

# Black Hat SEO Crash Course V1.1

## Introduction

If you have spent any significant amount of time online, you have likely come across the term Black Hat at one time or another. This term is usually associated with many negative comments. This book is here to address those comments and provide some insight into the real life of a Black Hat SEO professional. To give you some background, my name is Brian. I've been involved in internet marketing for close to 10 years now, the last 7 of which have been dedicated to Black Hat SEO. As we will discuss shortly, you can't be a great Black Hat without first becoming a great White Hat marketer. With the formalities out of the way, lets get into the meat of things, shall we?

## What is Black Hat SEO?

The million dollar question that everyone has an opinion on. What exactly is Black Hat SEO? The answer here depends largely on who you ask. Ask most White Hats and they immediately quote the Google Webmaster Guidelines like a bunch of lemmings. Have you ever really stopped to think about it though? Google publishes those guidelines because they know as well as you and I that they have no way of detecting or preventing what they preach so loudly. They rely on droves of webmasters to blindly repeat everything they say because they are an internet powerhouse and they have everyone brainwashed into believing anything they tell them. This is actually a good thing though. It means that the vast majority of internet marketers and SEO professionals are completely blind to the vast array of tools at their disposal that not only increase traffic to their sites, but also make us all millions in revenue every year.

The second argument you are likely to hear is the age old ,“the search engines will ban your sites if you use Black Hat techniques”. Sure, this is true if you have no understanding of the basic principals or practices. If you jump in with no knowledge you are going to fail. I'll give you the secret though. Ready? Don't use black hat techniques on your White Hat domains. Not directly at least. You aren't going to build doorway or cloaked pages on your money site, that would be idiotic. Instead you buy several throw away domains, build your doorways on those and cloak/redirect the traffic to your money sites. You lose a doorway domain, who cares? Build 10 to replace it. It isn't rocket science, just common sense. A search engine can't possibly penalize you for outside influences that are beyond your control. They can't penalize you for incoming links, nor can they penalize you for sending traffic to your domain from other doorway pages outside of that domain. If they could, I would simply point doorway pages and spam links at my competitors to knock them out of the SERPS. See??? Common sense!

So again, what is Black Hat SEO? In my opinion, Black Hat SEO and White Hat SEO are almost no different. White hat web masters spend time carefully finding link partners to increase rankings for their keywords, Black Hats do the same thing, but we write automated scripts to do it while we sleep. White hat SEO's spend months perfecting the on page SEO of their sites for maximum rankings, black hat SEO's use content generators to spit out

thousands of generated pages to see which version works best. Are you starting to see a pattern here? You should, Black Hat SEO and White Hat SEO are one in the same with one key difference. Black Hats are lazy. We like things automated. Have you ever heard the phrase "Work smarter not harder?" We live by those words. Why spend weeks or months building pages only to have Google slap them down with some obscure penalty. If you have spent any time on web master forums you have heard that story time and time again. A web master plays by the rules, does nothing outwardly wrong or evil, yet their site is completely gone from the SERPS (Search Engine Results Pages) one morning for no apparent reason. It's frustrating, we've all been there. Months of work gone and nothing to show for it. I got tired of it as I am sure you are. That's when it came to me. Who elected the search engines the "internet police"? I certainly didn't, so why play by their rules? In the following pages I'm going to show you why the search engines rules make no sense, and further I'm going to discuss how you can use that information to your advantage.

## **What Makes A Good Content Generator?**

This is the foundation of Black Hat. Years ago, Black hat SEO consisted of throwing up pages with a keyword or phrase repeated hundreds of times. As search engines became more advanced, so did their spam detection. We evolved to more advanced techniques that included throwing random sentences together with the main keyword sprinkled around. Now the search engines had a far more difficult time determining if a page was spam or not. In recent years however, computing power has increased allowing search engines a far better understanding of the relationship between words and phrases. The result of this is an evolution in content generation. Content generators now must be able to identify and group together related words and phrases in such a way as to blend into natural speech.

