Predictive Analytics auf Basis Unstrukturierter Daten

DGI Praxistage 2016
11. November 2016

Joachim Lackner
Business Development FSI
Expert System Deutschland GmbH

AGENDA

- Semantische Datenanalyse mit Expert System
- Anwendungen aus der Praxis
  - BPO im Schadenmanagement der Verssicherung
  - Produktmanagement /Entwicklung
- Expert System

- Fragen?

Predictive Analytics auf Basis unstrukturierter Daten

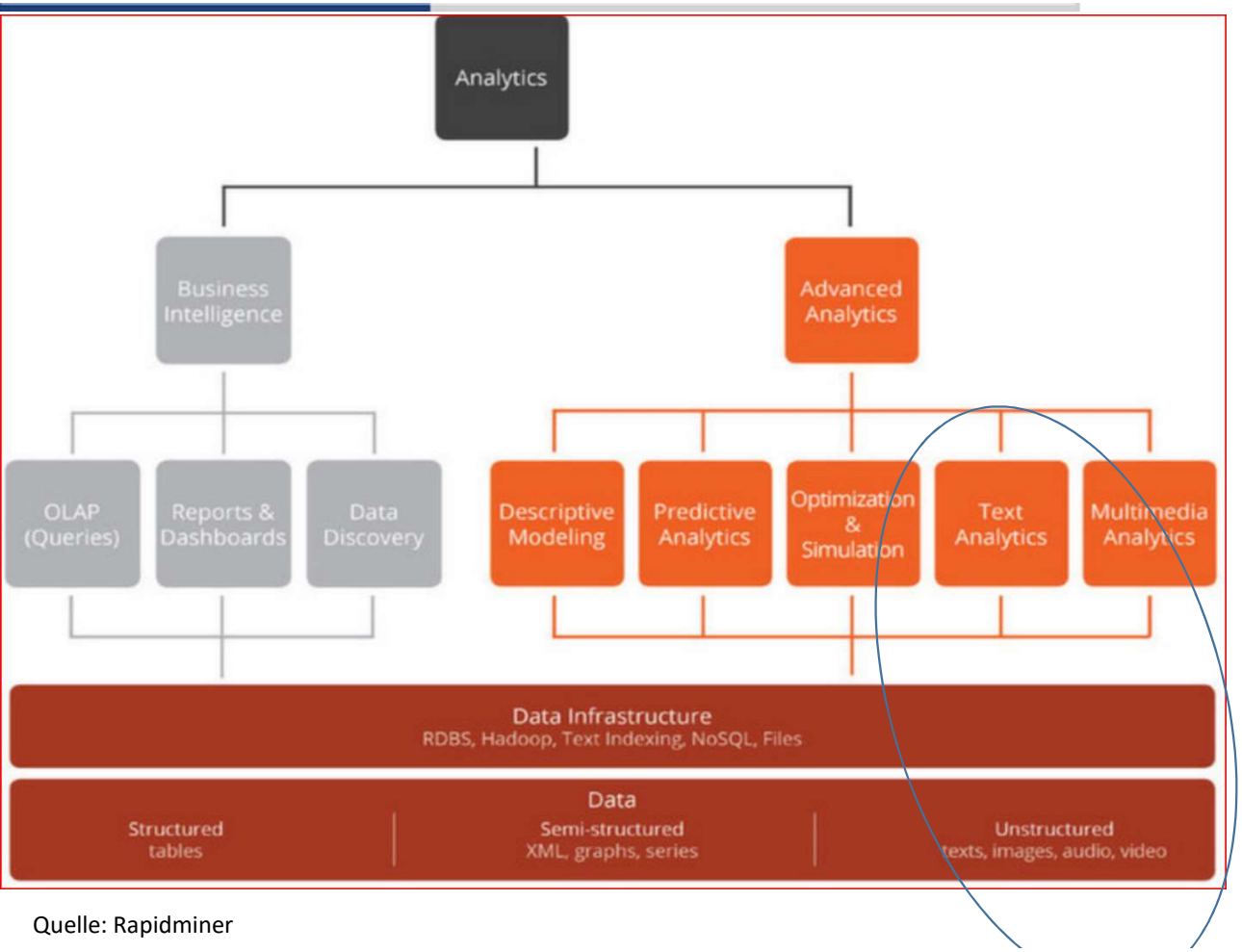
Wir kennen ihre Zukunft !?

