poolparty®

WHITE PAPER

Datenbestände mit semantischen Technologien bereinigen

Bruno Wildhaber (Competence Center Records Management)
Daniel Spichty (Competence Center Records Management)
Andreas Blumauer (Semantic Web Company)

Erstellt von

k s|rm

SEMANTIC WEB COMPANY

Inhaltsverzeichnis

Ausgangslage und Problemstellung

Daten werden oft als das 'neue Öl' bezeichnet, diese Meinung setzt sich allmählich in der Breite durch. Kaum ein Geschäftsmodell, welches heute ohne Daten als Grundlage funktioniert [1]. Compliance Anforderungen, wie der Datenschutz sind nur dann erfüllbar, wenn die Daten bekannt sind. Tatsächlich verursacht das Datenwachstum und die chaotische Datenbewirtschaftung in den meisten Organisationen ständig steigende Kosten und vor allem auch Unmut bei den Anwendern. Ein wesentlicher Aspekt dabei ist, dass immer mehr 'Datenmüll' ("digital Landfills" = digitale Mülldeponien) entsteht, also vollkommen obsolete Daten unnötiger Weise weiterhin gespeichert und verwaltet werden. Internationale Studien besagen, dass rund 70 bis 80 % aller gespeicherter Daten schlicht unnötig sind. Diese stehen beim Finden nach wertvollen Informationen im Weg und erzeugen unnötiges 'Rauschen'.

Die Kernfrage lautet also: **Wie kann man die wertvollen von den ROT-Daten trennen?** (ROT steht hier für Redundant, Outdated, Trivial, also Daten, welche doppelt vorhanden sind, veraltet oder keine Bedeutung für die Organisation haben).

Die heute vorhandenen Datenmengen lassen eine manuelle Bereinigung der Daten nur noch in Ausnahmefällen zu. Folglich ist eine automatisierte Lösung erforderlich. Ein wichtiger Ausgangspunkt bildet dabei die automatische Klassifikation aller Datenbestände, um ROT- Datenobjekte laufend eliminieren zu können.

In Folge beschreiben wir eine innovative Methode (MATRIO® Data Cleanup Method[1]) auf Basis semantischer Wissensmodelle. Die Bedeutung semantischer Technologien bringt Gartner auf den Punkt [2]: "Unprecedented levels of data scale and distribution are making it almost impossible for organizations to effectively exploit their data assets. Data and analytics leaders must adopt a semantic approach to their enterprise data assets or face losing the battle for competitive advantage." MATRIO® erlaubt das Durchführen performanter und präziser Datenanalysen, die ein zielgerichtetes Aussortieren obsolet gewordener Daten in heterogenen Datenbeständen erlaubt ohne dabei z.B. Compliance Regeln zu verletzen.

Herausforderungen und Erfolgskriterien

Vier Erfolgskriterien stehen im Mittelpunkt der Überlegungen:

1. Wie kann man zusätzlich zu einfachen Mechanismen (z.B. ein Ablaufdatum wurde überschritten) auch

1 MATRIO Methode: http://www.matrio.swiss

komplexe Regeln (z.B. basierend auf den eigentlichen Inhalt) zum automatischen Aufräumen nicht nur implementieren, sondern auch wartbar machen?

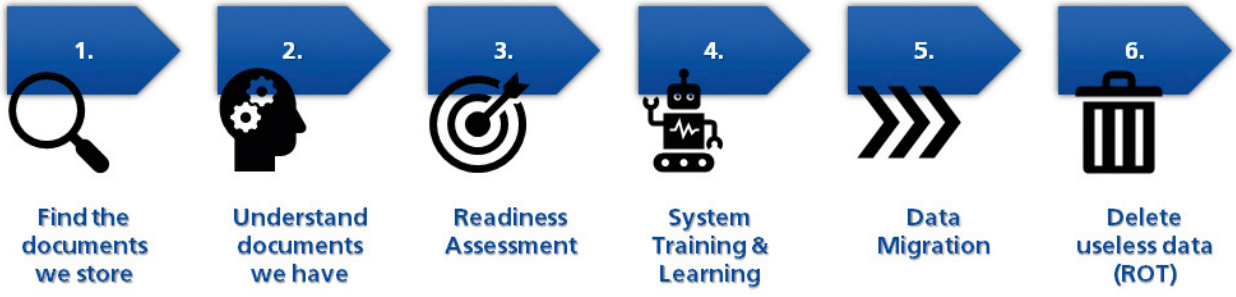
2. Wie kann dies auch dann gelingen, wenn man diese Regeln nicht nur auf einzelne Datensilos anwenden will, sondern auf Basis heterogener aber vernetzter Daten?
3. Wie kann man diese teilweise komplexen Operationen performant anwenden und möglichst präzise die Klassifikation durchführen, auch auf umfangreiche und verteilte Daten?
4. Wie kann der Lebenszyklus von Datenobjekten wieder zurückgewonnen werden, um damit die Daten von der Entstehung über die Archivierung bis zur automatisierten Löschung zu kontrollieren.