One of the more commonly used text spinners is known as Markov. Markov isn't actually intended for content generation, it's actually something called a Markov Chain which was developed by mathematician Andrey Markov. The algorithm takes each word in a body of content and changes the order based on the algorithm. This produces largely unique text, but it's also typically VERY unreadable. The quality of the output really depends on the quality of the input. The other issue with Markov is the fact that it will likely never pass a human review for readability. If you don't shuffle the Markov chains enough you also run into duplicate content issues because of the nature of shingling as discussed earlier. Some people may be able to get around this by replacing words in the content with synonyms. I personally stopped using Markov back in 2006 or 2007 after developing my own proprietary content engine. Some popular software that uses Markov chains include [RSSGM](#) and [YAGC](#) both of which are pretty old and outdated at this point. They are worth taking a look at just to understand the fundamentals, but there are FAR better packages out there.

So, we've talked about the old methods of doing things, but this isn't 1999, you can't fool the search engines by simply repeating a keyword over and over in the body of your pages (I wish it were still that easy). So what works today? Now and in the future, LSI is becoming more and more important. LSI stands for Latent Semantic Indexing. It sounds complicated, but it really isn't. LSI is basically just a process by which a search engine can infer the meaning of a page based on the content of that page. For example, lets say they index a page and find words like atomic bomb, Manhattan Project, Germany, and Theory of Relativity. The idea is that the search engine can process those words, find relational data and

determine that the page is about Albert Einstein. So, ranking for a keyword phrase is no longer as simple as having content that talks about and repeats the target keyword phrase over and over like the good old days. Now we need to make sure we have other key phrases that the search engine thinks are related to the main key phrase.

This brings up the subject of duplicate content. We know what goes into a good content generator, but we have the problem of creating readable yet unique content. Let's take a look at duplicate content detection.

## **Duplicate Content**

I've read seemingly hundreds of forum posts discussing duplicate content, none of which gave the full picture, leaving me with more questions than answers. I decided to spend some time doing research to find out exactly what goes on behind the scenes. Here is what I have discovered.

Most people are under the assumption that duplicate content is looked at on the page level when in fact it is far more complex than that. Simply saying that "by changing 25 percent of the text on a page it is no longer duplicate content" is not a true or accurate statement. Lets examine why that is.

To gain some understanding we need to take a look at the k-shingle algorithm that may or may not be in use by the major search engines (my money is that it is in use). I've seen the following used as an example so lets use it here as well.

Let's suppose that you have a page that contains the following text:

**The swift brown fox jumped over the lazy dog.**

Before we get to this point the search engine has already stripped all tags and HTML from the page leaving just this plain text behind for us to take a look at.

The shingling algorithm essentially finds word groups within a body of text in order to determine the uniqueness of the text. The first thing they do is strip out all stop words like and, the, of, to. They also strip out all fill words, leaving us only with action words which are considered the core of the content. Once this is done the following "shingles" are created from the above text. (I'm going to include the stop words for simplicity)

**The swift brown fox  
swift brown fox jumped  
brown fox jumped over  
fox jumped over the  
jumped over the lazy  
over the lazy dog**

These are essentially like unique fingerprints that identify this block of text. The search engine can now compare this "fingerprint" to other pages in an attempt to find duplicate content. As duplicates are found a "duplicate content" score is assigned to the page. If too many "fingerprints" match other documents the score becomes high enough that the search engines

flag the page as duplicate content thus sending it to supplemental hell or worse deleting it from their index completely.

**My old lady swears that she saw the lazy dog jump over the swift brown fox.**

The above gives us the following shingles:

**my old lady swears  
old lady swears that  
lady swears that she  
swears that she saw  
that she saw the  
she saw the lazy  
saw the lazy dog  
the lazy dog jump  
lazy dog jump over  
dog jump over the  
jump over the swift  
over the swift brown  
the swift brown fox**

Comparing these two sets of shingles we can see that only one matches ("the swift brown fox"). Thus it is unlikely that these two documents are duplicates of one another. No one but Google knows what the percentage match must be for these two documents to be considered duplicates, but some thorough testing would sure narrow it down ;).

So what can we take away from the above examples? First and foremost we quickly begin to realize that duplicate content is far more difficult than saying "document A and document B are 50 percent similar". Second we can see that people adding "stop words" and "filler words" to avoid duplicate content are largely wasting their time. It's the "action" words that should be the focus. Changing action words without altering the meaning of a body of text may very well be enough to get past these algorithms. Then again there may be other mechanisms at work that we can't yet see rendering that impossible as well. I suggest experimenting and finding what works for you in your situation.