Wir wissen (noch) nicht warum die Wahlvorhersage daneben lag !

Datenbasis ?, Datenmodell ? Datenart ?.....

Kann die Analyse unstrukturierter Daten ergänzend helfen ?

Predictive Analytics auf Basis unstrukturierter Daten



Quelle: Rapidminer

Predictive Analytics auf Basis unstrukturierter Daten als Inputfaktor

Unstrukturierte Daten bieten Wissen

Wissensextraktion mit Cognitive Computing*

*Cognitive Computing auch NLP, TextMining, Semantische Datenanalyse, Machine-Learning,

Vers. 19/04/2016

Predictive Analytics auf Basis unstrukturierter Daten

Semantik ?

Was macht diesen Ansatz so interessant ?

Technologie versteht die Zweideutigkeiten der Sprache als auch Textinhalte und Zusammenhänge

Gleiches Wort aber unterschied-liche Bedeutung	Verschiedene Wörter aber gleiche Bedeutung	Verschiedene Begriffe aber gleiche Bedeutung
Bank ?: Kreditinstitut oder Sitzgelegenheit?	kaufen oder erwerben ?	Hollywood oder die U.S. Film Industrie ?

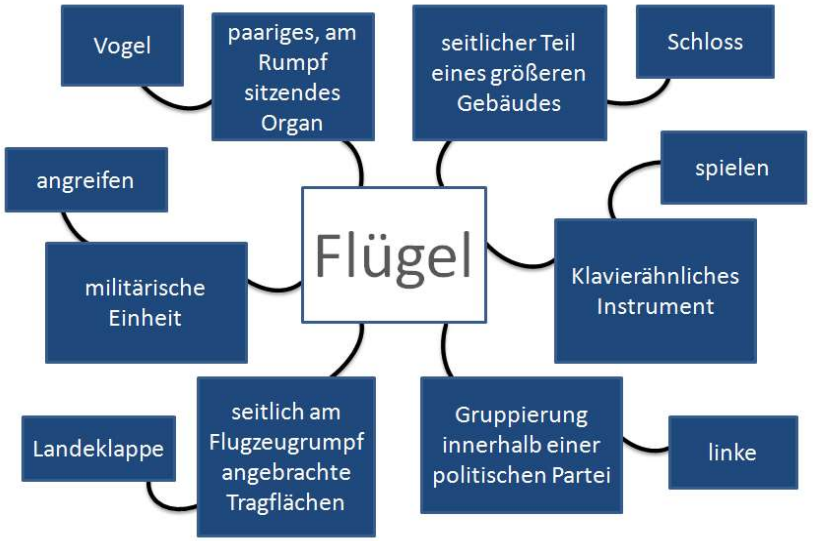
Semantik erschließt die DNA der Sprache.

Predictive Analytics auf Basis unstrukturierter Daten

Was ist eine semantische Plattform?

