

---

## UNIT 15 SEARCH ENGINES AND THEIR ABUSE

---

### Structure

- 15.1 Introduction
- 15.2 Objectives
- 15.3 What are Search Engines?
- 15.4 The Process: How a Search Engine Works
  - 15.4.1 Web Crawling
  - 15.4.2 Indexing
  - 15.4.3 Searching
  - 15.4.4 Ranking of Web Pages by Search Engine
- 15.5 Abuse of the Process: Spamdexing
  - 15.5.1 Hidden or Invisible Text
  - 15.5.2 Keyword Stuffing
  - 15.5.3 Meta Tag Stuffing
  - 15.5.4 Cloaking
  - 15.5.5 Page Hijacking
  - 15.5.6 Link Spam
- 15.6 Controlling Abuse of Searching Process through Law
- 15.7 Keyword-Linked Advertising and Trademark Infringement
- 15.8 Summary
- 15.9 Terminal Questions
- 15.10 Answers and Hints

---

### 15.1 INTRODUCTION

---

Information and Internet are synonyms. It is difficult to conceive the gigantic quantum of information present on the Internet. It runs in many billions of pages and still growing. According to some estimates the Internet is doubling its size every four months and with this increase in size comes a proportional increase in information.

Search engines have come to acquire a unique and important position on the overall Internet system. They have become the presenters of information that is provided by others. It is really interesting to know where all this information comes from or how it is being sorted and selected by those who make it available to us.

All good search engines index hundreds of millions of pages, and respond to tens of millions of queries each day. Google being the most used search engine today, in a lot of respects 'Googling' has become synonymous with searching on the Internet. It has become an autonomous concept and an independent form of leisure activity, similar to 'zapping' through television channels. Anybody who cannot be found via Google does not exist because of the Internet proverb, "if you are not indexed by a search engine you simply do not exist".

As search engine traffic achieved greater significance in online marketing plans, consultants were brought in who were well versed in how search engines perceive a web site. These consultants used a variety of techniques (legitimate and otherwise) to improve ranking for their clients. The disputes arose when people started using dishonest methods to manipulate the result of a search engine so as to get the best ranking on a search engine.

How far are search engines regulated? Is any regulation necessary or desirable? Search engines hardly find a mention in legal circles. However, search results that are tendered by a search engine upon a query could often be manipulated and distorted. The ones who obtain the highest position on a search engine may be using clever tactics to mislead search engines. Legal consideration of various issues surrounding search engines is indeed necessary to counter the abuse and to make search engines more transparent and dependable and trustworthy. It is important for users and consumers who depend on the information supplied by the search engines to understand and be aware of the working of search engines.

---

## 15.2 OBJECTIVES

---

After reading this unit, you should be able to:

- describe the term search engine;
- explain the working of a search engine;
- explain the different ways of the abuse of the process; and
- discuss the ways how abuses of search engine be controlled through law.

---

## 15.3 WHAT ARE SEARCH ENGINES?

---

A search engine is a program designed to help find information stored on a computer system such as the World Wide Web, or 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 references that match those criteria.

World Wide Web search engines are special sites that are intended and designed to help people find information stored on other sites. A search engine's main function is that of enabling access; it is a gateway to possibly relevant information on the Internet.

Let us analogise the Internet of today with a giant multidisciplinary library. The library will be of little use to anyone without a good cataloguing and indexing system. This system guides a visitor to the labyrinths of the library. When we consider Internet as a giant library then search engines take up the role of catalogues and librarians together. So, we are not alone when we look towards the Internet for a particular news or material. There are search engines to help us in our task and they are much faster in their operation than any other catalogue that we have known till now. In this manner search engines are responsible for making information on the Internet easily accessible. Search engines play an important role in information synthesis and acquisition in today's information society.

Please answer the following Self Assessment Question.

|                                   |                     |
|-----------------------------------|---------------------|
| <b>Self Assessment Question 1</b> | <i>Spend 3 Min.</i> |
| Define the term search engine?    |                     |
| .....                             |                     |
| .....                             |                     |
| .....                             |                     |
| .....                             |                     |
| .....                             |                     |
| .....                             |                     |

---

## 15.4 THE PROCESS: HOW A SEARCH ENGINE WORKS

---

There are mainly two ways in which people obtain access to any content on the Internet; first, they may either type in the URL of a web site on the location bar of their browser. The browser then takes you to the desired web page containing the content that you were looking for. The other way is to use a search engine. Search engines have become an integral part of the process through which virtually all Internet users seek to access information in cyberspace. Many times searching desired content may be the only option available with the users.