Auf Ebene des Governance Modells stehen u.a. folgende Fragen:

- Welchen Qualitätskriterien müssen die wertvollen Daten genügen und wer legt diese fest?
- Was alles kann "ROT"' bedeuten? Wer hat hier die Deutungshoheit? Wie werden diese ROT-Daten vernichtet (forciertes Löschen, ev. protokollierte Kassation)?
- Wer übernimmt die Verantwortung für das kontrollierte Vokabular (Stammdaten sind ein guter Startpunkt) und wie werden Taxonomien gepflegt?
- Wie ist mit Metadaten umzugehen, die nicht oder schwer interpretierbar sind, die inkomplett oder widersprüchlich sind?
- Wer in einer Organisation kann Datenobjekte als 'obsolet' qualifizieren, wie ist mit Unstimmigkeiten umzugehen? Wie kann dabei das Wissen der Domänenexperten angezapft und modelliert werden?
- Wie ist mit unverbundenen Datenfragmenten umzugehen, über die keine eindeutige Aussage getroffen werden kann, ob die Daten aussortiert werden sollten?

MATRIO® Data Cleanup Methode

Dies ist der Ablauf der MATRIO® Data Cleanup Methode:



Schritt eins: Identifikation der Dateiablage

In einem ersten Schritt geht es darum, die gespeicherten Dateien zu finden und grob zu beschreiben ("Find the documents we store"). Die Dateiablagen (Repositories) müssen identifiziert und Metriken gesammelt

werden. Das klingt zwar trivial ist aber in Cloud-Lösungen und dezentralen IT-Systemen durchaus anspruchsvoll. Netzwerklaufwerke, SharePoint Bibliotheken, One Drive, Google Docs etc. sind nur einige Vertreter von Repositories für Dateien. Die gesammelten Metriken über die Dateien sind erste wertvolle Informationen. Sie geben Auskunft darüber, um was für Dokumente es sich überhaupt handelt. Sind es gescannte Dokumente als PDF, für welche es keinen Volltext gibt (nur Bilder)? Handelt es sich um Containerformate wie Zip-Dateien, E-Mail Daten in PST Dateien? In diesem Schritt können neben den Umgebungs-Metadaten (z.B. Dateiattribute) auch bereits Inhaltliche Metadaten gelesen werden. Dies ist sinnvoll, weil damit auch Hashwerte der Dateien für die sichere Identifikation der Redundanzen gerechnet werden. Zudem kann nur mit inhaltlichen Metadaten eine erste Aussage über die weitere Verarbeitbarkeit der Dateien gemacht werden kann. Mit Passwort verschlüsselte Dateien können ebensowenig verarbeitet werden, wie korrupte d.h. nicht mehr lesbare Dateien. In diesem ersten Schritt werden die Ausnahmebehandlungen (Exception Handling) definiert und die Dateien für die weitere Verarbeitung vorbereitet: Flachklopfen von Containerformaten (z.B. auch rekursive Zip-Dateien, die, PST-Dateien enthalten, die wiederum MSG-Dateien, und MSG mit Anhängen, etc.), OCR für Bilder, PDF, etc. Der erste Schritt ist Industriestandard und entspricht dem Status Quo.