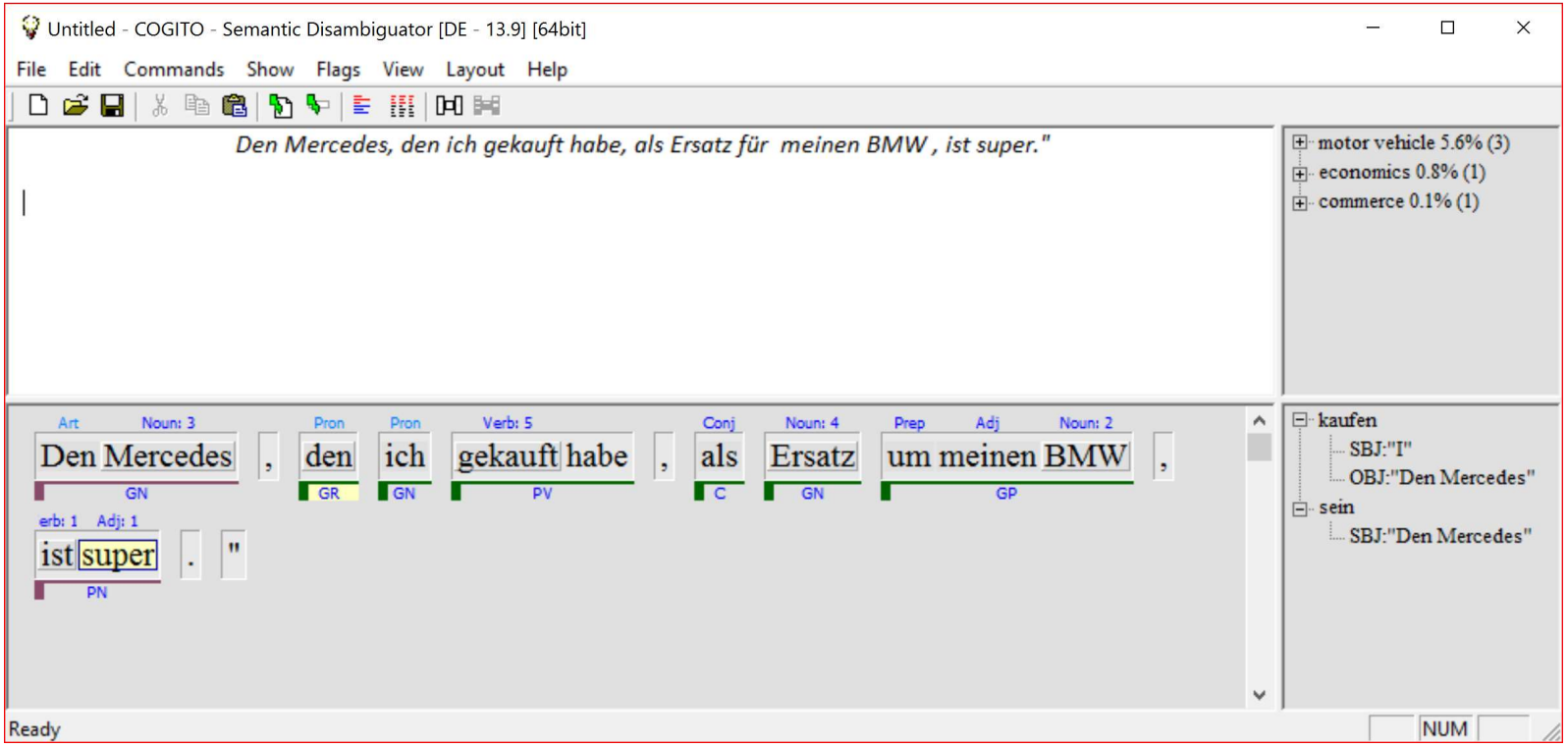
Semantische **Engine** und Semantisches **Netzwerk** (language ontology), um Zweideutigkeiten auszuräumen.

Unser semantisches Netzwerk ist eine umfangreiche „Concept-Map“, die sich aus Wortdefinitionen und Wortassoziationen zusammensetzt.



