IAB/ABC International Spiders & Bots List

The following "best practices" are to be used as a guide when implementing the IAB's Spiders and Bots list. These are not requirements, but recommended best practices to ensure that the appropriate level of testing is performed prior to implementing the IAB Spiders & Bots list. Additional information about the IAB Spiders & Bots list can be found in Appendix I.

Obtaining / Implementing the IAB Spiders & Bots List

There are actually two text files that make up the IAB Spiders & Bots list - one for qualified browsers (the Whitelist) and one for known robots (the Blacklist). Please note that the following should be applied based on your implementation of the IAB Spiders & Bots list. Refer to Appendix I for additional information about the International IAB/ABC Spiders & Bots List.

The IAB Spiders & Bots lists are updated on or before the 25th of each month, and are available via FTP (ftp.iab.net).

All updates to the list are summarized at the bottom of each list by month (the last line always represents the most recent update). Although all changes to the list go through a thorough vetting process by The Alliance for Audited Media (AAM), companies are encouraged to employ the following "best practices" prior to implementing the list. These procedures will ensure that updates to the list do not adversely affect your filtration and measurement processes.

If an entry on the list should be considered for removal, or if you've identified a new spider or valid user-agent string that should be added to the list, please contact AAM. Their email address is spiders.bots@auditedmedia.com. You may also submit a change request through the IAB website by clicking on the following link - https://www.iab.com/guidelines/iab-abc-international-spiders-bots-list.

Best Practices

The IAB Spiders & Bots list should be treated like any other configuration file that is used by your measurement system. As such, the use of the list as well as changes to the list should follow your standard operating procedures related to change management. In other words, there should be a formal process in place for downloading the file, testing the file and updating the "production" file.

1. It is recommended that all changes to the list be tracked in some fashion to ensure that the changes are properly documented. Changes are typically tracked through a ticketing system and therefore, you should open a new ticket whenever you make a change to the list.

Note: It is also recommended that the "production" copy of the list be updated on a monthly basis.

2. Once a ticket has been opened, the next step in the process requires that the new list be downloaded from the IAB website or FTP server. The main benefit of using the FTP server is that the process can be easily automated so that it occurs every month.

3. Once the file has been downloaded, the first test should be a simple "diff" check between the updated file and the previous file. This can be a manual process (i.e., text pad allows you to quickly compare files under the "Tools" menu option) or it can be automated.

4. The next test should include an analysis of the user-agents that will be filtered when using the new list. Using a sample of current production data (i.e., a list of user-agent strings extracted from production data is sufficient), identify those user-agent strings that are flagged by the updated IAB Spiders & Bots list.

    a. Review the output to ensure that the results are consistent with expectations. If the change results in activity being filtered that appears to be valid, then further investigation is strongly recommended.

       Note: It is helpful to include record counts by user-agent in this part of the analysis as materiality should be taken into consideration.

    b. Look for possible "exceptions" that may not be identified as such on the list of known robots. This refers to situations where an entry on the list may have one or more exceptions (i.e., situations in which the string should be considered valid). An example is provided in the Other Notes section of Appendix II.

       Note: The use of "exception" strings are required by MRC's Invalid Traffic Detection and Filtration Guidelines Addendum.

    c. Report any observed anomalies or unexpected results to AAM immediately for their review.

    Note: If you find one or more entries that require further review, it is acceptable to exclude these entries from the update pending further review; however, we do recommend that you update your production copy with the remainder of entries that do not require any additional follow up.

5. Document findings in your tracking system, have the change approved and follow the appropriate steps to push the configuration file into production. Finally, close the ticket.

Emergency Changes to the List

The IAB and AAM have implemented emergency procedures to update the Spiders & Bots list upon notification of a change that requires immediate attention. The affected lists will be emailed to all subscribers in addition to making them available via the FTP server. If you identify an entry on the list that should be removed immediately or a new bot that should be added immediately, contact AAM at spiders.bots@auditedmedia.com. Please note that emergency updates will only be considered if the change is considered critical and material to the business as well as time sensitive (i.e., cannot wait until the next month's update).

Appendix I - About the International IAB/ABC Spiders & Bots List

The IAB has enlisted AAM to perform evaluation and management services associated with maintaining and publishing the industry Spider and Robot list, the use of which is strongly encouraged for compliance to the IAB's Client Side Counting (CSC) Measurement Guidelines. We have also formed a Spider & Robot Policy Board to oversee and approve all list modifications (see list below).