It is interesting to know how a search engine works. When you type an expression say “law schools in India” in the Google search engine, it gives you about 29,900,000 results and ranks them according to some criterion and it gives you this list in a flash — it takes hardly a second. Is the search engine attempting to search the entire WWW when you ask for a particular expression? Definitely no. Had it been so it would not have taken so less a time to complete the search process. But then what is the process? The process is essentially composed of three elements — web crawling, indexing and searching.

### 15.4.1 Web Crawling

This automated process uses intelligent ‘sleuths’ called spiders, bots or crawlers. These sleuths surf the Internet using criteria set previously by the search-engine provider. Their task is to find various words and expressions that have been used on the Internet and also noting on which web site they appear. Web search engines work by storing information about a large number of web pages, which they retrieve from the WWW itself.

### 15.4.2 Indexing

After web crawling is done, indexing determines what counts as relevant words or combinations of words; and non-relevant information, such as fillers and punctuation marks, is deleted. The contents of each page are then analysed to determine how it should be indexed (for example, words are extracted from the titles, headings, or special fields called meta tags). Data about web pages is stored in an index database for use in later queries. This enriched information forms the ultimate basic material

for the search engine. Some search engines, such as Google, store all or part of the source page (referred to as a cache) as well as information about the web pages, whereas some store every word of every page it finds, such as AltaVista. By no means all the information that is present on the Internet is found and indexed by search engines. According to some estimates, individual search engines index only about 15% of all the information present on the Internet and all the search engines together cover no more than 40% of all the available information.

Please answer the following Self Assessment Question.

|  |                     |
|--|---------------------|
| <b>Self Assessment Question 2</b>                  | <i>Spend 3 Min.</i> |
| Name the different ways how a search engine works? |                     |
| .....  |                     |
| .....  |                     |
| .....  |                     |
| .....  |                     |
| .....  |                     |
| .....  |                     |

### 15.4.3 Searching

When a user comes to the search engine and makes a query, typically by giving key words, the engine looks up the index and provides a listing of best-matching web pages according to its criteria, usually with a short summary containing the document’s title and sometimes parts of the text.

### 15.4.4 Ranking of Web Pages by Search Engine

The usefulness of a search engine depends on the relevance of the results it gives back. While there may be millions of Web pages that include a particular word or phrase, some pages may be more relevant, popular, or authoritative than others. Most search engines employ methods to rank the results to provide the ‘best’ results first. How a search engine decides which pages are the best matches, and what order the results should be shown in, varies widely from one engine to another. The methods also change over time as Internet usage changes and new techniques evolve.

Google is currently the most popular search engine. Its success was based in part on the concept of link popularity and PageRank. How many other web sites and web pages link to a given page is taken into consideration with PageRank, on the premise that good or desirable pages are linked to more than others. The PageRank of linking pages and the number of links on these pages contribute to the PageRank of the linked page. Google and most other web engines utilize not only PageRank but more than 150 criteria to determine relevancy.

---

## 15.5 ABUSE OF THE PROCESS: SPAMDEXING

---

‘Search engine optimisation’ is a set of methods aimed at improving the ranking of a website in search engine listings. Search engine optimisation practices should legitimately focus on building better sites, and using honest methods of promotion.

But eager to achieve the highest rankings sometimes the whole process of search engine optimisation is abused. Moreover, what constitutes an honest, or ethical, method is an issue that has been the subject of numerous debates. This abuse of the process is known as spamdexing. ‘Spamdexing’ is a combination of two words ‘spamming’ and ‘indexing’.

