results gaining 11th position in week one with entries from Reddit in 28th and 80th. It is worth noting that there were no entries from the remaining social bookmarking sites StumbleUpon, Facebook or Delicious in the first ten pages of results on either Google or Yahoo for the duration of the study.

TABLE4. ALDERVALE WEBSITE PLACEMENT IN SEARCH RESULTS

	Wk 1	Wk 2	Wk 3	Wk 4	Wk 5	Wk 6
Google	7	2	2	4	4	10
Yahoo	-	-	1	1	2	2
MSN(Live)	-	-	-	-	-	-

In a similar way to the Aghacommon site the Aldervale site, once indexed, was able to maintain its position in Yahoo coming first and second in weeks three and four, and then second and third in weeks five and six. However, the site was in eighth and ninth position in week one in Google but by week two it had risen to second and third, then dropped back to fourth before falling back as far as tenth by the end of the study in week six.

The search results data collected for Aldervale from Google and Yahoo seem to indicate that both search engines found the links from Reddit and Digg, added them to their index, then followed them and as a result they were able to crawl and index each site successfully. This is in line with our expectation about how the search engines to operate but this was not the case with MSN(Live).

In the search results of MSN(Live) the entry from the Reddit website linking to Aldervale was as high as eighth in the search results after just one week. It then rose to fourth position in week two, dropped back to fifth in week three then disappeared completely from the results by week four. During the entire test period MSNBot actually crawled the Aldervale site 15 times but never indexed it once. No pages from the Aldervale site were ever found in the search engine during the study.

Therefore, we can only deduce that Microsoft have either developed a search engine that is smarter than Google and Yahoo because it didn't index the Aldervale site, realising that we were 'spamming' the social bookmarking sites, or the search engine has serious problems. Judging by the quality of the results returned by MSN(Live) during the course of this study we believe it to be the latter.

C. Results for Moyraverty

The submission strategy for the Moyraverty site was to buy Pay Per Click (PPC) advertisements on each of the main search engines Google, Yahoo and MSN(Live) and see if the robots of each would crawl the site, add it to their index and then give it prominence. Therefore, the Moyraverty site was effectively the crucial site for this study in that the activity and behaviour of the search engine robots on this site would answer the question of whether or not participating in PPC programmes would improve the indexing and placement of a website in the search engine results.

An Adwords account was opened and activated on Google along with an AdCenter account on MSN(Live). The Yahoo Search Marketing account was opened but not fully activated as the cost was prohibitive. The results of these accounts being opened and ads created, budget set and keywords bid for was in the end, nothing. The Moyraverty site was never indexed in any of the three search engines during the entire study period. It was crawled by Yahoo Slurp just three times, never crawled by MSNBot and only crawled once by Googlebot on the very last day of the test period. No entries for any page in the Moyraverty site ever appeared in the search results on any of the search engines.

We never expected this submission method to produce such poor results. We had thought that the sites would at least be added to the index of each of the search engines and that some results would be found in the search results. It now seems clear that the participation in PPC advertising programmes does not affect the indexing and placement of that website in the search results of Google, Yahoo or MSN(Live).

D. Analysis and Findings
  1) The most effective submission method

Figure 1 shows the total numbers of each sites crawled by all search engine robots during the study period. Obviously the most effective search engine submission method in terms of this

data was the inward link building of the Aldervale site. It was crawled 122 times by search engine robots with Aghacommon a total of 74 times and Moyraverty just 4 times. This was against the anecdotal theory that participation in PPC programmes would get the site crawled quicker and indexed better than the other two methods.
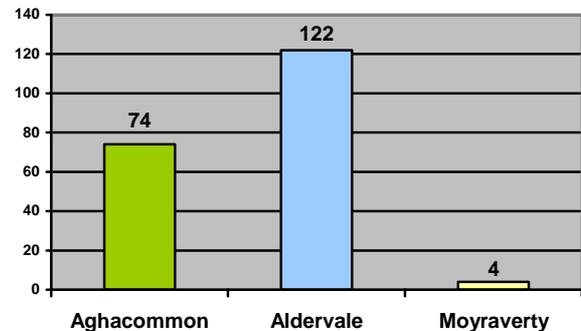

Figure1. Number of crawls on each test site

The inward link building also gave the Aldervale site greater prominence when the key phrase of 'Craigavon guest house' was used in Google. Aldervale appeared on the first page of results for that key phrase reaching as high as fourth position in 50,600 search results before dropping out of the top ten and onto page two in the last week of the study. The highest position for Aghacommon in Google during the test period was 22nd and no results were ever found for Moyraverty, not surprising as the site was never indexed by any of the search engines in this study.

