



## AI to automatically detect sensitive documents

Using advanced machine learning techniques, Text IQ's Deep Concept Analysis (DCA) aids in the identification and review of sensitive documents.

### How it works

The engine identifies documents within a data set that contain words and phrases that are of *similar meaning* to known words and phrases. Customers utilize DCA to search for documents related to 1) the use of embarrassing language, 2) detection of code words, and/or 3) documents triggering privacy concerns (i.e. PII/PHI). After review, customers can export all relevant document IDs to flag and/or withhold from production.

All examples are pulled from Text IQ's Deep Concept Analysis running against the publicly available Enron data set

### Offensive Language Detector

Searching "idiot" returns the word "chicks" and identifies a highly offensive company email

idiot

idiot Emails

i'm really Emails

tubby Emails

muttered Emails

got Emails

season six Emails

**chicks Emails**

dammit Emails

fw fw Emails

blonde Emails

anyway Emails

female news anchor who Emails

huh Emails

nonchalantly Emails

his dad Emails

joke Emails

just kidding Emails

exclaimed Emails

girlfriend Emails

redneck Emails

I was a stud in NYC but I pretty much get about on average phone numbers a night and at least hot chicks that say that they want to go home with me every night I go out I love the buyside I have bankers calling me everyday with opportunities and they pretty much cater to my every whim you know golfing events lavish dinners a night out clubbing The guys I work with are also all chill I live in the same apt building as my VP and he drives me around in his Porsche of in all of Korea to work and when we go out What can I say live is good CHUNG IS KING of his domain here in Seoul

So all of you fuckers better keep in touch and start making plans to come out and visit my ass ASAP I'll show you guys an unbelievable time My contact info is below Oh by the way someone's gotta start fedexing me boxes of domes I brought out about but I think I'll run out of them by Saturday

Later  
CHUNG

Peter Chung  
The Carlyle Group

### Code Word Detector

off balance sheet

off balance sheet Emails

balance sheet Emails

partnerships Emails

off balance sheet partnerships Emails

special purpose entities Emails

accounting Emails

**raptors Emails**

liabilities Emails

company's balance sheet Emails

debt Emails

ljm co investment Emails

its balance sheet Emails

off balance sheet transactions Emails

raptor

raptor Emails

condor Emails

raptor ii Emails

jedi Emails

ljm Emails

raptor's Emails

monetized Emails

write down Emails

raptor iii Emails

financing vehicle Emails

raptor partnerships Emails

raptor entities Emails

Searching for "off balance sheet" returns the code word "Raptor"; searching "Raptor" returns multiple other code words