Nr. 1 / Januar 2007
http://www.sicherheitspolitik.ch

SICHERHEITSPOLITIK

der Verwaltung, Wirtschaft und Bevölkerung

Sicherheit benötigt frühzeitige Beurteilungen.

BST Sicherheitstechnik AG
Lagerhausweg 10, 3018 Bern
Tel. 031 997 10 10, Fax 031 997 55 50
http://www.bst-sicherheitstechnik.com

Ronald Schulze

Zimmermann Security GmbH
Beratung bei Inventurdifferenzen
Schloss-Str. 37, 8803 Rüschlikon
http://www.zimmermann-security.ch

Informations- und Krisenmanagement

Presdok AG
Mimosenstr. 5, CH-8057 Zürich
044 312 10 50
http://www.presdok.ch

© http://www.fallucchi.ch

Spezialisten finden Wissen!

Ronald Schulze

Ausgehend von einem kurzen technologischen Überblick wird eine Lösung beschrieben, wie der täglich zunehmenden Datenflut entsprochen werden kann. Hauptaugenmerk wird auf die Möglichkeiten gelegt, innerhalb dieser Unmengen von Daten[1] das darin repräsentierte Wissen zu extrahieren und somit einer Bewertung bzw. einer Relevanz für den aktuellen Fall zuzuführen. In möglichst kurzer Zeit soll erkannt werden: ist in dem vorliegenden Datenmaterial das für mich (z.B. beweisrelevante) interessante Wissen enthalten? Wenn ja, in welcher Form (Struktur) repräsentiert es sich, wo findet man es?

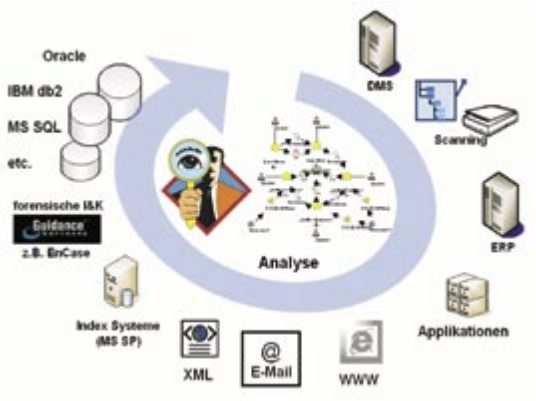
Ronald Schulze
Jahrgang 1961, Betriebswirt (FH) / Wirtschaftsinformatiker, postgradualer Studiengang für Management, Marketing und Recht der Europäischen Union an der Akademie für Internationale Wirtschaft Berlin, Project & Account Manager der Ontos International AG in Nidau (Schweiz), Kontakt: ronald.schulze@ontos.com

Die Problematik von Massendaten

Zu verarbeitende bzw. zu betrachtende Daten haben zwei Haupteigenheiten: eine unterschiedliche Genese (woher kommen die Daten) und einen unterschiedlichen Aufbau - unstrukturiert oder strukturiert (semi-strukturiert).

Über 80% der Daten unterschiedlicher Genese sind unstrukturierter Art (Dokumente, Memos, E-Mails, Internetseiten usw.).

Heute werden jedoch etwa 80% der Tätigkeit während einer Ermittlung/ Analyse für die Datenaufbereitung verbraucht; nur 20% der Arbeitszeit kann für die eigentliche Analyse genutzt werden.



Was passiert mit den Daten?

Die zu betrachtenden Daten werden mittels einer NLP-Software[2] verarbeitet und die darin enthaltenen Objekte (Entitäten)[3] jeweils mit deren Eigenschaften (Attribute; z.B. Name, Vorname, Geschlecht usw.) identifiziert.

Da jedoch in einem umfangreichen Dokument bzw. einer Dokumentensammlung die einzelne Person nicht immer namentlich beschrieben wird, kommt hierbei der Erkennung der sogenannten Anapher[4] eine grosse Bedeutung zu.