Schritt zwei: Beschreiben der Unternehmensdaten

Im zweiten Schritt geht es um das Verstehen und Beschreiben der Unternehmensdaten (“Understand and describe the documents we have”). Dieser Schritt kann parallel zum ersten gestartet werden und findet auf der normativen Ebene statt. Bereits vorhandene Aktenpläne sind ein guter Startpunkt, greifen jedoch zu kurz, da diese lediglich gesetzlich zu archivierende Dokumente beschreiben. Zudem machen Aktenpläne keine Aussagen zur Vertraulichkeit von Dokumenten oder auch nicht darüber, ob es sich um Personendaten oder sogar besonders schützenswerte Personendaten handelt.

Mit Hilfe von Taxonomien und Verfahren des Text Minings werden Dokumente analysiert und nach festgelegten Kriterien klassifiziert, d.h. in Kategorien eingeordnet. Zur näheren Beschreibung von Unternehmensdaten (vor allem strukturierte Daten) können Taxonomien in Kombination mit Ontologien verwendet werden. Eine Ontologie erstellt einen Rahmen, der eine Wissensdomäne beschreibt, indem sie Klassen, Beziehungen und Einschränkungen festlegt, die auf die Konzepte und Entitäten wirken [3]. Ontologien und Taxonomien bilden die Basis für ein gemeinsames Vokabular (“Semantic Layer”), das sicherstellt, dass der verwendete Wortschatz ein Themengebiet (“Wissensdomäne”) stets konsistent beschreibt.

Stammdaten oder auch gepflegte Glossare sind eine gute Quelle für den Aufbau eines gemeinsamen, kontrollierten Vokabulars. Ausgeklügelte semantische Technologien wie PoolParty Semantic Suite[2] sind in der Lage, die verschiedenen vorhandenen Quellsysteme zu verbinden und stellen Funktionen zur Verfügung, um die notwendigen Taxonomien, Ontologien und unternehmensweiten Wissensgraphen für die Beschreibung der Daten zu erstellen.

---

2 PoolParty Semantic Suite: https://www.poolparty.biz/

Schritt drei: Einschätzung der Durchführbarkeit

Das Readiness Assessment stellt sicher, dass alle notwendigen Grundlagen für den erfolgreichen Data Cleanup vorhanden sind. Dabei werden auf der normativen, strategischen, taktischen und operativen Ebene systematisch allfällige Gaps identifiziert und Massnahmen für das Schliessen der Gaps festgelegt. Die Ermittlung des Reifegrades der Information Governance einer Organisation ist Teil dieses Schrittes und kann zu Beginn des Data Cleanups erfolgen (siehe dazu auch das separate Kapitel “Voraussetzung: Maturität ermitteln”).

Schritt vier: System trainieren

Im vierten Schritt wird das System trainiert (“System Training & Learning”). Mit repräsentativen Stichproben aus den vorhandenen Dateien beginnt damit die Lernphase. Dabei werden Dinge (Entitäten) aus den Dokumenten extrahiert und miteinander verknüpft. Die statistisch basierten Ansätze für die automatische Klassifikation sind in der künstlichen Intelligenz eine weit verbreitete Methode, um Systeme zu trainieren. Wird dieses Verfahren aber mit semantischen Methoden der automatischen Metadaten-Anreicherung kombiniert, kann ein Klassifikator vielfach schon mit kleineren Mengen von Trainingsdaten trainiert werden. Dabei werden Dokumente, die zum Trainieren des Klassifikators dienen um semantische Metadaten angereichert, was in vielen Fällen auch dazu führt, dass der Klassifikator präziser wird. Ist das zuvor festgelegte Konfidenzniveau in der Lernphase schliesslich erreicht, kann mit dem nächsten Schritt begonnen werden. Der zweite (“Understand and describe the documents we have”) und dieser Schritt werden solange in Iterationen durchlaufen, bis die Beschreibung der Daten die Wirklichkeit genügend gut abbildet. Mit Hilfe semantischer Metadaten, die in Form von Wissensmodellen beschrieben werden, können die Ergebnisse der Klassifizierung deutlich verbessert und Metadaten kontrolliert zugeordnet werden.