The last paragraph here is the real important part when generating content. You can't simply add generic stop words here and there and expect to fool anyone. Remember, we're dealing with a computer algorithm here, not some supernatural power. Everything you do should be from the standpoint of a scientist. Think through every decision using logic and reasoning. There is no magic involved in SEO, just raw data and numbers. Always split test and perform controlled experiments.

## **Cloaking**

So what is cloaking? Cloaking is simply showing different content to different people based on different criteria. Cloaking automatically gets a bad reputation, but that is based mostly on ignorance of how it works. There are many legitimate reasons to Cloak pages. In fact, even Google cloaks. Have you ever visited a web site with your cell phone and been automatically directed to the mobile version of the site? Guess what, that's cloaking. How about web pages

that automatically show you information based on your location? Guess what, that's cloaking. So, based on that, we can break cloaking down into two main categories, user agent cloaking and ip based cloaking (IP Delivery).

User Agent cloaking is simply a method of showing different pages or different content to visitors based on the user agent string they visit the site with. A user agent is simply an identifier that every web browser and search engine spider sends to a web server when they connect to a page. Above we used the example of a mobile phone. A Nokia cell phone for example will have a user agent similar to: User-Agent: Mozilla/5.0 (SymbianOS/9.1; U; [en]; Series60/3.0 NokiaE60/4.06.0) AppleWebKit/413 (KHTML, like Gecko) Safari/413

Knowing this, we can tell the difference between a mobile phone visiting our page and a regular visitor viewing our page with Internet Explorer or Firefox for example. We can then write a script that will show different information to those users based on their user agent.

Sounds good, doesn't it? Well, it works for basic things like mobile and non mobile versions of pages, but it's also very easy to detect, fool, and circumvent. Firefox for example has a handy plug-in that allows you to change your user agent string to anything you want. Using that plug-in I can make the script think that I am a Google search engine bot, thus rendering your cloaking completely useless. So, what else can we do if user agents are so easy to spoof?

## **IP Cloaking also known as IP Delivery**

Every visitor to your web site must first establish a connection with an ip address. These ip addresses resolve to dns servers which in turn identify the origin of that visitor. Every search engine crawler must identify itself with a unique signature viewable by reverse dns lookup. This means we have a sure fire method for identifying and cloaking based on ip address. This also means that we don't rely on the user agent at all, so there is no way to circumvent ip based cloaking (although some caution must be taken as we will discuss). The most difficult part of ip cloaking is compiling a list of known search engine ip's. Luckily software like [Blog Cloaker](#) and [SSEC](#) already do this for us. Once we have that information, we can then show different pages to different users based on the ip they visit our page with. For example, I can show a search engine bot a keyword targeted page full of key phrases related to what I want to rank for. When a human visits that same page I can show an ad, or an affiliate product so I can make some money. See the power and potential here?

So how can we detect ip cloaking? Every major search engine maintains a cache of the pages it indexes. This cache is going to contain the page as the search engine bot saw it at indexing time. This means your competition can view your cloaked page by clicking on the cache in the SERPS. That's ok, it's easy to get around that. The use of the meta tag noarchive in your pages forces the search engines to show no cached copy of your page in the search results, so you avoid snooping web masters. The only other method of detection involves ip spoofing, but that is a very difficult and time consuming thing to pull off. Basically you configure a computer to act as if it is using one of Google's ip's when it visits a page. This would allow you to connect as though you were a search engine bot, but the problem here is that the data for the page would be sent to the ip you are spoofing which isn't on your computer, so you are still out of luck.

The lesson here? If you are serious about this, use ip cloaking. It is very difficult to detect and by far the most solid option.

## Software

**SSEC or Simplified Search Engine Content**( <http://www.simplifiedsec.com/?=ebook> ):

This is one of the best IP delivery systems on the market. Their ip list is updated daily and contains close to 30,000 ip's. The member only forums are the best in the industry. The subscription is worth it just for the information contained there. The content engine is also top notch. It's flexible, so you can chose to use their proprietary scraped content system which automatically scrapes search engines for your content, or you can use custom content similar in fashion to SEC above, but faster. You can also mix and match the content sources giving you the ultimate in control. This is the only software as of this writing that takes LSI into account directly from within the content engine. This is also the fastest page builder I have come across. You can easily put together several thousand sites each with hundreds of pages of content in just a few hours. Support is top notch, and the knowledgeable staff really knows what they are talking about. This one gets a gold star from me.