Spamdexing or search engine spamming is the practice of deliberately and dishonestly manipulating search engines to increase the chance of a web site or a Web page being placed close to the beginning of search engine results, or to influence the category to which the page is assigned. Since high ranking is what is desired by most web sites, many designers of web pages try to get a good ranking in search engines and design their pages accordingly. Spamdexing refers exclusively to practices that are dishonest and mislead search and indexing programs to give a page a ranking it does not deserve. It has to be seen in contradistinction of search engine optimisation which includes techniques for making a website indexable by search engines, without misleading the indexation process. Following are some of the ways in which spamdexing is practiced:

Please answer the following Self Assessment Question.

|                                   |                     |
|-----------------------------------|---------------------|
| <b>Self Assessment Question 3</b> | <i>Spend 3 Min.</i> |
| What is spamdexing?               |                     |
| .....                             |                     |
| .....                             |                     |
| .....                             |                     |
| .....                             |                     |
| .....                             |                     |
| .....                             |                     |

### 15.5.1 Hidden or Invisible Text

One way of fooling a search engine is to match certain keywords by making them the same colour as the background or by using a tiny font size or hiding them within the HTML code such as “no frame” sections. This is useful to make a page appear to be relevant for a web crawler in a way that makes it more likely to be found. Hiding text out of view of the visitor could be done in many different ways. Text coloured to blend with the background, positioning keywords to place text “behind” an image & mash; and therefore out of view of the visitor are all common techniques. But as of 2005, some of these invisible text techniques can be detected by major search engines.

### 15.5.2 Keyword Stuffing

Keyword stuffing is considered to be an unethical Search engine optimisation technique. Keyword stuffing occurs when a web page is loaded with keywords in the content. This involves the insertion of hidden, random text on a webpage to raise the keyword density or ratio of keywords to other words on the page. Keyword stuffing used is to obtain maximum search engine ranking and visibility for particular phrases. Inserted text sometimes includes words that are frequently searched such

as 'sex' even if those terms bear little connection to the content of a page, in order to attract traffic to advert-driven pages.

Older versions of indexing programs simply counted how often a keyword appeared, and used to determine relevance levels. A word that is repeated too often may raise a red flag to search engines. Most modern search engines have the ability to analyse a page for keyword stuffing and determine whether the frequency is above a 'normal' level.

### 15.5.3 Meta Tag Stuffing

A meta tag is a coding statement in the Hypertext Markup Language (HTML) that describes some aspect of the contents of a Web page. The meta tag is placed near the top of the HTML in a Web page as part of the heading. The information that you provide in a meta tag could be used by search engines to index a page so that someone searching for the kind of information the page contains will be able to find it. There are several kinds of meta tags, but the most important for search engine indexing are the keywords meta tag and the description meta tag. The keywords meta tag lists the words or phrases that best describe the contents of the page. The description meta tag includes a brief one- or two-sentence description of the page. Meta tagging could also be used with a dishonest intention whereby a web site uses a trademark of another business entity thereby wishing to divert web traffic to his own site.

Meta tags such as these have been the focus of a field of marketing research in search engine optimisation. In the mid to late 1990s, search engines were reliant on meta tag data to correctly classify a web page. Webmasters quickly learned the commercial significance of having the right meta tag, as it frequently led to a high ranking in the search engines - and thus, high traffic to the web site. In the early 2000s, search engines have veered away from reliance on meta tags since many web sites used inappropriate keywords in order to increase their search engine ranking. Many search engines, however, still take meta tags into consideration when delivering results. Currently, Google does not use meta tags to index web sites. Techniques have also been developed by some search engines in order to penalize web sites considered to be cheating the system. For example, a website repeating the same keyword several times may have its ranking decreased by a search engine trying to eliminate this practice.

### 15.5.4 Cloaking

Cloaking is another search engine optimisation technique in which the content presented to the search engine crawler is different from that presented to the users' browser. cloaking is often used as a spamdexing technique, to try to trick search engines into giving the relevant site a higher ranking; it could also be used to trick search engine users into visiting a site based on the search engine description which turns out to have substantially different – or even illegal – content.

Primarily when a web page is designed it is designed keeping in mind the end user. But now, in order to get the attention of a search engine, webmasters also keep in mind the search engine while designing a web page. This results in pages with too many keywords and other factors that might be search engine 'friendly', but make the pages inconvenient for actual end users to consume. So, designers sometimes design pages solely for the search engines backed by pages for actual end users. As

such, cloaking is an important technique to allow webmasters to split their efforts and separately target the search engine spiders and end users. This technique also has a potential to be used illegally.

### 15.5.5 Page Hijacking

Page hijacking is a form of spamdexing. It is the act of copying a random but popular webpage with the intent to feed the copied page to web crawlers. The intent is that the copied page appears in search engine results, and when the users click on it, the visitors are redirected to a different, often unrelated, website. Page hijacking is a form of cloaking, and it is possible because web crawlers detect duplicates as they index web pages, and if two pages have the same content, they keep only one of the URLs and reject the other. So, there is a chance that the original Web Page is rejected by a search engine while taking the duplicate.