Wesentlich interessanter und bedeuteungsvoller ist die Erkennung der in den Daten beschriebenen Beziehungen zwischen den Objekten, in welcher Relation stehen diese zueinander? Auch dieses geschieht mit Hilfe der genannten NLP-Software. Entsprechend des integrierten Wissensmodelles (Ontologie) kann schlussendlich aus einem unstrukutierten Dokument eine Struktur geschaffen wurden, die das darin enthaltene und beschriebene Wissen repräsentiert.

Zur Vergegenwärtigung: Dieser Prozess ist nur sinnvoll anwendbar, wenn die Aufgabe besteht, innerhalb von zur Verfügung stehenden (unstrukturierten) Massendaten[5] kurzfristig zu erkennen: Ist darin relevantes Wissen enthalten? Wenn ja - wo?

Eine einzelne Akte bzw. einige DIN A4-Seiten eines Berichtes sind sicherlich auf der herkömmlichen Art und Weise analysierbar. Aber in welchem Falle ist man in der glücklichen Lage, mit nur so wenigen Ausgangsdaten, in den meisten Fällen, regelrecht kämpfen zu müssen?!

Ein möglicher Arbeitsablauf

Stellt man sich vor, ein Rechtsanwalt erhält im Zuge seiner Tätigkeit innerhalb eines Wirtschafts-Strafverfahrens Akteneinsicht, so u.a. mehrere Leitz-/Bundesordner in Kopie. Dazu kommen in den meisten Fällen umfangreiche Dokumentsammlungen externer Wirtschaftsprüfungsgesellschaften bzw. unternehmensinterne Dokumente. Nun soll er kurzfristig bewerten, sind darin Informationen über seinen Mandanten (eine Person oder Firma) enthalten, die beweisrelevant oder gar entlastend sind.

Ausgehend vom bekannten Namen (der Person oder der Firma) müsste man nun alle Dokumente unter folgenden (oder noch mehreren) Massgaben lesen: Kommt die Person/Firma darin vor, wenn ja, in welcher Beziehung zu anderen Personen/Firmen/Konten usw. wird diese genannt bzw. wie wird dieses beschrieben. Die Angabe einer Quellreferenz (Fundort, welches Dokument; Seite) sollte nicht vergessen werden. Alle diese Erkenntnisse werden dann durch den Bearbeiter im jeweiligen Dokument markiert und für eine weitere Auswertung manuell in ein dafür geeignetes System[6] übertragen.

Ist dieses realistisch bzw. in der heutigen Zeit überhaupt noch machbar? Viele solcher Notwendigkeiten scheitern

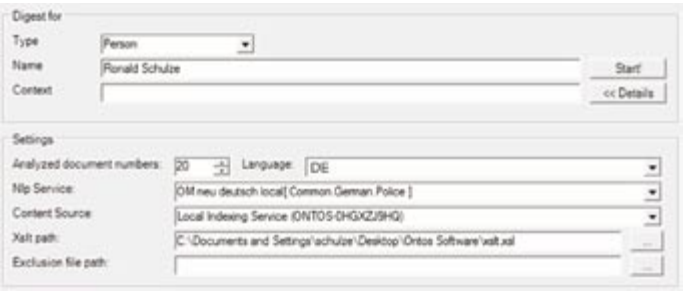
doch schon allein am Zeitmangel (drohender Fristverzug) bzw. am fehlendem Personal. Weiterhin: Einem Rechtsanwalt zur Verfügung gestellte Akten aus einem laufendem Verfahren kann dieser auch nicht «bis in alle Ewigkeit» behalten, Eile ist also geboten. Ausserem: Wem soll man es zumuten, sich durch Berge o.a. Dokumente zu kämpfen, dabei stets im Hinterkopf zu haben, wonach zu suchen ist?!

Der Lösungsansatz

Die zur Verfügung gestellten oben genannten Daten müssen in elektronischer Form vorliegen, das heisst eingescannt und mittels einer OCR-Software nachbearbeitet sein. Dann muss angegeben werden, wo sich diese elektronischen Dokumente befinden (Verzeichnissystem, SharePoint-Server, lokales Indexiersystem des Computers[7]).

