

**Table 1**Frequency of reporting of keywords/phrases in the dictionary used in automated content analysis of SOX 404 reports with ITWs<sup>a</sup>.

	Keyword(s)/phrases <sup>a</sup>
<b>IT context indicator keywords and phrases</b> which are not associated with a specific ITW category	<b>information technology</b> (161); <b>user*</b> (51); <b>software</b> (48); <b>accounting system*</b> (43); <b>information system*</b> (43); <b>financial application*</b> (42); <b>automated</b> (34); <b>computer*</b> (28); <b>financial application programs and data</b> (24); <b>database*</b> (21); <b>comput* control*</b> (20); <b>reporting system*</b> (17); <b>payroll process*</b> (16); <b>ERP</b> (15); <b>IT general controls</b> (13); <b>system* control*</b> (9); <b>billing system*</b> (8); <b>network</b> (8); <b>inventory process*</b> (7); <b>financial reporting system</b> (6); <b>general ledger system*</b> (6) <b>perpetual inventory records</b> (6); <b>enterprise resource planning</b> (5); <b>inventory system*</b> (4); <b>Oracle</b> (4); <b>platform*</b> (4); <b>processing of financial data</b> (4); <b>system-generated report*</b> (4); <b>computing</b> (3); <b>data entry</b> (3); <b>ITC*</b> (3); <b>ITGC</b> (3); <b>online</b> (3); <b>accounts payable system*</b> (2); <b>application source code</b> (2); <b>application system</b> (2); <b>billing process*</b> (2); <b>computer-generated</b> (2); <b>inventory costing system*</b> (2); <b>IT controls</b> (2); <b>IT department</b> (2); <b>IT function</b> (2); <b>payroll system*</b> (2); <b>point-of-sale</b> (2); <b>processing file*</b> (2); <b>SAP</b> (2); <b>transactional control*</b> (2); <b>accurately enter*</b> (1); <b>application based</b> (1); <b>CIO</b> (1); <b>current system</b> (1); <b>data processing</b> (1); <b>enterprise business system</b> (1); <b>enterprise resource platform</b> (1); <b>financial accounting IT applications</b> (1); <b>financial software system</b> (1); <b>financial system application*</b> (1); <b>generated by the system</b> (1); <b>hardware</b> (1); <b>input* to model*</b> (1) <b>IT staff*</b> (1); <b>IT support staff</b> (1); <b>Microsoft_AX</b> (1); <b>model input*</b> (1); <b>MRP</b> (1); <b>operating system</b> (1); <b>programmer*</b> (1)
<b>IT weakness (ITWs) categories</b>	
<i>Keywords/phrase in bold font type are also IT context indicator keywords/phrases<sup>a</sup></i>	
Access	access (168); <b>user access</b> (40); <b>access control*</b> (33); <b>restrict* access</b> (30); <b>password*</b> (18); <b>system access</b> (16); <b>inappropriate access</b> (14); <b>access rights</b> (9); <b>logical access</b> (9); <b>security access</b> (8); <b>access privilege*</b> (4); <b>physical access</b> (3); <b>security setting</b> (3); <b>system right*</b> (2); <b>network access</b> (1); <b>prevent management override</b> (1); <b>system privle*</b> (1); <b>user identification</b> (1)
Monitoring	oversight (384); review* (258); monitor* (185); supervis* (126); examin* (73); logging (4); scrutiny (1)
Design issues	design* (387); reporting requirement* (50); complexit* (23); lack of effective (19); assumption*used (11); formal process* (9); manual process* (8); inadequate system* (7); <b>legacy</b> (6); manual intervention* (6); disparate (5); manually (5); decentralized (4); interface* (4); not integrated (4); incompatible application* (3); lack of a unified (3); non-integrated (3); audit trail* (2); manual performance (2); do not appropriately address the requirements (1); functional business requirement* (1); inadequate recording and report* (1); invest sufficiently in technology (1); large number of manual process* (1); manual* intensive* (1); over-reliance on (1); properly integrated (1); reporting capabilities (1); reporting limitations (1); system capabilities (1); user dependence (1) test* (385); implementation (46); change management (37); development (35); <b>program change*</b> (23); <b>program development</b> (16); configure* (15); change control* (11); conversion (10); migrat* (9); system change* (7); changes to financial (5); system* development (5); version control (5); approval of chang* (4); accuracy of calculat* (3); system* conversion* (3); authorized change* (2); changes to production application* (2); placed into production (2); <b>program* error*</b> (2); project management (2); set-up (2); software chang* (2); application error* (1); authorization of chang* (1); changes to program* (1); deactivat* (1); develop and validate (1); maintenance control* (1); recording changes (1); <b>testing of program chang*</b> (1); track* chang* (1); updated and maintained (1)
Change and development	<b>spreadsheet*</b> (119); <b>end-user comput*</b> (11); <b>end user comput*</b> (4); <b>user developed application</b> (2); <b>cell protection</b> (1) segregat* (182); incompatible duties (17); incompatible responsibilit* (2);
End-user computing	
Segregation of incompatible functions	