Schritte fünf und sechs: Daten Migration und Löschen von ROT Daten

Der fünfte (“Data Migration”) und der sechste Schritt (“Delete useless data”) sind die logische Fortführung der vorhergehenden Schritte. Das Aufräumen der Dateien (Schritt 5) erfolgt hoch automatisiert. Dabei werden nicht nur die obsoleten Dateien verlässlich ausgesondert, sondern die wertvollen Dokumente erhalten ihren Lebenszyklus zurück. Im Rahmen dieses Aufräumens können Dokumente auch für die Migration in ein ECM System bzw. elektronisches Archiv vorbereitet werden, indem die notwendigen Business Metadaten aus der Taxonomie und Semantik automatisch generiert werden. Das Loswerden der obsoleten Daten ist Teil des letzten und sechsten Schrittes. Ein risikobasierter Ansatz legt die konkrete Umsetzung fest. Von einer forcierten, unmittelbaren und automatisierten Datenlöschung über eine durch Informationseigner gesteuerte und mit Bewilligungsprozess unterstützte, hoch-kontrollierte Vernichtung bis zum friedlichen Sterben der Daten über die Zeit ist alles denkbar.

Mit einer lückenlosen Kontrollkette und vollständiger Dokumentation wird die Konformität und Ordnungsmässigkeit sichergestellt.

Anwendungsfälle

Fallbeispiel Pharmaunternehmen

Eine international tätige Pharmafirma muss den nationalen Gesundheitsbehörden umfangreiche Unterlagen für die Zulassung und die periodische Überprüfung der Medikamente einreichen. Die eingereichten Unterlagen liegen unkontrolliert auf verschiedenen Dokumentenablagen wie Netzwerklaufwerk, SharePoint etc, welche über die Zeit durch Firmenübernahmen und Re-Organisationen entstanden sind. Für die kontrollierte Ablage wurde ein ECM System und ein elektronisches Archivsystem aufgebaut. Jedoch muss für die Migration der 2 Millionen Dateien eine Lösung her. Mit der oben beschriebenen Methode konnten im ersten Schritt 50 % der Daten als redundant erkannt werden. Es müssen damit “nur” 1 Million Dateien klassifiziert und migriert werden. Zudem wurde im Schritt 1 festgestellt, dass 20 % der Daten in Containern gespeichert sind (7z und PST Dateien). Weitere 10 % der Daten sind E-Mail Daten und haben in den Dokumentenablagen nichts verloren (gemäss Firmenpolicy). In der Vorbereitungsphase wurden die Containerformate ausgepackt und die E-Mail Daten in das E-Mail Archiv verschoben.

Fallbeispiel NGO

Eine weltweit tätige Schweizer Stiftung für technische Zusammenarbeit ist in 36 Ländern ausschliesslich in der internationalen Entwicklungszusammenarbeit tätig und führt eigene und mandatierte Projekte durch.

Auf Grund der hohen Personalfluktuation ist das Onboarding neuer Projektmitarbeiter kritisch: Wie bekommt der neue MA möglichst schnell Zugriff auf die relevanten Projektdaten? In einer hoch verteilten Umgebung eine grosse Herausforderung, zumal klassische Knowledge-Management Praktiken hier oftmals nicht wirken. Entweder ist der Vorgänger schon weg oder befindet sich fernab in einem anderen Land. Dies führt auch dazu, dass die verteilten Daten auf unterschiedlichsten Systemen liegen und keine einheitlichen Ablagestrukturen definiert wurden. Das Zusammensuchen von Daten wird entsprechend schwierig und für den neuen Mitarbeiter zur Sisyphusarbeit. In einem solch dynamischen Umfeld muss gewährleistet werden, dass zumindest die Beschreibung der Daten vorgegeben wird. Dies geschieht mittels einer zentral gesteuerten und aktiv betreuten Taxonomie, welche durch die Analyse von erfassten Dateninhalten (Textanalyse) laufend ergänzt wird. Dies ermöglicht dem neuen Projektleiter, seine Daten und Dokumente einfacher zu finden und zu verwalten. Im Idealfall hat er zusätzlich ein Dokumenten Management System zur Verfügung, welches das Lifecycle Management ideal ergänzt. Nebst den inhaltlichen Metadaten werden rechtliche Metadaten eingebunden, z.B. die Kennzeichnung von Dokumenten im Zusammenhang mit Projektgenehmigungen oder solche, die auf Grund nationaler Gesetzesvorschriften über eine vordefinierte Dauer archiviert werden müssen. Auch die Zusammenführung von Einzeldokumenten zu Dossiers erfolgt mit Hilfe kontrollierter Metadaten.

