**#10. Illegal HTML exists**: Some spam messages include a code for identification in the text of the message. The text is entered outside the HTML tags so as to hide the code from the recipient. There is no legitimate reason to add text outside HTML tags, so the mere presence of illegal HTML can be treated as suspicious.

**#9. Message body contains small font size**: In order to circumvent Bayesian filters and filters that block messages with only images, spammers enter ‘normal’ text at the bottom of the message in order to appear legitimate. Some spammers include this text in small font size.

**#8. Message subject contains email address or recipient name**: Either the complete email address or part of the email address (the part before the domain) is added to the subject in order to personalize the message and trick the recipient into thinking that it is a legitimate message. For legitimate mails there is no reason to enter the recipient’s email address in the subject, so the presence of this is a pretty sure sign of spam.  
 **#7. Message body is base64 encoded**: Spammers use base64 to encode the message headers and body so that spam filters are not able to read the content and perform any filtering. Most email clients will decode the message so that the message can still be read by the recipient.

**#6. Sender address contains number or character sequence**: Spammers use automated programs to register thousands of email addresses. Since they are generated in bulk, they often include number or character sequences such as FRfJIrqOpV@hotmail.com or bob36189624@gmail.com. At first spammers used number sequences but when most spam filters started to block these types of addresses they changed to using character sequences which are harder to detect.

**#5. From: and Reply To: address are different**: This is a common feature of spam mails, but it is also very common with newsletters. The importance of this characteristic should be minimized since it is also found in legitimate emails.

**#4. Message body contains remote image**: In order to avoid spam messages from being blocked by word filters, spammers include an image in their message that cannot be filtered for words. In addition, upon opening the email message the image is downloaded from the spammer’s website. Since each message contains a unique ID, the spammer will know exactly which recipient has viewed the mail. This indicates which email addresses are ‘live’ and can be sent even more spam.

**#3. Message contains only HTML body**: HTML messages usually include a plain text version of the email so that recipients with email clients that cannot read HTML can still view the message in plain text. However, many spammers tend to send HTML messages without this plain text body part. This is done to save on size and to force recipients to read the HTML version which automatically opens an image and connects to a web site when the message is opened. Newsletters also tend to send messages without a plain text body part, so it is important to use a white list of allowed newsletters so as not to catch any false positives.

**#2. Message contains many or only tags**: Some spammers try to circumvent content filters by placing lots of HTML comment tags within the email body text. In this way, content filters will not recognize the spam words since they are separated by comment tags. The recipient however, will not see the comment tags since these are not displayed when viewing the message in HTML. Therefore it is important to use an email filter that can filter emails by removing HTML tags first.

**#1. Recipient’s email address is not in the To: or Cc: fields**: Red Earth Software found this to be the most commonly found characteristic in current spam messages. The reason for this is that the recipient’s email address is hidden in the Bcc: field or X-receiver field, along with a substantial number of other email addresses. Spammers do this in order to conceal the fact that the mail was sent to a large number of recipients, and presumably so as not to publish their email list. Some persons might add recipients to the Bcc: field for sending out â€˜legitimateâ€™ mailings, but these will tend to be of a more personal nature (which you might wish to block anyway) since most professional companies do not use this method for sending newsletters or mailings. Note however that if you do block emails without a local recipient in the To: or Cc: field, you will be blocking all bcc: messages.

Bottom line: Many spam filters check for the existence of these characteristics (and more) and use these to determine whether the message should be identified as spam. Some characteristics are strong indicators that a message is spam, others really cannot be taken into account at all since they can also exist in legitimate emails. A system checking for spam characteristics can be very effective, but must make use of a sophisticated scoring system in able to flag spam correctly, applying a different weight for each characteristic.