### **Fantomaster Shadowmaker:**

This is probably one of the oldest and most commonly known high end cloaking packages being sold. It's also one of the most out of date. For \$3,000.00 you basically get a clunky outdated interface for slowly building HTML pages. I know, I'm being harsh, but I was really let down by this software. The content engine doesn't do anything to address LSI. It simply splices unrelated sentences together from random sources while tossing in your keyword randomly. Unless things change drastically I would avoid this one. This software probably worked great when it was developed back in 1999, but today it leaves much to be desired.

### **SEC (Search Engine Cloaker):**

Another well known paid script. This one is of good quality and with work does provide results. The content engine is mostly manual making you build sentences which are then mixed together for your content. If you understand SEO and have the time to dedicate to creating the content, the pages built last a long time. I do have two complaints. The software is SLOW. It takes days just to setup a few decent pages. That in itself isn't very black hat. Remember, we're lazy! The other gripe is the ip cloaking. Their ip list is terribly out of date only containing a couple thousand ip's as of this writing. Rumor has it that the developers are MIA meaning updates are unlikely.

### **Blog Cloaker** ( <http://www.simplifiedses.com/?=ebook> ):

Another solid offering from the guys that developed SSEC. This is the natural evolution of that software. This mass site builder is based around wordpress blogs. This software is the best in the industry hands down. The interface has the feel of a system developed by real professionals. You have the same content options seen in SSEC, but with several different redirection types including header redirection, JavaScript, meta refresh, and even iframe. This again is an ip cloaking solution with the same industry leading ip list as SSEC. The monthly

subscription may seem daunting at first, but the price of admission is worth every penny if you are serious about making money in this industry. It literally does not get any better than this.

### **BlogSolution:**

Sold as an automated blog builder, BlogSolution falls short in almost every important area. The blogs created are not wordpress blogs, but rather a proprietary blog software specifically written for BlogSolution. This “feature” means your blogs stand out like a sore thumb in the eyes of the search engines. They don't blend in at all leaving footprints all over the place. The licensing limits you to 100 blogs which basically means you can't build enough to make any decent amount of money. The content engine is a joke as well using rss feeds and leaving you with a bunch of easy to detect duplicate content blogs that rank for nothing.

### **Link Building**

As we discussed earlier, Black Hats are Basically White Hats, only lazy! As we build pages, we also need links to get those pages to rank. Lets discuss some common and not so common methods for doing so.

### **Blog ping:**

This one is quite old, but still widely used. Blog indexing services setup a protocol in which a web site can send a ping whenever new pages are added to a blog. They can then send over a bot that grabs the page content for indexing and searching, or simply to add as a link in their blog directory. Black Hats exploit this by writing scripts that send out massive numbers of pings to various services in order to entice bots to crawl their pages. This method certainly drives the bots, but in the last couple years it has lost most of its power as far as getting pages to rank. Still a powerful indexing tool, but be sure to supplement the results with some real backlinks.

### **Trackbacks:**

Another method of communication used by blogs, trackbacks are basically a method in which one blog can tell another blog that it has posted something related to or in response to an existing blog post. As a black hat, we see that as an opportunity to inject links to thousands of our own pages by automating the process and sending out trackbacks to as many blogs as we can. Most blogs these days have software in place that greatly limits or even eliminates trackback spam, but it's still a viable tool. The real key is to blend in and avoid being caught by spam filters. To do that, you need to actually post content related to the original post. [SSEC](#) automates this by searching for blogs directly related to each of your keyword pages. Once found, the software posts a trackback with related content and a link back to your page. Methods like this avoid spam detection and also give you a nice themed link. These links are two way, so don't expect them to be as powerful as a non reciprocal one way link.

Most people are not aware of the ability to quickly and easily find link partners in search engines using simple search patterns. I'm going to share some pointers here.

First I will post a simple way to find blogs that allow trackbacks:

keyword "TrackBack URL for this entry"  
keyword "Trackback address for this post"  
keyword "index.php/trackback"  
keyword "wp-trackback.php"

In the above examples you simply place the phrase you are searching for in place of keyword.  
For example

cancer "TrackBack URL for this entry"

This searches google for the word cancer, but also requires that the phrase TrackBack URL for this entry be included on the page as well. You can use patterns like this to find blogs, guestbooks, etc quickly and easily. You can even narrow down your results by top level domain extension. For example:

"keyword phrase" site:.org

This finds the phrase "keyword phrase" in the search engine, and limits the results to only .org domains.