### 15.5.6 Link Spam

Link spam takes advantage of link-based ranking algorithms, such as Google's PageRank algorithm, which gives a higher ranking to a website the more other highly-ranked websites link to it. This has led to another spamdexing practice known as link farm. On the World Wide Web, a link farm is any group of web pages that all hyperlink to every other page in the group. Search engines such as Google recommend that webmasters request relevant links to their sites that is conduct a link campaign, but avoid participating in link farms. Search engines try to identify specific attributes associated with link farm pages and filter those pages from indexing and search results. In some cases, entire domains are removed from the search engine indexes in order to prevent them from influencing search results.

Please answer the following Self Assessment Questions.

|                                   |                         |
|-----------------------------------|-------------------------|
| <b>Self Assessment Question 4</b> | <i>Spend 3 Min.</i>     |
| a) Define cloaking?               | .....<br>.....<br>..... |
| b) Define page hijacking?         | .....<br>.....<br>..... |

---

## 15.6 CONTROLLING ABUSE OF SEARCHING PROCESS THROUGH LAW

---

Abuse of searching process or various forms of spamdexing have been countered by the owners of trademarks by using the trademark doctrine of “initial interest confusion”. While the benchmark of trademark infringement traditionally has been a demonstration that consumers are likely to be confused by the use of a similar or identical trademark to identify the goods or services of another, the doctrine called

initial interest confusion allows liability for trademark infringement solely on the basis that a consumer might initially be interested, attracted, or distracted by a competitor's, or even a non-competitor's, product or service.

In a hypothetical Internet situation a user searches for a trademark 'Tata' but by using the above stated spamdexing tools, like stuffing the trademark Tata within the contents of his site or by including it in its meta tags, a competitor achieves a search engine listing even higher than the actual site of Tata. Now, the user who attracts web traffic which was meant for Tata is not ultimately confusing the user as when the user visits his site he comes to know that it is not that of Tata's. But the user may settle to buy competing products from that site only. Such a situation could be an actionable claim under this theory of initial interest confusion. Initial interest confusion is being used with increasing frequency, especially on the Internet, to shut down speech critical of trademark holders and their products and services, to prevent comparative advertisements and to otherwise limit information and choices available to consumers.

In *Brookfield Communications v. West Coast Entertainment* [174 F.3d 1036, 1062 (9th Cir. 1999).], West Coast Entertainment registered the domain name [www.moviebuff.com](http://www.moviebuff.com) despite the fact that Brookfield had a registered trademark for its entertainment industry software product 'MovieBuff.' The United States Court of Appeals for the Ninth Circuit found that although users arriving at West Coast's [moviebuff.com](http://moviebuff.com) website would realise that the site was not associated with Brookfield's MovieBuff software, the initial interest confusion caused by West Coast's use of the mark caused some damage to Brookfield. The court noted that "the use of another's trademark in a manner calculated 'to capture initial consumer attention, even though no actual sale is finally completed as a result of the confusion, may be still an infringement'. This holding shows that any action taken to divert the consumer's attention to one's product by using the name and reputation of a competitor's product is improper for the fact that there is only initial consumer confusion does not alter the fact that the infringer would be misappropriating the trademark-holder's acquired goodwill.

But it is not always that a party alleging initial interest confusion will be successful in a court of law. In *Playboy Enterprises, Inc. v. Terri Welles* [279 F.3d 796, 801 (9th Cir. 2002)] the defendant was a former Playboy playmate who described herself as such on her personal website. Playboy asserted that the use of the 'Playboy' and 'Playmate' marks may give visitors to Welles' site the wrongful impression that the site was sponsored by Playboy. The court thought otherwise and held that the use of Playboy's marks both on the website and in the website's metatags was purely nominative and therefore fair use.

