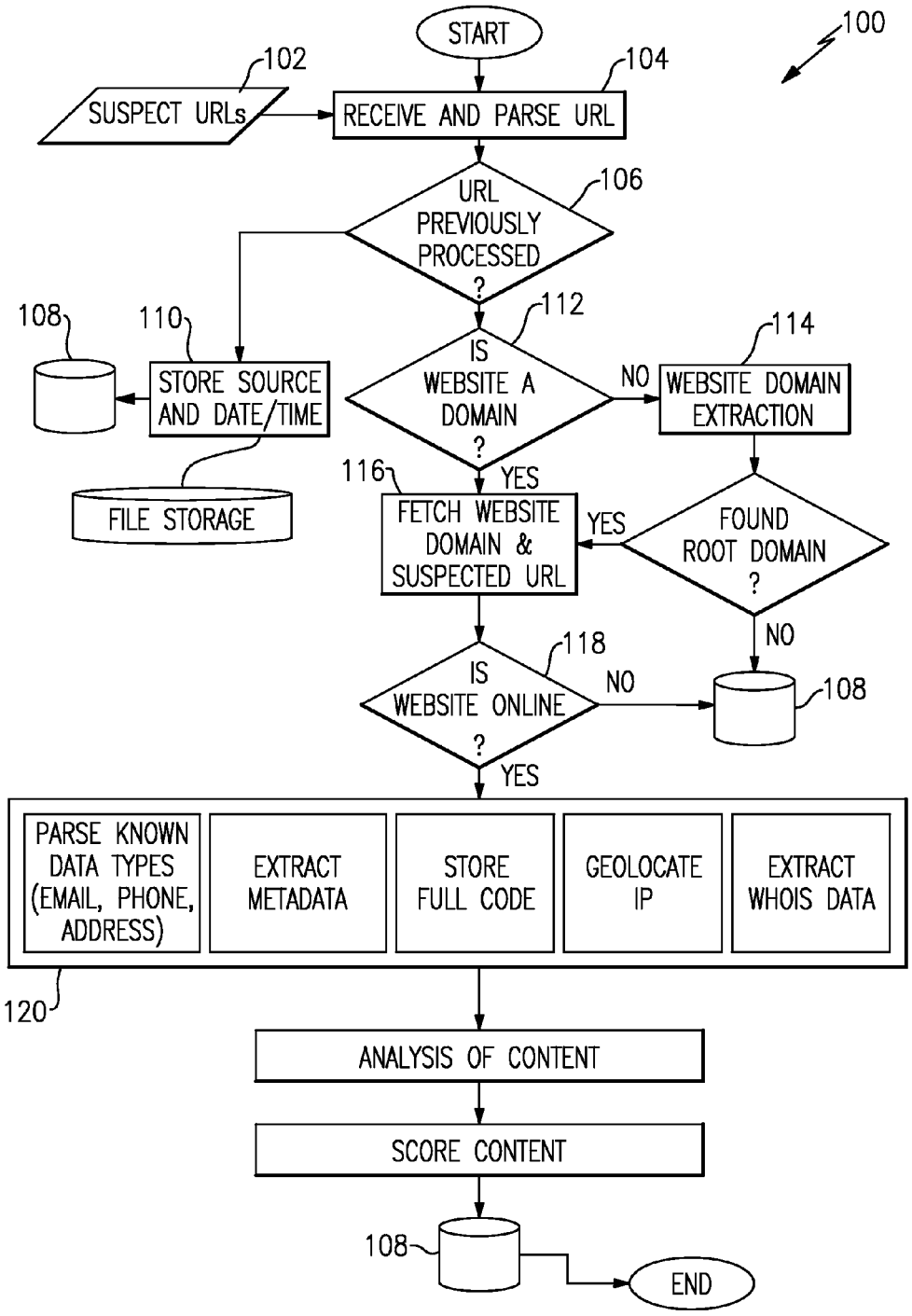
100
START
102
SUSPECT URLs
104
RECEIVE AND PARSE URL
106
URL PREVIOUSLY PROCESSED ?
108
110
STORE SOURCE AND DATE/TIME
FILE STORAGE
112
IS WEBSITE A DOMAIN ?
NO
114
WEBSITE DOMAIN EXTRACTION
YES
116
FETCH WEBSITE DOMAIN & SUSPECTED URL
YES
FOUND ROOT DOMAIN ?
NO
118
IS WEBSITE ONLINE ?
NO
108
YES
PARSE KNOWN DATA TYPES (EMAIL, PHONE, ADDRESS)
EXTRACT METADATA
STORE FULL CODE
GEOLOCATE IP
EXTRACT WHOIS DATA
120
ANALYSIS OF CONTENT
SCORE CONTENT
108
END

FIG.2