When the same key phrase was used in Yahoo the results were quite different. Aghacommon was the first to appear gaining ninth position in 67,000 search results by week two. Aldervale joined Aghacommon in week three gaining eighth position compared to tenth for Aghacommon. The placement of both sites remained relatively unchanged during the study so much so that by week six the Aldervale site was eighth and the Aghacommon site was ninth in 81,100 search results.

From the data collected on the activity of robots and the placement of the sites in the search results of the respective search engines it can be concluded that the most successful method for submission to Google is clearly the building of inward links from popular sites. The inward link method was also more successful in Yahoo but only very slightly with just a single place between Aldervale and Aghacommon at the end of the test period.

  2) The most active search engines

We also monitor and collect the total number of crawls each search engine carried out for all websites, which can be used as a metric to measure the performance of each search engine. Figure 2 shows the pie chart of the search engine robot activity

in this study. The most active search engine is Google with Googlebot performing 109 crawls which amounts to 54.5% of the total robot activity on all three test sites. Yahoo is second with 76 crawls or 38% of robot activity and MSN(Live) the last with 15 crawls amounting to just 7.5% of robot activity.
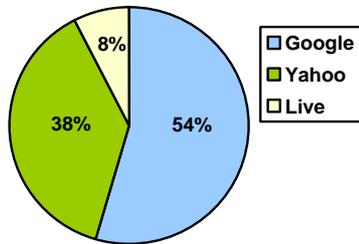

Figure 2. Search engine robot activity

V. DISCUSSIONS

One valuable and unexpected finding of the empirical study is that participation in PPC advertising programmes does not affect the indexing and ranking placement of a website in the search results of the major search engines. This discovery is backed by the fact that the Moyraverty site has never been indexed by any of the three search engines in this study. At several search engine marketing seminars, that we have attended, speakers have suggested that if a website URL were to be used in a Pay Per Click (PPC) advertising programme on the main search engines like Google, Yahoo and MSN(Live) then the indexing of the site was a certainty. The experiments have demonstrated this was not the case. If the finding does hold, surely it would be to the detriment of their advertising programme because as soon as a high ranking organic listing is achieved there is little or no need to pay for advertising.

Equally the finding that the most successful search engine submission method was the building of inward links from the social bookmarking sites is extraordinary and beyond expectation. We had expected there to be at least one entry from each of the 'folksonomy' sites. Actually the sites Reddit and Digg were the only ones present in the search results with both managing to achieve prominent positions in the search results for a short period of time. This does seems strange that no inward links were ever reported for a site that relied exclusively on inward linking to gain a listing in the Google and Yahoo search engines in the first place.

It's almost as if the search engines did not want to draw attention to the fact that inward linking from the folksonomy sites is an option. Perhaps with good sense otherwise more people would use this method as a form of spamming. Unfortunately, it was unable to tell whether or not the other folksonomy sites were ever used to access the Aldervale site. This was because an inward link search on Google and Yahoo consistently produced no results thus making it impossible to determine which folksonomy sites were impacting on the indexing and ranking placement of the Aldervale site.

The study has proved that the direct submission of a site URL to each search engine was by contrast an effective method of getting indexed in both Google and Yahoo but not MSN(Live). This was something of a surprise in that anecdotal evidence has suggested that the 'submit your site' pages were almost a placebo for inexperienced webmasters and they didn't actually do anything.

The experimental results have confirmed general public perception that Google is by far the most active search engine with Yahoo the second. A close examination reveals that Google is the most volatile of the three search engines with entries moving up and down the search results much more frequently than the more consistent Yahoo. It appears to take slightly longer to get a website indexed in Yahoo but once it is in the index it seems likely to stay there for longer. The MSN(Live) search engine is poor. Its robot called MSNBot only managed to perform 15 out of 199 crawls across all three test sites and the results returned during the study period were very poor and often dominated by illicit sites and newsfeeds. This is in conformity with some other studies' results that Microsoft is losing ground to both Google and Yahoo.