Fallbeispiel aus dem privaten Umfeld

Eine private Fotosammlung soll aufgeräumt und neu strukturiert werden. In Zeiten hochperformanter Kameras als Bestandteil jedes mobilen Endgeräts häufen sich in jeder Familie schnell an die 100.000 Fotos an. Gelinde gesprochen sind davon vielleicht 25% auch interessant genug, um sie in das Langzeitarchiv der Familie aufzunehmen. Sind nun mehrere Familienmitglieder als Fotografen, als “Qualitätsmanager” und als User in diesen Vorgang involviert, so nimmt die Komplexität und der Umfang der anzuwendenden Klassifikationsregeln schnell zu, wiewohl auch das zugrundeliegende Governance Modell schnell kompliziert werden kann.

Eine wesentliche Herausforderung in diesem Beispiel ist die Deduplizierung des Bestandes, wobei bei ‘sehr ähnlichen’ Fotoaufnahmen von unterschiedlichen Fotografen die Regel zum Teil sehr komplex aussehen können.

Unterschiedliche Klassifikationsregeln (was der eine in den Folder ‘Strand’ aufnimmt, würde der andere im Ordner ‘Badespass’ ablegen) führen schnell zu inkonsistenten Metadaten, was später in der Cleanup-Phase das grösste Hindernis darstellt.

Selbst dieses kleine Beispiel zeigt bereits, wie wichtig konsistente semantische Metadaten sind, die mit Hilfe kontrollierter Vokabulare bzw. Taxonomien erzeugt werden sollten, anstatt freihändig vergeben zu werden.

Wie semantische Technologien funktionieren

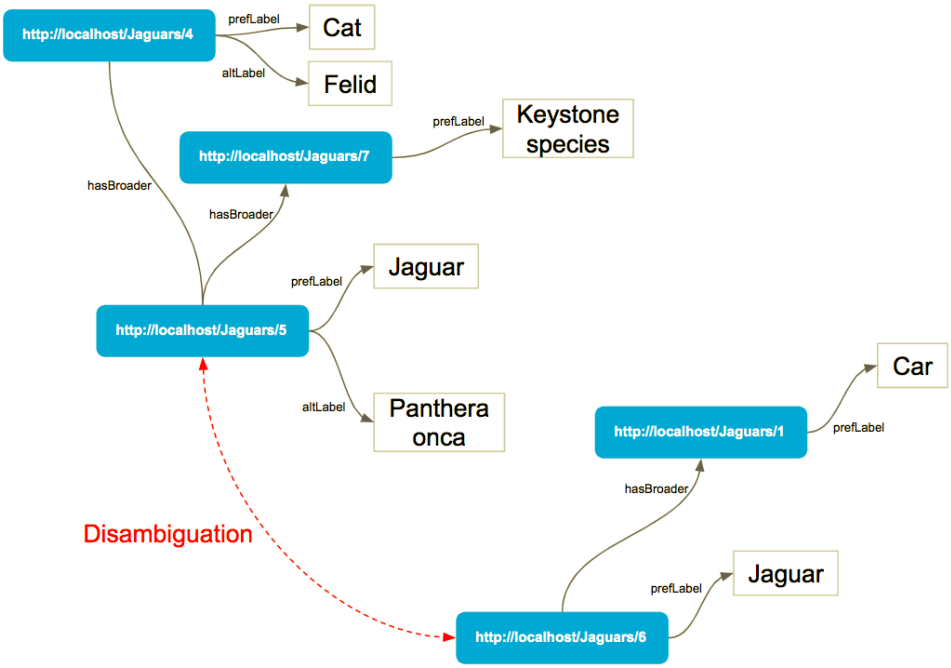
Semantic Web Technologien basieren auf offenen Standards des World Wide Web Consortium (W3C)[3], welche die Grundlage für das Web of Linked Data bilden. Basierend auf denselben Technologien, die im herkömmlichen Web zum Austausch und zur Darstellung von Dokumenten dienen (URIs, HTML und http), ermöglicht das Semantic Web die Verwaltung und den Austausch von Wissen mittels sogenannter Wissensgraphen.