In January 2006, the IAB and ABC UK (Audit Bureau of Circulations) agreed to merge their two lists to create a comprehensive ad- and site traffic-related spiders & bots list. This single list is now called the International IAB/ABC Spiders & Bots List.

There are two text files included in the list, one for qualified browsers and one for known robots. These lists are intended to be used together to comply with the IAB's "dual pass" approach to filtering as defined in the Ad Impression Measurement Guidelines[1] (i.e., identify valid transactions using the qualified browser list and then filter/remove invalid transactions using the known robots list).

Proper use of the list should follow this order:

Whitelist > Blacklist > Blacklist Exceptions

A user-agent must first match an entry on the Whitelist and not match an entry on the Blacklist to be considered not a bot. If a user-agent does not match any entry on the Whitelist, it must be considered a bot. If a user agent passes the Whitelist and then matches a Blacklist entry, it must be considered a bot unless it also matches at least one of the exception strings associated with the matched Blacklist entry.

The use of the Whitelist, Blacklist, and Blacklist Exceptions are required by the MRC's Invalid Traffic Detection and Filtration Guidelines.

At the beginning of each file is a comment section that describes the format of the file and how to use it.

The list is updated monthly to reflect changes that are brought to the attention of AAM, ABC and the IAB Policy Board.

For any questions or comments on how to use the list, email
spiders.bots@auditedmedia.com

---

[1] Filtration Guidelines - URL, user agent, and client browser information is used to exclude robots based on exact matches with a combination of two sources: (1) The IAB Industry Robot List and (2) a list of known Browser-Types published by the IAB. In the case of (1), matches are excluded from measurements. For item (2) matches are included in measurements. (Note that filtration occurring in third party activity audits is sufficient to meet this requirement.)

Appendix II - IAB/ABC International Spiders & Bots List Technical Details

Special characters in each text file:

 "#" - Comment (only at the start of a line)
"|" - Field separator
"," - Separator when multiple exceptions are present

Valid Browsers (a.k.a. "the Whitelist")

The numbers below correlate the column or field of each entry on the list that are separated by a pipe "|".

1) Case insensitive UA string to match
2) Active flag
    1 = pattern is active and should be matched
    0 = pattern is inactive, and should ignored
3) Start-of-string flag
    1 = pattern must occur at the start of the UA string
    0 = pattern may appear anywhere within the UA string
4) Inactive Date: mm/dd/yyyy format (only present if string has been inactivated)

Sample Entry: Mozilla/|1|1

Known Bots (a.k.a. "the Blacklist")

The numbers below correlate the column or field of each entry on the list that are separated by a pipe "|".

1) Case insensitive UA string to match
2) Active flag
    1 = pattern is active and should be matched
    0 = pattern is inactive, and should ignored
3) Comma-separated list of exception patterns
4) An additional flag was added to this list in November 2005 to identify those user agent strings on this list that would not pass the valid user-agent test and therefore, are redundant if both lists are used.
    1 = this entry is not needed for those who use a two-pass approach
    0 = this entry is always needed for both one-pass and two-pass approaches
5) Another flag was added to this list when the IAB and ABC merged their two lists (01/06) to identify those strings that primarily impact page impression measurement vs. those strings that primarily impact ad impression measurement (or both). The flags are as follows:

        0 = this entry primarily impacts page impression measurement

        1 = this entry primarily impacts ad impression measurement

        2 = this entry impacts both

6) Start-of-string flag

        1 = pattern must occur at the start of the UA string

        0 = pattern may appear anywhere within the UA string

7) Inactive Date: mm/dd/yyyy format (only present if string has been inactivated)

Sample Entry: googlebot|1||0|2|0

Other Notes

The 3rd column supports an "exception" feature, which is used to specify patterns that should be considered valid even though they are being flagged by the entry. For example, if the UA string includes "obot," the transaction should be marked invalid unless the string "robotics" is present, which allows for the counting of US Robotics cobranded browsers. There may be more than one exception for each pattern separated by a comma.

Sample Entry: obot|1|robotics|0|0|0

The 5th column attempts to associate the robot with page impressions or ad impressions (or both) but should be used only as a guide. Application of this list should be based on an analysis of the activity itself before excluding any entries.