Nunmehr spezifiziert man den Objekttyp, wonach gesucht werden soll (z.B. Person) und vergibt das eindeutige Suchkriterium (z.B. Familienname der Person). Die Angabe, wieviele der vorliegenden Dokumente prozessiert werden, ist möglich und hat demnach auf die Verarbeitungsdauer Einfluss.

Eine Besonderheit noch: Es kann spezifiziert werden, dass bei einer bestimmten Objektkonstellation (z.B. Personen-Personen-Beziehung)[8] das System automatisch einen Warnhinweis ausgibt oder eine vordefinierter Arbeitsablauf (workflow) automatisch gestartet werden soll. Somit kann hier durchaus von einen proaktiven System gesprochen werden.



Nach Prozessierung aller Dokumente erhält der Benutzer ein von ihm vordefinierten Bericht bzw. eine Zusammenfassung, einschliesslich Angabe der Fundstelle (des Quelldokumentes) sowie der betreffenden Textpassage, woraus das System die Objekt-Objekt-Beziehung erkannte.

Die Ergebnisse werden in einer separaten Datenbank (Wissensdatenbank; KnowledgeBase) gespeichert, wobei diese jedes derzeit verfügbare relationale Datenbankmanagementsystem darstellen kann.

Über alle in der Datenbank abgespeicherten Elemente ist eine echte semantische Suche möglich. Hierbei wird u.a. spezifiziert, dass z.B. der Text «Otto» der Vorname einer Person sein soll (im Gegensatz dazu «Otto» als Bestandteil des Firmennamens eines Versandhandels).

Sämtliche Ansätze, dieses mit einer Volltextrecherche (auch unter Ausnutzung von Fuzzy-Logic oder Boolean'schen Operatoren) zu realieren, sind nicht möglich.

Die Visualisierung aller Ergebnisse in einem Beziehungsgeflecht (Chart oder kognitive Karte/CMAP), eine visuelle Navigation innerhalb des Beziehungsgeflechtes und die Übergabe der gefundenen Objekte und Relationen (aus der KnowledgBase) an eine relationale Datenbank sind weitere Funktionalitäten.

Neben den oben aufgeführten Datenquellen ist auch die Recherche im Internet (z.B. via Google) möglich; die sonst erzielten Treffermengen aus einer Google-Suche stellen demnach die Datenquelle dar, die mittels des o.a. Verfahrens adäquat be- und verarbeitet werden.

Übergang zur strukturierten Welt - Beseitigung des «Medienbruches»

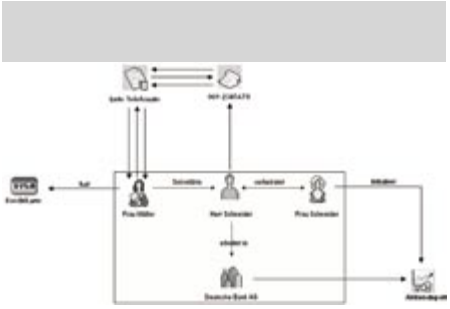
Im Zuge von Ermittlungen bzw. Sonderuntersuchungen hat man es aber nicht nur mit unstrukturierten Daten zu tun, deren sonst arbeitsintensive Aufbereitung zur dann folgenden Verifizierung nun als abgeschlossen betrachtet werden kann.

Wäre es nicht ideal, zu dem zuerst gefundenem Wissen, den erkannten Objekten und Beziehungen, weitere Informationen (Wissen) aus bereits bestehenden stukturierten Datenhaltungen zu erhalten?

Selbstverständlich ist dieses möglich, indem von jedem gefundenen Objekt, zu dem ich weitere Informationen aus der «strukturierten Welt» erhalten möchte, eine Datenverbindung (connect) zu genau der Datenquelle herstellen, die von Interesse ist. So läge es ja nahe, bei erkannten Personen-Firmen-Beziehungen auch zu hinterfragen, wie z.B. die Gesellschafterverhältnisse in der erkannten Firma sind oder in welchen weiteren Firmen die erkannte Person evtl. eine Gesellschafterrolle inne hat. Hierfür bieten verschiedene, unter dem Begriff «public sources» subsummierbare, mehrfach kostenpflichtige Anbieter,[9] entsprechende Möglichkeiten an.

