Bad things can happen on your site

- WHITEPAPER -

Be smart. You must know what must be done on your site and what not.
This is a list of the most common issues that sites have, a checklist for negative things on your site.

1. Duplicate content

Most of the sites tend to have this problem and can be most of the times be filtered out of the google search results. However due to Penguin, Panda and all sorts of recent updates, duplicate content on sites can cause major stress.

Let's look at the **causes that can generate duplicate content:**

- If you decide to copy information from other sites than the duplicate content is obvious. Don't do that! It's bad for Google, it's bad for you too.

- Duplicate content can be generated by technical issues:
  - **Language pages like:**
    http://www.example.com/what-is-seo.html
    http://www.example.com/en/what-is-seo.html

  - **Capitalization of words in URL**
    http://www.example.com/what-is-seo.html
    http://www.example.com/en/what-is-SEO.html

- o **http and https**
  http://www.example.com/what-is-seo.html
  https://www.example.com/what-is-seo.html

- o **Linking to both the url-friendly link and to the link with parameters. The url friendly link below is generated through .htaccess:**
  http://www.example.com/index.php?article_id=123
  http://www.example.com/my-new-article-123.html

- o **Same product in more that one category. Many ecommerce sites tend to have this one.**
  http://www.example.com/electronics/mp3-player-sony.html
  http://www.example.com/audio/mp3-player-sony.html
  If you find yourself in this situation you could try to remove the category from the URL for the final product page.

- o **Duplication due to location, color or other attributes.**
  If your site has different URLs for white laptop, blue laptop, black laptop but the same content, or your site offers a service and you have a page for every district than you are very close to something named seo index spam.

- o **Every kind of autogenerated content**, spinned content, parsing of RSS feeds etc are not new practices. They are well documented and even if they work a little at first they don't last on the long run.
  Creating thousands of pages (ex: using free romanian DEX database on the internet, using postal code databese etc) with no real content is not going to improve your rankings.

How to check for duplicate content:

- the most simple way to check is just to take a snippet of text from your site and than copy paste it with exact match on google (use quotation marks at the begging of the search and at the end to achieve the exact match search).
- copyscape.com is a very good resource. It is free for a few tries, but if you need it for hundreds of pages be prepared to pay.
- webmaster tools -> Optimization -> Duplicate Title tags, or Duplicate meta description. You can gen some great insights from this.
- if everything else fails you could give a try at:
  - o http://www.duplichecker.com/
  - o http://plagiarisma.net/
  - o http://www.plagium.com/
  - o http://www.webconfs.com/similar-page-checker.php (just checks the duplication percent)

How to fix it

- rel=canonical tag can help a great deal with the technical part.
- create fresh new content, engaging for the user

2. Content is King

Content is probably the most important thing you need to focus on. A site with good SEO but poor spammy content is nothing than a spammy site. A site with good content but no SEO is just a step away from recognition.

These are just a few issues related to this matter:

- **Keyword stuffing** – repeating the same words or phrases so often that it sounds unnatural.
  We sell custom cigar humidors. Our custom cigar humidors are handmade. If you're thinking of buying a custom cigar humidor, please contact our custom cigar humidor specialists at custom.cigar.humidors@example.com.
  Overoptimization can be an issue now and could cause a penalty.
  http://www.mattcutts.com/blog/avoid-keyword-stuffing/
  Using the same alternate text for images comes close to keyword stuffing. The alternate text of any picture should describe that picture, not something else. I would even recommend not to add alternate text to pictures smaller that 100x100 px or to those that are part of your general template.
  TIPS AND TRICKS: use synonyms for keywords. Google understands the meanings of words and can determine relevancy.

- **Hidden text.**
  Hiddent text is hidden keyword stuffing.
  Text is often hidden using:

  o `<input type="hidden"`

  o css `display:none` attribute or `visibility:hidden;` (inline css or external css file)

  o misuse of tag closing. Adding text in the html file before ending of a tag.

  o text on a background with the same color

  o css `position:absolute; left:-1000px;`

  o hiding a DIV below the visible DIV.

  There are legitimate uses of display:none or tabbed browsing. If a text is hidden and you can access the text from button/link „view more" than it could be white hat.