Here are a few more examples you can play around with for guest books and places to comment spam.

inurl:guestbook.php keyword  
inurl:gbook.php keyword  
inurl:light.cgi keyword

inurl:suggest\_link.php  
inurl:add\_url.php  
inurl:add\_post.php

"PHP Guestbook" inurl:ardguest.php" +keyword  
phpBook Ver inurl:guestbook.php +keyword  
"Achim Winkler" inurl:guestbook.php +keyword  
"KISGB" inurl:kisgb -inurl:.html "public entries" +keyword  
"powered by xeobook" admin +keyword

or just a good old

"Powered by nameofscript"

When you visit blogs, guestbooks, forums, etc they almost always have patterns or footprints. Something that is the same on each page. Check the footer, check the submission forms. Look for these phrases and it will help you better target your searches and in turn help deliver a larger number of potential link partners.

**EDU links:**

A couple years ago Black Hats noticed an odd trend. Universities and government agencies with very high ranking web sites often times have very old message boards they have long forgotten about, but that still have public access. We took advantage of that by posting millions of links to our pages on these abandoned sites. This gave a HUGE boost to rankings and made some very lucky Viagra spammers millions of dollars. The effectiveness of this approach has diminished over time, but the power is still there.

So how do you find these links?

Simple!

Go to google and search for the following include the quotes:

**"Discussion Submission Form" site:.edu**

Change the .edu to .gov or .org to get other top level domains from the same (it will work with any domain extension so be creative). Now change your google settings to show 100 links per page. Copy and paste that url into the form on the dashboard and you just entered 100 spammable message board url's. Here is another search that works.

**"Requirements Discussion Submission Form" site:.edu**

There are others, you just have to examine the pages you find. Check the source code and find a common foot print. Once you do simply modify the above search with your new found footprint text. Make sure it is returning the message board submission forms like these are, then submit your links. Doesn't get much easier than that.

**Forums and Guest books:**

The internet contains millions of forums and guest books all ripe for the picking. While most forums are heavily moderated (at least the active ones), that still leaves you with thousands in which you can drop links where no one will likely notice or even care. We're talking about abandoned forums, old guest books, etc. Now, you can get links dropped on active forums as well, but it takes some more creativity. Putting up a post related to the topic on the forum and dropping your link in the BB code for a smiley for example. Software packages like Xrumer made this a VERY popular way to gather back links. So much so that most forums have methods in place to detect and reject these types of links. Some people still use them and are still quite successful. The key here is volume. Submit enough links and you are bound to find some gold.

**Link Networks:**

Also known as link farms, these have been popular for years. Most are very simplistic in nature. Page A links to page B, page B links to page C, then back to A. These are pretty easy to detect because of the limited range of ip's involved. It doesn't take much processing to figure out that there are only a few people involved with all of the links. So, the key here is to have a very diverse pool of links. Take a look at [Link Exchange](#) for example. They have over 300 servers all over the world with thousands of ip's, so it would be almost impossible to detect. A search engine would have to discount links completely in order to filter these links

out. Another option is to build a large diverse network of blogs, forums and directories all spread across different servers and ip's. This gives you a large network with which you can create some good one way links. This avoids the label of a link farm in most cases, but again, the key here is in the diversity of the sites from which the links originate.

### **Exploiting the social web for links:**

The web 2.0 craze started a couple years ago, and with it came more social interaction on web sites. Sites like [delicious](#) , Digg, Pligg, and hundreds of others allow members to join and interact with the site in various ways. Many of these sites allow outside links and content to be published. Digg and Pligg sites (also known as social media sites) allow you to submit news stories which in turn provide valuable links back to your pages. Now normally these links are to white hat sites, but we can just as easily exploit this for black hat use. Of course, as with everything else, the key is automation. Software like [Bookmarking Demon](#) automates the process by signing up for and posting links to over 100 social media and bookmarking sites. This saves you hours of work and provides hundreds if not thousands of one way incoming links.