---

## 15.7 KEYWORD-LINKED ADVERTISING AND TRADEMARK INFRINGEMENT

---

Keywords advertisement is a practice adopted by search engines which allows one to display one's advertisements along with search results about a particular keyword. Keyword-linked advertising is a way to provide relevant, client-specific advertisements to Internet users. When visitors to a search engine enter their search terms, advertisements are placed on the search results page depending on the search terms entered. For example, a user is seeking to search for "sports shoes in India". A few shoe making companies have placed their bid for such keywords and their



advertisements are displayed along with the search results. This is quite a harmless exercise. But the difficulty arises only when someone is searching for a trademark say “Bata” and a competing shoe making company has bid its own advertisements to appear along with the search results.

A major source of income for many of the most popular search engines is the sale of advertisements placed on the search results page. In April 2004 Google began to allow advertisers to purchase the right to have their advertisements linked to trademarked terms entered into the search field, even if the advertisers were not the trademark holders. However, the search engines’ practice of selecting advertisements to display based on the search terms entered by the user has recently come under fire for being violative of trademarks owner rights. Trademark owners are unhappy when competitors’ advertisements are linked to the owner’s trademarks. They allege that search engines are allowing competitors to take unfair advantage of interest associated with the marks.

But does it mean that all kinds of advertisement linked to keyword search are unlawful. It may be unlawful in certain situations and perfectly in harmony with trademark law in other. The primary issue in online advertising cases tends to be whether the advertisement creates confusion as to its source. The likelihood of confusion plays an important role in determining whether a competitor’s use of a mark is an infringing use. Courts internationally have found that the initial confusion as to the source of a product or service may still be enough to assess liability for trademark infringement. The test in *Brookfield case* is whether the accused infringer used the plaintiff’s mark “in a manner calculated to capture initial consumer attention” would be applicable in such situations as well.

In the context of keyword-linked search results, infringing advertisements are the ones that fail to identify the true source of the advertisement, either by falsely identifying the advertisement as being from the trademark holder or by giving no indication as to the source of the advertisement. If there is no uncertainty as to the source of the advertisements, there is no likelihood of confusion.

Since the search engines are not themselves performing the infringing activity, the basis for liability would be as a contributor to infringement. The search engine would be liable if it intentionally induces another to infringe a trademark, or if it continues to supply its product to one whom it knows or has reason to know is engaging in trademark infringement. Many search engines themselves have come forward to help the trademark owners in this regard. Google has been willing to perform a limited investigation and respect ‘reasonable’ requests to remove trademark terms from the bidding process. Search engine provider Yahoo limits its keyword-linked advertisement sales even further, only permitting use by an advertiser whose website refers to the trademark or its owner in a permissive nominative manner without creating confusion, or if the keyword is generic or merely descriptive. Such practices that take an active, although responsive, approach to preventing infringement could ensure that legitimate competitive uses of the keyword-linked advertising are protected.

Let us now summarize the points covered in this unit.

---

## 15.8 SUMMARY

---

- Search engines have come to acquire a unique and important position on the overall Internet system.

- World Wide Web search engines are special sites that are intended and designed to help people find information stored on other sites.
- A search engine's functioning is essentially composed of three elements—web crawling, indexing and searching.
- Spamdexing refers exclusively to practices that are dishonest and mislead search and indexing programs to give a page a ranking it does not deserve.
- Keyword stuffing, meta tag stuffing, cloaking, page hijacking and link spam are some of the means through which spamdexing is practised.
- Through the trademark doctrine of 'initial interest confusion' abuse of search engine's process can be countered.

---

## 15.9 TERMINAL QUESTIONS

---

1. Why getting good ranking on search engines is so important in today's business world?
2. How the process of search engines can be abused and what all remedies are available to counter the same?
3. In which situations keyword linked advertising can lead to trademark infringement?

---

## 15.10 ANSWERS AND HINTS

---

### Self Assessment Questions

1. A search engine is a program designed to help find information stored on a computer system such as the world wide web or a personal computer.
2. Web crawling, Indexing, Searching.
3. Spamdexing or search engine spamdexing is the practice of differently and dishonestly manipulating search engine to increase the chance of a web site of a web page being placed close to the beginning of search engine results to influence the category to which the page is assigned.
4.
  - a) Cloaking is search engine optimisation technique in which the content presented to the search engine is different from that presented to the user's browser.
  - b) Page hijacking is the act of copying a random but popular web page with the intent to feed the copied page to web crawling.

### Terminal Questions

1. Refer to sub section 15.3.4 of the unit.
2. Refer to sections 15.4 and 15.5 of the unit.
3. Refer to section 15.6 of the unit.