- **Pagination.**
  One of the problem with pagination is splitting link popularity. Fortunately there is a quick fix for this. Implementing relnext and relprev tags on the actual pagination:
  http://googlewebmastercentral.blogspot.com/2011/09/pagination-with-relnext-and-relprev.html
  These tags tell Google to treat different pages of the pagination as components of the same article. Google will index every page of the pagination but will redirect the juice to the first page of the pagination.
  The rel=canonical tag can work here as well.

- **Positioning and relevancy.**
  Content well below the fold will not act as well as one right away visible.
  A higher bounce rate could be an indication of a low quality page, shallow content or poor presentation.

- **Footer content.**
  Many sites abuse the footer section of their sites with
    o keyword phrases stuffed over and over again
    o links from other sites footers. Group of links together with no content around.
   o list of links that use different anchor text from the main navigation

3. Black hat techniques

Black hat techniques are still used, but are less frequent now. They are intentionally used 99% of the times.

Keyword stuffing and hidden content that we discussed before are black hat techniques.

About Black Hat

There is a fine line between black and white in SEO. Some good idea can just be a bad idea for Google. Look at square A and square B. Judge the colors! Which is White Hat and which Black Hat? The color of block A is the same as color block B.
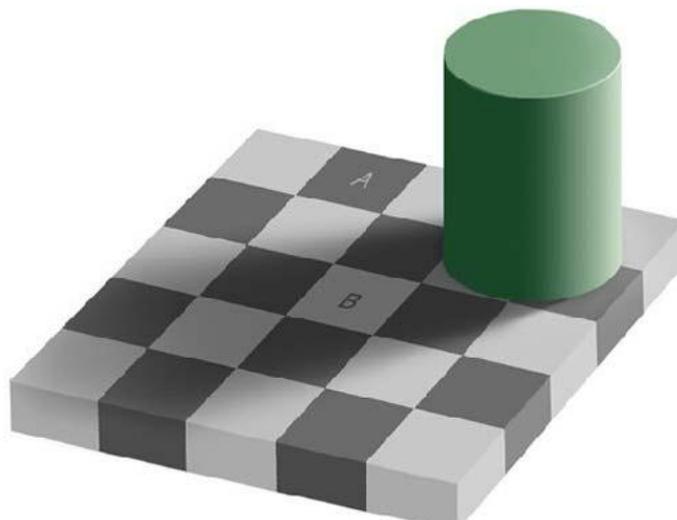
3.1) Cloaking

Cloaking involves using the IP address or features of the HTTP request (e.g. the User-Agent field) in order to identify and deliver unique content to a specific IP or User-agent.

Example:

- `<img src=nofilehere.gif onerror=window.open('http://mysite.com','_top')>`

- When surfers visit the page, they will be sent to "mysite.com" because the requested image file does not exist (therefore there will be an error). The spiders and search engines, on the other hand, will all see the original page
- Many high profile sites and companies use Cloaking to send different content to different IP addresses. IP delivery is the safer and preferred way to cloak
- There is a less chance of getting banned using IP delivery
- There is a 99% chance of getting banned using stupid JavaScript redirects.

Fake PR with cloaking

So you want to get into link exchanging, but no one wants to trade with you because your page rank sucks.

1. Buy a domain.
2. 301 Redirect the domain to a big site
3. Index the first page of the domain (make a link to it from a good site)
4. Put sneaky cloaking code
5. Wait for Google PR update

- ```php
  <?php
  if(ip is in list) (list of google IPs){
          header("HTTP/1.1 301 Moved Permanently");
          header("location:http://www.mysite.com");
          exit;}
  else
  { header('Content-Type: text/html; charset=ISO-8859-1');
  //Show your site
  }; ?>
  ```

-

3.2) Doorway pages

Doorway pages are pages that have been optimised to rank highly on the search engines for one or two particular keywords. Clicking on any of these can automatically redirect the user to another site (meta refresh or JavaScript or even .htaccess)

Trusted parasite hosts are critically important to making your doorway pages rank, and rank quickly.

As a rule of thumb, if you can't reach the page by following the site navigation or if a user can't get to it just by clicking, then it is a doorway page. You are not supposed to "visit" the page. Instead, you are just supposed to find it in the search results and then click through to get to the site in question.