In den Skizzen unten wird, ausgehend vom ersten, aus den unstrukturierten Daten gewonnenen Wissen dargestellt (erste Skizze), dass zwischen zwei Personen Telefonate stattgefunden haben bzw. eine Person auch ein Konto bei der Bank hat. Das Wissensnetz erweitert sich sukzessive (zweite Skizze).

Erstes gewonnenes Wissen aus unstrukturierten Daten



Weitere Informationen aus strukturierten Datenhaltungen oder Datenquellen

Anwendungsbeispiele

Nutzung für Vorfeldermittlungen

Vorfeldermittlungen[10] sind Ermittlungen zur Klärung der Frage, ob die Einleitung eines Ermittlungsverfahrens zulässig ist, also ob ein Anfangsverdacht besteht.

Massnahmen aufgrund von Vorfeldermittlungen können sich aus den Polizeigesetzen ergeben oder sie bedürfen keiner Rechtsgrundlage, da sie gar nicht in Rechte eingreifen (Recherche aus allgemein zugänglichen Quellen, z. B. Internet).

Jedoch gibt es für das Mittel der Vorfeldermittlungen auch Auffassungen pro und contra. Solche Untersuchungshandlungen und Ermittlungen unterhalb der Schwelle eines Anfangsver-

dachts unterliegen nicht der Regelung der Strafprozessordnung (StPO). Deren Schutzmechanismen greifen erst ein, wenn zureichende tatsächliche Anhaltspunkte für das Vorliegen einer Straftat bestehen. So wird jedoch öfters auf den weitaus problematischeren Teil der Datenerhebungen und den daraus resultierenden Erkenntnisgewinn, der ausserhalb der prozessual geregelten Vorgänge erfolgt, hingewiesen. Hier wird dann auch auch sehr schnell von einer regelrechten «Rasterfahndung» gesprochen, was dieses aber nicht darstellt.

Es gibt jedoch einige typische und klassische Deliktsbereiche (z.B.: Kapitalanlage- oder Vermittlungsbetrug, I&K-Straftaten),[11] wo gerade eine Vorfeldermittlung das Mittel der Wahl sein sollte. So kann die regelmässige Auswertung der Anzeigen/Artikel in der Tagespresse, entsprechender Fachpublikationen (z.B. Gerlach-Report) oder bestimmter Internetseiten erste Hinweise auf Straftaten geben. Anzeigen mit unrealistischen Renditeversprechen, Anzeigen unter Chiffre oder mit Firmenanschriften im Ausland sollten dann genauer überprüft werden.

Die Auswertung so gewonnener Informationen führt wie in anderen «klassischen» Deliktsbereichen auch, nahezu zwangsläufig zur Namhaftmachung von Personen oder Organisationen, die wiederholt bzw. besonders intensiv in diesem Deliktsbereich in Erscheinung treten. Dabei beziehen sich doch diese Massnahmen stets auf die sog. Klärung eines Anfangsverdachts, ist daher letztlich anlassgebunden, ist somit vereinbar und nachvollziehbar mit der prinzipiellen Aufklärungsaufgabe und erspart unter Umständen evtl. betroffenen Personen den Beschuldigtenstatus.

Ein möglicher Ablauf: Nach Bekanntwerden einer der o.a. Deliktsform bzw. Festellung, dass diese in einen bestimmten Territorium gehäuft auftreten, sollten diese «public sources»[12] genauer betrachtet werden, mit deren Hilfe eventuelle Straftäter ihre Opfer «rekrutieren». Nach Auswertung dieser Daten (meistens in unstrukturierter Form vorliegend) werden diese dann mit den bereits erfassten Vorgängen (gespeichert in relationalen Datenhaltungen) verbzw. abgeglichen.