Ein semantischer Wissensgraph[4] setzt sich aus Konzepten und ihren Beziehungen zusammen. Ein Konzept verfügt über natürlichsprachige Benennungen (sogenannte ‘labels’) und einen eindeutigen Bezeichner, den sogenannten Uniform Resource Identifier (URI). Die Benennungen und andere Literale dienen der Speicherung und Darstellung menschlich lesbarer Information, und werden darüber hinaus für das automatische Textmining verwendet. Konzepte sind digitale Vertreter von konkreten Gegenstände, oder von abstrakten Gedanken. Somit ist dieser Ansatz für alle Arten von Entitäten universell anwendbar.

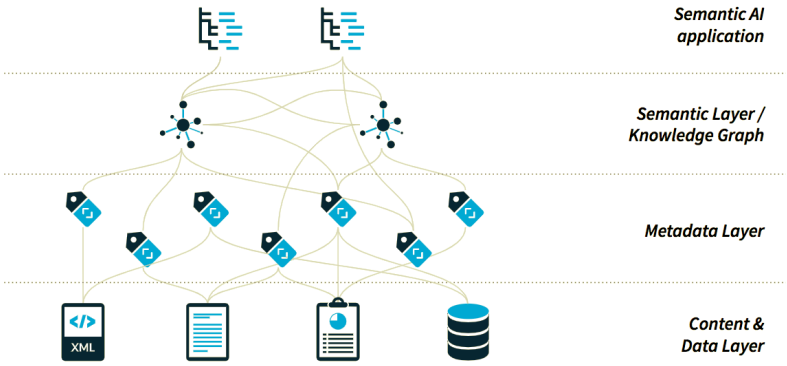
---

3 W3C: Semantic Web: https://www.w3.org/standards/semanticweb/

4 What is a Knowledge Graph? https://www.poolparty.biz/what-is-a-knowledge-graph



Die Entitäten werden als Baum bzw. Netzwerk organisiert und beschrieben, beispielsweise in Form eines kontrollierten Vokabulars. Konzepte sind also als Dinge zu verstehen, die verschiedene Begriffe (Deskriptoren und deren Synonyme) angefügt haben. Zudem können sie assoziative - oder Eltern-Kind-Beziehungen zu anderen Konzepten eingehen. Damit verfügen sie über einen Bedeutungskontext, der eindeutige Zuordnungen erlaubt (Disambiguation), womit Entscheidungen getroffen werden, die mit kognitiven Prozessen vergleichbar sind. Dieser Prozess ist immer transparent, da das Modell die Beziehungen und Beschreibungen der Konzepte sichtbar, kontrollierbar, nachvollziehbar und auf Standards beruhend, änderbar macht.



Daten automatisiert aufräumen bzw. entsprechend klassifizieren können, gelingt dann, wenn die Bedeutungen und auch Zusammenhänge von Daten und Metadaten exakt (genug) bestimmt werden können. Dies gelingt mit Hilfe semantisch basierter Wissensgraphen.

Unser Ansatz: Datenräume aufräumen mit semantischen Technologien

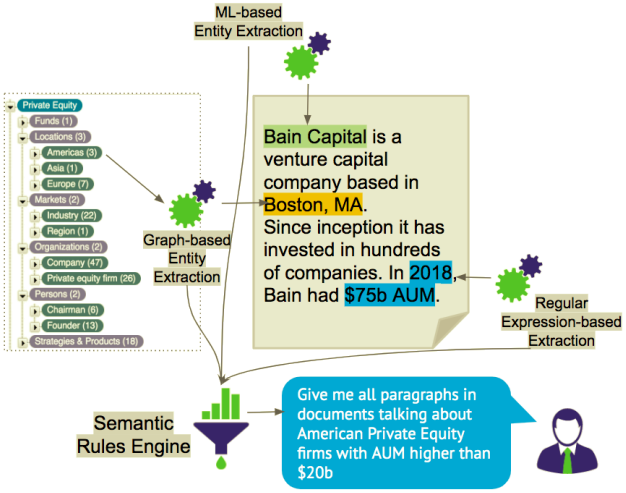
Voraussetzung: Reifegrad (Maturität) ermitteln

Es versteht sich von selbst, dass nicht jede Organisation dieselben Voraussetzungen für den Einsatz der MATRIO® Data Cleanup Methode mitbringt. Es lohnt sich deshalb, vor dem Einsatz einige Eckwerte zu ermitteln. Dabei geht es grundsätzlich um den Umgang mit Daten, sowohl aus Sicht der Daten-Landschaft (Wo sind meine Daten?) wie auch der logischen oder semantischen Betrachtung (Welche Daten habe ich?) sowie der Frage der Organisation und dem Umgang mit Daten (Steuerung der Datenhaltung durch Richtlinien und andere Vorgaben). Insgesamt stellt sich die Frage, welchen Reifegrad die Information Governance Organisation aufweist.