### **Money Making Strategies**

We now have a solid understanding of cloaking, how a search engine works, content generation, software to avoid, software that is pure gold and even link building strategies. So how do you pull all of it together to make some money?

### **Arbitrage and ppc:**

Think of this as buy low, sell high. It's simple in nature, slightly more complex in practice. The idea here is that you are using a site building program of your own creation, or one from the list we discussed earlier to build pages with long tail keywords that are then redirected to a money page with a higher price, but related keyword. That's the key here. So lets take SSEC for example. I head over to my keyword tool of choice and type in my high priced keyword. Lets use red widget as an example. My keyword tool gives me hundreds of long tail variations of the keyword red widget which I then use to setup my SSEC cloaked pages. My pages get built on several throw away domains, then the human traffic gets redirected to a landing page focused on my main keyword red widget which is of course loaded with ads and looks completely White Hat. I get my clicks, make money, rinse and repeat. Now, lets assume you are doing this with thousands of main money keywords because one of the keys to black hat is volume. Well, you need to be able to build these landing pages quickly and easily, right? After all, Black hat's are lazy. So, you find a good tool. In this example I use a program called [Landing Page Builder](#) . Landing Page Builder dynamically generates landing pages based on the traffic you send it. You load up your money keyword list, setup a template with your ads or offers, then send all of your doorway/cloaked traffic to the index page. The Landing Page Builder shows the best possible page with ads based on what the incoming user searched for. Couldn't be easier, and it automates the difficult tasks we all hate. This is the same method all of the PPC arbitrage people use to make a killing, but the difference here is that our traffic is free and natural instead of paid. I've personally made well over \$100,000 with this exact method.

## **Affiliate Marketing:**

We all know what an affiliate program is. There are literally tens of thousands of affiliate programs with millions of products to sell. The most difficult part of affiliate marketing is getting well qualified targeted traffic. That again is where good software and cloaking comes into play. Some networks and affiliates allow direct linking. Direct Linking is where you setup your cloaked pages with all of your product keywords, then redirect straight to the merchant or affiliates sales page. This often results in the highest conversion rates, but as I said, some affiliates don't allow Direct Linking. So, again, that's where Landing Pages come in. Either building your own (which we are far too lazy to do), or by using something like Landing Page Builder which automates everything for us. Landing pages give us a place to send and clean our traffic, they also prequalify the buyer and make sure the quality of the traffic sent to the affiliate is as high as possible. After all, we want to make money, but we also want to keep a strong relationship with the affiliate so we can get paid.

I can't stress the importance of diversification enough here. Never put all of your eggs in a single basket. Find a network or merchant that converts, and expand on that. Find several merchants in that same niche so you always have a steady stream of income if one merchant should cancel your account.

## **Niche Selection:**

Niche selection plays an important role in any type of online marketing be it white hat or black hat. You have to take a moment and think about your niche as a consumer would. Take large appliances for example. It's unlikely that a shopper will purchase a new refrigerator online. They will probably research the item online, but the final purchase is likely to be made in store. Items like computer parts, cosmetics, and even clothing are far more likely to be purchased online, especially if the price online is lower. Take some time to research your target audience, think like a consumer and your niche selection is easy.

## **Keywords versus data feeds:**

Now that we have our niche selected, we need to have a method for promoting it. Many affiliates offer up product data feeds which gives you all of the info you need with little or no effort. The drawback to that is the fact that every one of your competitors has that same info, so it diminishes the value and increases the competition. Not to say data feeds don't work, quite the contrary, they can work very well. The point to take away from this is that there is likely an increased level of competition in the SERPS for the keywords and phrases found in these data feeds. This is where doing your own keyword research can pay off. By putting in the time and effort required to research your own keywords, you are likely to drive more traffic to your offers. Now, in order to turn that traffic into conversions you need to pay attention to the quality of that traffic. Avoid unrelated or broad keywords. This again is where the long tail comes into play.

## **Conclusion:**

As we can see, Black Hat Marketing isn't all that different from White Hat marketing. We automate the difficult and time consuming tasks so we can focus on the important tasks at

hand. I would like to thank you for taking the time to read this. I plan to update often. In the mean time, you can follow me on my personal blog over at <http://www.blackhat360.com/?=ebook> . Be sure to register and post on the forums <http://www.blackhat360.com/forums/?=ebook> if you have any comments or questions.