Wie dann die weitere Verfahrensweise gegenüber bekanntgewordenen involvierten Personen oder Firmen ist, hängt vom internen Arbeitsablauf der ermittelnden bzw. recherchierenden Stelle ab.[13]

Allein schon deshalb, dass sich zu einem frühen Zeitpunkt um das Angebot «gekümmert» wird, kann dazu führen, dass ein unseriöses Angebot vom Markt verschwindet. Auch das Aufwand-Nutzen-Verhältnis von einer Vorfeldermittlung zur späteren Ermittlungstätigkeit (Einleitung eines Ermittlungsverfahrens mit allen Konsequenzen) muss eindeutig zu Gunsten des Ersterem betrachtet werden.

Nutzung für Bekämpfung der Geldwäsche[14]

Rechtliche Grundlagen und Erfordernisse: Die Europäische Kommission hat am 01.08.2006 die Richtlinie 2006/70/EG mit Durchführungsbestimmungen für die 3. EU-Geldwäsche-Richtlinie erlassen, welche am 24.08.2006 in Kraft getreten sind.

Diese Durchführungsbestimmungen regeln insbesondere, wer als so genannte Politisch exponierte Person (PEP) anzusehen ist. Nach diesen Vorgaben gelten nur für ausländische PEPs entsprechende Sorgfaltspflichten und somit für die Kreditwirtschaft erhöhte Anforderungen im Hinblick auf diesen Personenkreis.

Bisherige Verfahrensweise: Herkömmliche und bisher bekannte Lösungen verschiedener Hersteller basieren darauf, dass ein Finanzinstitut seine potenziellen und existierenden Kunden anhand von öffentlich zugänglichen Listen, so genannten Watch-Listen (Black-PEP-Listen oder Worldcheck™-Listen), prüft.

Diese enthalten Informationen zu Personen, die wegen Terrorismus oder anderer Verbrechen gesucht oder verdächtigt werden bzw. die in politisch und wirtschaftlich exponierten Funktionen tätig sind.

Es ist jedoch bekannt, dass nicht die Personen, die eventuell in einer der genannten Listen erfasst sind, mit einem Finanzinstitut eine direkte Geschäftsbeziehung anbahnen werden. Vielmehr schaffen sie sich eine regelrechte «Korona» von Strohmännern und Scheinfirmen, die in den Watch-Listen nicht aufgeführt sind.

Eine Prüfung durch das Finanzinstitut wird in diesen Fällen keine negativen Befunde erbringen.

Der Lösungsansatz: Die Ontos-Lösung geht weiter: sie prüft eine Person zusätzlich gegen öffentlich verfügbare Informationsquellen ab und erkennt, ob eine Beziehung zu einer anderen Person (oder Institution), die in den Watch-Listen als suspekt geführt wird, besteht.

Im Falle einer solchen Beziehung analysiert die Ontos-Lösung diesen Sachverhalt und generiert automatisch (individuell an die Nutzerspezifika anpassbare) Reports mit entsprechenden Warnhinweisen (sogenannten alerts) bzw. es kann daraus ein vordefinierter Arbeitsablauf (workflow) ausgelöst werden.

Fallbeispiel: Es soll die Person «Ronald Schulze» geprüft werden.

1. Eingabe «Ronald Schulze» als Suchkriterium, danach Auswahl der Datenquellen, worin nach Informationen zu dieser Person suchen gesucht werden sollen.
2. Der Crawler durchsucht die ausgewählten Datenquellen und übergibt die Ergebnisse (unstrukturierte Daten) an den OntosMiner. Dieser verarbeitet diese unstukturierten Daten und erkennt darin enthaltene Objekte (z.B. Personen, Firmen, Örtlichkeiten usw.), deren Eigenschaften und Ausprägungen sowie eventuell beschriebene Beziehungen (z.B. Personen-Personen, Personen-Firmen, Firmen-Firmen, Personen-Ort usw.).

Im obigen Beispiel wurde erkannt, dass über Beziehungen des «Ronald Schulze» zu drei Personen («Achim Müller», «Klaus Lehmann» und «Willi Land») innerhalb der unstrukturierten Daten berichtet wurde.

3. Nun wird in den vorher ausgesuchten Datenquellen (z.B. PEP, WorldCheck(TM) oder interne «black lists») nach
a) der Person «Ronald Schulze» und
b) den dazu gefundenen und in einer Beziehung zu «Ronald Schulze» stehenden o.a. 3 Personen gesucht.