Im Rahmen eines Online-Assessments erheben wir u.a. die folgenden Parameter und halten in einem Bericht fest, wie es um die Reife der Information Governance steht und welche Aktionen empfohlen werden:

- Übersicht über die tatsächlich gespeicherten Daten
- Aussagen zur Datenqualität: Festlegen der Kriterien
- Bereitschaft zur übergreifenden Zusammenarbeit in der Organisation
- Vorgaben / Policy / Weisungen
- Bereitschaft zur Investition in Daten
- Bereitschaft zur Nutzung neuer Technologien / AI
- Komplexität der Dateninfrastruktur und der Datenhaltung
- Strategie im Umgang mit Daten
- Organisation (Operating Model: Personen, Prozesse, Rollen)
- Wissensmodelle / Knowledge Management
- Verantwortlichkeit für Daten / Informationen
- Willen des Managements zur Investition in “Wissen”

Der Beitrag semantischer Technologien



Zusätzlich zur Daten- und Metadatenschicht verwaltet das Unternehmen semantische Wissensmodelle, die zur automatisierten Analyse von Dokumenten und Datenobjekten herangezogen werden können. Diese Verfahren werden mit Möglichkeiten der Datenanalyse und Textklassifikation basierend auf maschinellem Lernen kombiniert. So werden auch komplexere semantische Bedeutungsmuster für die Maschine interpretierbar [4].

Wirtschaftlichkeit

Es gibt wenige Verfahren und Systemumstellungen, deren Nutzen so einfach berechnet werden kann wie die Bereinigung von Daten. Dabei bietet sich einerseits die ROI (Return on Investment) Betrachtung an, gleichzeitig gibt es auch eine grosse Zahl von nicht quantitativen Nutzen, die sich auflisten lassen.

Bei der Berechnung des ROI setzen wir nur auf jene Grössen, die sich tatsächlich auch messen lassen. Wie wir wissen, sind die Kosten der IT oftmals höchstens als globaler Posten oder auf Projektebene ausgewiesen. Wir müssen folglich mit Grössen arbeiten, die sich einfach ableiten lassen. Dabei eignen sich z.B. die Gesamtkosten der IT in Relation zur totalen in der Organisation gespeicherten Datenmenge. Grundsätzlich kann man davon ausgehen, dass jede gespeicherte Dateneinheit eine Vielzahl an Sekundärkosten erzeugt. Oder anders gesagt: Jedes GB an gespeicherten Daten erzeugt direkte Kosten (Speicherkosten, Betriebskosten), die anteilsmässig umgelegt werden können. Die reinen Speicherkosten sind dabei normalerweise vernachlässigbar, von Interesse sind die Umlagen, wie z.B. Kosten für die Sicherheit, für den Betrieb, das Monitoring, die Wartungskosten etc. Geht man davon aus, dass 80% der Daten einer durchschnittlichen Organisation überflüssig sind, dann ergibt sich schon ein gewaltiges Potenzial an Kosten, die gespart werden können. Ob man nun den Ansatz pro GB auf CHF 100 oder 500 veranschlagt, die Summe wird erklecklich. Noch einfacher wird es bei statistischen Zahlen, die uns auf Grund internationaler Studien vorliegen. Gemäss Studien verbringt der durchschnittliche “Knowledge Worker” mit einer Stunde Suchen -- pro Tag! Kann die Suchzeit um die Hälfte reduziert werden ergibt sich eine gewaltige Kosteneinsparung. Auch diese Messgrösse lässt sich einfach ermitteln und für die Kommunikation, auch gegenüber der Geschäftsleitung und dem Verwaltungs- oder Aufsichtsrat, nutzen.

