

## APPENDIX I

# LIST OF INTERNET ROBOTS, CRAWLERS, AND SPIDERS

---

The growing use of internet robots, crawlers, and spiders has the potential to artificially inflate usage statistics. Only genuine, user-driven usage should be reported in COUNTER usage reports. Usage of full text articles that is initiated by automatic or semi-automatic bulk download tools such as Quosa or Pubget should only be recorded when the user has clicked on the downloaded full-text article in order to open it.

**Activity generated by internet robots, crawlers, and spiders must be excluded from all COUNTER usage reports.**

This list of internet robots, crawlers, and spiders was published in April 2016 and updated July 2016. Please note it is rationalised, removing some previously redundant entries (e.g. the text ‘bot’—msnbot, awbot, bbot, turnitinbot, etc.—is now collapsed down to a single entry ‘bot’).

The list is available at: <https://github.com/atmire/COUNTER-Robots> Each line in the file contains a regular expression (regex), which is a type of text string that describes a search pattern. When using the exclusion list, **all the regexes should be matched case-insensitively.**

For further information on regular expression matching, see: <http://www.regular-expressions.info/quickstart.html>.

*Please let us know of any user agents that should be included in this list or to suggest other amendments.*