Im Beispiel steht ein Eintrag (Treffer) über die Person «Klaus Lehmann».

Wichtig: Im dargestellten Fall bestand zur ursächlich abzuprüfenden Person «Ronald Schulze» kein Eintrag, aber zu einer der o.a. drei gefundenen, mit dem «Ronald Schulze» in Beziehung stehenden Personen, ein «Negativeintrag». Bisherige Lösungen gäben in diesem Falle keine Warnmeldung aus, weil Beziehungen zu Dritten, ausgehend von der ursächlich zu prüfenden Person, nicht betrachtet werden.

4. Es wird ein an die individuelle Situation angepasster Report mit entsprechendem Warnhinweis auf Beziehungen zu suspekten Personen (im Bespiel «Klaus Lehmann») ausgegeben. Von diesem aus kann dann eine weitere Recherche durchgeführt werden.

Schlussfolgerungen

Die beschriebene Softwarelösung ist geeignet für:

- die semantische Analyse von unstrukturierten (sowie strukturierten) Daten mit dem Ergebnis der Darstellung von darin enthaltenen Objekten (Entitäten), deren Eigenschaften und Ausprägungen (Anapher) sowie den beschriebenen Beziehungen untereinander und
- eine gleichwertige semantische Verarbeitung von verschiedensprachigen Dokumenten innerhalb eines Analysevorganges.

Sie ermöglicht

- umfangreiche Recherchemöglichkeiten innerhalb von «public sources», wie Internet, RSS-Feeds, Newsgroups, «deep web» etc. unter Ausnutzung von Crawlern, Digestern, Summarizern sowie NLP-Software und Ontologien zu fahren,
- über alle bisher genannten gewonnenen Daten eine echte semantische Suche durchzuführen und
- mittels offener Architektur (SOA, Web Services, XML etc.) Daten auszutauschen und in Drittprodukte zu integrieren.

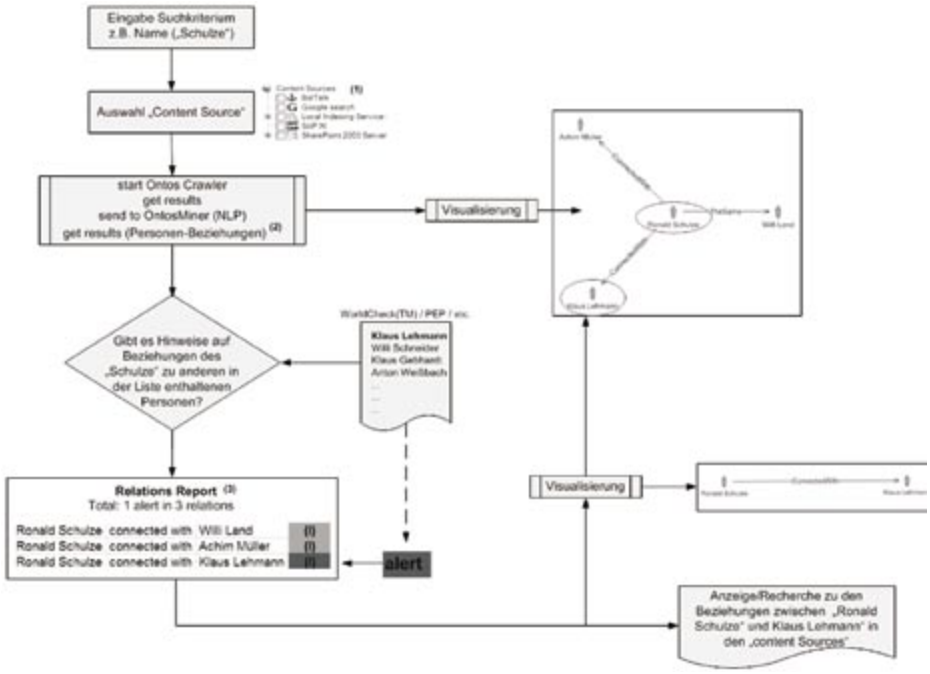
Die Datenquellen können unterschiedlicher Herkunft sein, so u.a.