(continued on next page)

Table 1 (continued)

	Keyword(s)/phrases <sup>a</sup>
Policies	policies (387); policy (42); lack of documented procedures (1)
Documentation	document* (174)
Masterfiles	<b>master file*</b> (23); payroll changes (7); <b>census data</b> (4); <b>data file*</b> (4); <b>master data</b> (4); <b>payroll data</b> (4); <b>vendor data</b> (3); vendor management (3); <b>master record*</b> (2); <b>masterfile*</b> (2); <b>master-file*</b> (2); vendor setup (2); vendor set-up (2); <b>price table*</b> (1); <b>standing data</b> (1); <b>vendor accounts</b> (1); <b>vendor file*</b> (1); <b>vendor listing</b> (1)
Staffing sufficiency and competency	train* (113); experienc* (80); knowledg* (74); *sufficient complement of personnel (40); *sufficient number of (24); *sufficient personnel (24); skill* (21); inadequate staff* (10); competenc* (5); adequate staffing (3); *sufficient complement of staff (3); understaffed (3); inadequate IT staff (2); inadequate personnel (2); inadequate IT support staff (1); limited number of personnel (1); personnel limitation* (1); shortage of resources (1); turnover of personnel (1);
Backup	<b>backup*</b> (13); <b>disaster</b> (9); <b>back-up*</b> (8); <b>offsite</b> (6); <b>back up*</b> (5); <b>off-site</b> (3); <b>record* storage</b> (1); <b>remote location</b> (1); <b>removable media</b> (1); <b>rotation media</b> (1); <b>uninterruptible power</b> (1)
Operations	<b>computer operation*</b> (11); <b>information system* operation*</b> (3); <b>software licens*</b> (3); <b>IT operation*</b> (2); <b>operating procedure*</b> (2); <b>operations report*</b> (1)
Security (other than access)	<b>physical security</b> (7); <b>information security</b> (5); <b>it security</b> (4); <b>encrypt*</b> (3); <b>firewall</b> (3); <b>security management</b> (3); <b>security setting</b> (3); <b>control* over security</b> (2); security and data protection (2); <b>security issue*</b> (2); <b>system security</b> (2) <b>antivirus</b> (1); <b>anti-virus</b> (1); electronic transmission (1); fire (1); <b>intrusion detection</b> (1); <b>network vulnerability assessment*</b> (1); <b>security breach</b> (1); security configuration (1); <b>security incident*</b> (1)
Outsourcing	<b>service provider*</b> (14); <b>outsourc*</b> (7); <b>SAS 70</b> (3); <b>third-party service</b> (3); <b>data center*</b> (1); <b>out-sourc*</b> (2);

<sup>a</sup> Keywords/phrases in the dictionary used to search for IT weakness indicators near (in the same sentence and within 20 words either side) of IT indicators. IT indicators keywords/phrases are shown in bold font. Less restrictive searches reported in Table 3, were conducted for IT weakness indicators anywhere in sentences that contain IT indicators. In more restrictive searches reported in Table 3, ITWs are ignored if also within 20 words (as determined by the content analysis software) of “exclusion keyword/phrases” in sentences with IT indicators. Exclusion keyword(s)/phras(es), frequently associated with SOX 404 “boilerplate” definitions and non-ITW contexts are: adherence to policies; capitalization polic\*; capitalized software; communicat\* policies; communication to employees; compliance with; corporate governance policies; costs of computer software; do not execute; did not operate effectively; documentation regarding; documentation supporting; enforce; enforcement of; ensure the proper operation; established company policies; execute its policies; failed to apply; four aspects of information technology general controls; four basic information technology; four basic IT; in accordance with; information technology area\*; information technology controls are policies; internally developed software; IT areas; ITCS are policies; ITCS include four basic; multiple element software arrangement\*; policies did not operate effectively; revenue recognition polic\*; software accounting policies; software capitalization; software development costs; source document\*; sufficiently document; support operations; supporting documentation; the company’s polic\*; the policy of; and were not followed.

with tools (word frequency count, key-word-in-context, similarity measure, text extraction, and drag and drop categorization tools) to help the researcher group keywords/phrases into categories. However, coding may be unreliable and insufficiently grounded in theory if categories are formed using automated text mining without someone with subject matter expertise reviewing and refining the linkages. While today’s automated content analysis toolkits have sophisticated data visualization, mapping, and networking tools, expert categorization and search rule development will continue to distinguish automated content analysis from other text analytics methodologies. Fig. 1 summarizes the software features necessary and the expertise leveraged along the content analysis/text analytics spectrum.

Using automated content analysis to study ITWs in SOX 404 reports presents a unique opportunity for comparing the efficacy of manual vs. automated content analysis of unstructured text. There are several hundred SOX 404 reports in which auditors describe ITWs in unstructured text: a number sufficient for development and validation of the categorized dictionary of keywords/phrases required for automated content analysis. Further, there exists a manual/computer assisted content analysis of ITWs (Klamm and Weidenmier Watson, 2009) to which the results of the automated search may be compared. Since recent