Bei den qualitativen Zielen steht die Datenqualität im Vordergrund und der damit erzielbare Nutzen. In erster Linie aber die Einhaltung der immer strenger werdenden Gesetze rund um die Datenhaltung. So ist z.B. die Datensparsamkeit eine Schlüsselforderung jedes modernen Datenschutzgesetzes. Gleiches gilt für die Forderung nach der jederzeitigen Abrufbarkeit und Löschbarkeit von Personendaten. Nach einem Daten-Cleanup kann man mit hoher Sicherheit eine Aussage darüber treffen, wo Personendaten liegen, welchen Stand sie haben und ob sie gelöscht werden können: Schlüsselforderungen der Datenschutzgesetze, die heute kaum

eine Organisation erfüllen kann.

Weiterführende Literatur

[1] Information Governance: Ein Leitfaden mit Checklisten, Mustern und Vorlagen. Kompetenzzentrum Records Management, Bruno Wildhaber. https://www.amazon.de/Information-Governance-Leitfaden-Checklisten-Vorlagen-ebook/dp/B011IWV8L8

[2] How to Use Semantics to Drive the Business Value of Your Data. Gartner, Guido De Simoni.https://www.gartner.com/en/documents/3894095/how-to-use-semantics-to-drive-the-business-value-of-your

[3] Taxonomies vs. Ontologies. Forbes (Cognitive World), Kurt Cagle.https://www.forbes.com/sites/cognitiveworld/2019/03/24/taxonomies-vs-ontologies/

[4] Natural Language Processing with PoolParty. Semantic Web Company, Andreas Blumauer.https://www.poolparty.biz/wp-content/uploads/2017/03/Natural-Language-Processing-with-PoolParty.pdf

Kompetenzzentrum Records Management

Das Kompetenzzentrum Records Management (KRM) konzentriert seine Dienstleistungen auf Information Governance und betreibt das erste Kompetenzzentrum in Europa. Wir schaffen für unsere Kunden Mehrwerte, indem wir traditionelle Konzepte der IT- und Informationsmanagement-Wissenschaft mit der neuen Welt der Engagementsysteme (Social Media, Mobilität) verbinden. Wir bauen Brücken zwischen hochspezialisierten Bereichen und fördern ganzheitliche und unternehmensweite Lösungen. Wir verlieren jedoch nie den praktischen Sinn für Lebensfähigkeit, Pragmatismus und schnelle Erfolge. In der Kommunikation mit unseren Partnern wird immer ein interdisziplinärer Ansatz verfolgt, der sowohl unternehmerisches visionäres Denken als auch die erforderlichen Fachkompetenzen in allen angrenzenden Disziplinen beinhaltet. Der Hauptsitz von KRM befindet sich in Zürich, Schweiz. Um mehr zu erfahren, besuchen Sie https://informationgovernance.ch/.

Semantic Web Company

Semantic Web Company (SWC) ist der führende Anbieter von Graph-basierten Wissenstechnologien und die umfassendste semantische Middleware-Plattform auf dem globalen Markt. SWC ist Anbieter der PoolParty Semantic Suite, einer innovativen und sofort einsetzbaren Technologieplattform, die Unternehmen beim Aufbau und der Verwaltung von Knowledge Graphs als Grundlage für verschiedene KI-Anwendungen unterstützt.

Semantic Web Company wurde von KMWorld mehrfach als „Unternehmen, das im Wissensmanagement zählt" ausgezeichnet, und laut Gartner ist die PoolParty Semantic Suite ein „repräsentatives Produkt" für „Hosted AI Services". Neben Credit Suisse, Roche, Philips und der Asian Development Bank (ADB) werden zahlreiche Global 2000-Unternehmen durch die Semantic Web Company bei der Einführung ihres Wissensmanagements und ihrer KI-Strategie unterstützt.

Semantic Web Company hat seinen Hauptsitz in Österreich und eine Niederlassung in Großbritannien. Weitere Informationen finden Sie unter www.semantic-web.com und www.poolparty.biz oder folgen Sie ihnen auf LinkedIn und Twitter.

150+ customers trust us.

Featured Customers

PEARSON PHILIPS THE WORLD BANK IBRD • IDA | WORLD BANK GROUP

Awards and Recognitions

KMWorld 100 COMPANIES That Matter in Knowledge Management

KMWorld Trend-Setting Product

poolparty®

Phone (EU): +43 1 4021235
Phone (USA): +1 (415) 800-3776
Mail: info@poolparty.biz

Neubaugasse 1
1070 Vienna, Austria
www.poolparty.biz