- Online-Datenquellen (LAN, WAN, Internet, E-Mail),
- Local Indexing Services,
- SharePoint Server,
- sonstige polizeiliche Informations- und Auskunftssysteme (Bund/Land) / operative Ermittlungsdatenbanken, Vorgangsbearbeitungssysteme und auch
- ERP/CRM-Systeme

bzw. kann darauf während des Prozesses zugegriffen werden.

Einsatzmöglichkeiten sind besonders in den Bereichen gegeben,

- wo grundsätzlich bzw. auch «ad hoc» mit sehr vielen unstrukturierten, aber auch strukturierten Daten gearbeitet wird (z.B. Wirtschaftskriminalität / Korruptionsbekämpfung nach Beschlagnahmung umfangreicher Asservate (z.B. forensische I&K), Geldwäschebekämpfung, Umsatzsteuerkarussell oder im Bereich der «spurenintensiven Ermittlungen») und / oder
- kurzfristig Informationen aus verschiedenen Quellen mit unterschiedlicher Struktur zu bewerten sind.



Möglicher workflow

Literatur, Quellen, Fussnoten

Internet: Wikipedia; eigene Recherchen; «Kriminalistenfachbuch», Schmidt Römhild Verlag.

1. Der Begriff Daten subsummiert hierbei alle elektronisch vorliegenden Daten, also vom eingescannten und mittels OCR-Software (Optical Character Recognition, auch Texterkennung) editierbaren Papierdokument bis hin zu auf einem Datenträger gespeicherten Dateien (Textverarbeitung, Internetseiten, E-Mail-Korrespondenz usw.). Siehe auch: forensische I&K
2. NLP: Natural language processing; In der Computerlinguistik (englisch: Computational Linguistics oder Natural Language Processing) wird untersucht, wie natürliche Sprache mit Hilfe des Computers algorithmisch verarbeitet werden kann. Sie ist Teilbereich der künstlichen Intelligenz und gleichzeitig Schnittstelle zwischen Sprachwissenschaft und Informatik. Quelle: Wikipedia.
3. z.B. Personen, Firmen, Ortsbezeichnungen, Datum, Geldbeträge usw.
4. Herr Müller arbeitet bei der Firma Schneider GmbH. Er ist mit Frau Christa Schneider verheiratet.
5. z.B. beschlagnahmte papierne Asservate (Leitz-/Bundesordner) nach Durchsuchungsaktionen bzw. sichergestellte Computerfestplatten bzw. Treffermengen bei einer Internetrecherche.
6. Sicherlich in den meisten Fällen in eine Tabellenkalkulation oder Datenbank.
7. local indexing system: Windows legt von allen Festplattenlaufwerken Indizes an, um die Suche nach Dateien zu beschleunigen.
8. Ausgangspunnkt der Recherche ist die Suche nach Informationen zur Person «Müller». Das System soll nun einen Warnhinweis ausgeben, insofern eine Beziehung dieser Person zu einer Person «Scheider» oder «Lehmann» erkannt wurde. Diese Indikatoren werden im System in einer separaten Stopliste geführt; wo diese gespeichert ist, gibt man im Feld «Exclusion file path» an: Gleiches ist für jeglichen Objekttyp (Firma, Konto usw.) anwendbar.
9. z.B. verschiedene, online recherchierbare Wirtschafts-und Firmeninformationsdatenbanken
10. auch: Initiativermittlungen
11. Siehe: Marbella-Connection («Firmenbestatter»), Computerbetrug mit Hilfe von Mehrwertdiensten
12. Hierzu können auch, neben den bereits oben aufgeführten, auch das Internet, Zeitungen, Zeitschriften, Newsletter, Diskussionsforen usw. mit einbezogen werden.
13. z.B. Anforderung von Prospekten unter einer Legende, Erwecken von Scheininteressente, offenes Ansprechen der Person/des Unternehmens (Präventionsfaktor)
14. auch: anti money laundering (AML)
15. siehe: public sources ◆