VI. CONCLUSIONS

In this paper we have developed a systematic methodology for an empirical study to evaluate whether or not active participation in PPC advertising programmes would improve the indexing and ranking placement of a website comparing to the organic search results. We have designed and conducted extensive experiments, and analysed the collected data during the study period. Initial results indicate that inward links embedded in social bookmarking sites are the most effective submission method and PPC advertising programmes does not at all affect the indexing and ranking placement of a website in the search results. Among these search engines selected for study Google is the most active search engine but its ranking placement in the result list are the most volatile. Yahoo, on the other hand, has the most stable ranking placement but is slow to get a website indexed.

In the future we plan to explore the impacts of the change of metadata in the <Meta> element on the ranking placement, and to investigate the quality of the search results to see which is the most accurate. More search engines such as BaiDu [24] could be included in the research to make experiment results widely applicable.

REFERENCES

[1] K.A. Houghton and H. Winklhofer, "The Effect of Website and E-Commerce Adoption on the Relationship between SMEs and their Export Intermediaries", International Small Business Journal, vol.2, no.4, pp.369, 2004.

[2] D.C. Green, "Search Engine Marketing: Why it benefits us all", Business Information Review, vol.20, no.4, pp.196, 2003.

[3] Generate business from your e-marketing plan, 2007. <http://www.nibusinessinfo.co.uk/bdotg/action/detail?r.l1=1073861197&r.l3=1075387753&r.t=RESOURCES&type=RESOURCES&site=191&itemId=1075390391&r.i=1075390432&r.l2=1073866263&r.s=sc>.

[4] D. Lewandowski, H. Wahlig, and G. Meyer-Bautor, "The freshness of web search engine databases", Journal of Information Science, vol.32, no.2, pp.142, 2005.

[5] K. Burnett, K.B. Ng and S.A. Park, "Comparison of the two traditions of Metadata development", Journal of the American Society for Information Science, vol.50, no.(13, pp.1209, 1999.

[6] Keywords for search engines, 2007. <http://www.w3schools.com/HTML/HTML_Meta.asp>.

[7] F. Harvey, "Paid searches find the route to online riches", The Financial Times, 1st July 2003.

[8] D.H. Goh and R.P. Ang, "Are pay for performance search engines relevant?", Journal of Information Science, vol.28, no.5, pp.349, 2002.

[9] Wikipedia, *Pay per click*, http://en.wikipedia.org/wiki/Pay_per_click, 2007.

[10] Web ad spend overtakes newspapers. 2007. http://news.bbc.co.uk/go/pr/fr/-/1/hi/business/6502773.stm

[11] Wikipedia, Search Engines http://en.wikipedia.org/wiki/Search_engine, 2007.

[12] M. Schwartz, "Search Engines, Spiders and Web Crawlers", Computerworld, 2000, http://www.accessmylibrary.com/coms2/summary_0286-27737930_ITM

[13] Centuar Publishing. 2007. Trends: Google sees almost three times as many searches as its rivals. New Media age.

[14] I. Watson, "Google gradually losing search engine market share", Sunday Business (London), 2006.

[15] D. Sullivan, "Submitting to directories: Yahoo & the Open Directory", 2007. http://searchenginewatch.com/showPage.HTML?page=2167881

[16] D. Sullivan, "Submitting to search crawlers: Google, Ask & Microsoft's Live search", http://searchenginewatch.com/showPage.HTML?page=2167871, 2007.

[17] C. Franklin, "How search engines work", 2007. http://computer.howstuffworks.com/search-engine2.htm

[18] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine", Computer Networks and ISDN Systems, vol.30, no.17, pp.107–117, 1998.

[19] J. Mostafa, "Seeking Better Web Searches", Scientific American [online]. http://sciam.com/print_version.cfm?articleID=0006304A-37F4-11E8-B7F483414B7F0000, 2005.

[20] HTML <Meta> tag, http://www.w3schools.com/tags/tag_Meta.asp, 2007

[21] Yahoo! Directory Submit, https://ecom.yahoo.com/dir/submit/intro/, 2007.

[22] T. Weinberg, "Does Google Pay Per Click Influence Organic Search Results?" http://www.seroundtable.com/archives/013662.HTML, 2007.

[23] Google dominates world search, http://news.bbc.co.uk/go/pr/fr//1/hi/technology/7039114.stm, 2007.

[24] BaiDu, http://www.baidu.com/