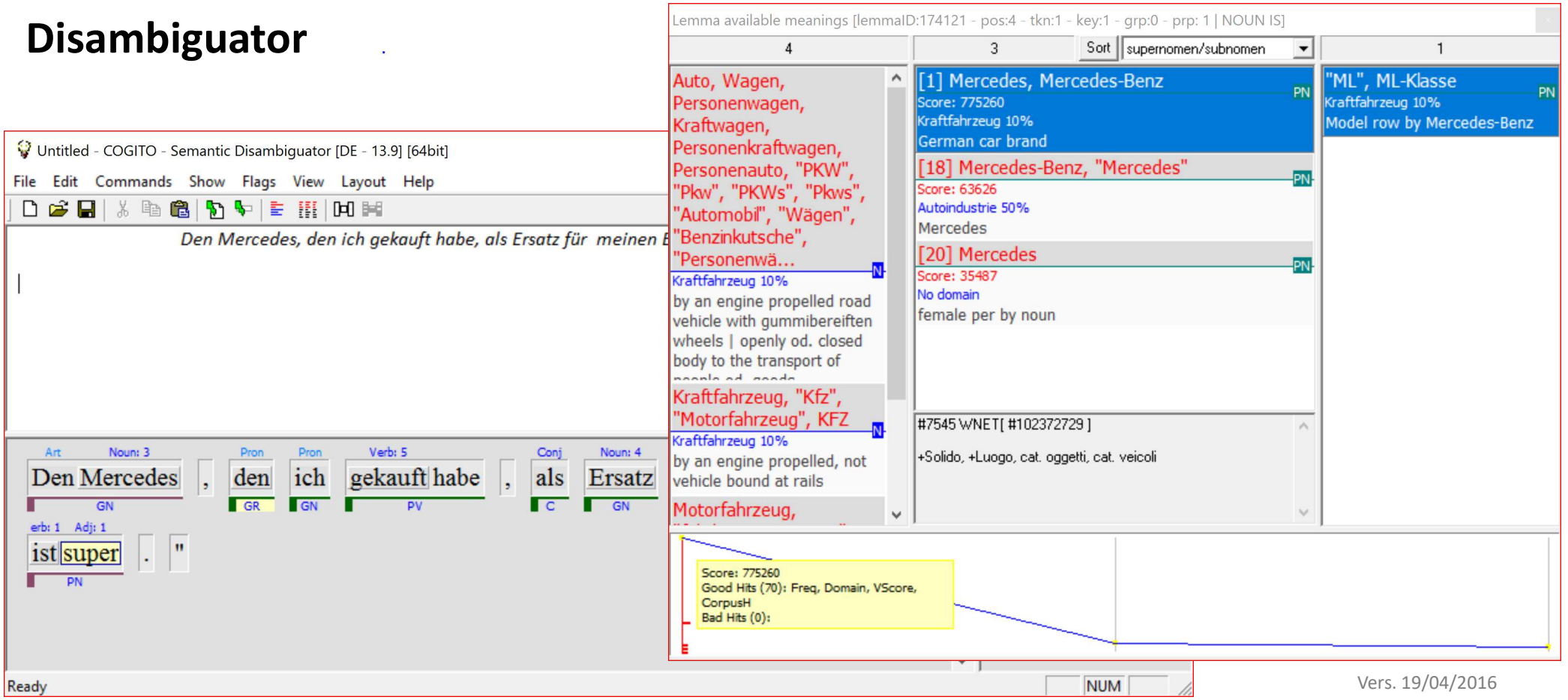
2 Millionen Konzepte und ihre unterschiedlichen Bedeutungen
10 Millionen Assoziationen zwischen diesen Bedeutungen

Disambiguator



Predictive Analytics auf Basis unstrukturierter Daten

Disambiguator



Vers. 19/04/2016

Predictive Analytics auf Basis unstrukturierter Daten

Was ist eine semantische Plattform?

Tiefgehende, linguistische Analyse eines Textes, um den Kontext verstehen zu können.

Analysestufen

Morphologische Analyse	Wortformen	Hund, Hundefänger, kalter Hund
Grammatikalische Analyse	Wortklasse	„Die Lebensdauer ausgewachsener Fliegen beträgt wenige Tage.“ (Substantiv)
		„Wir fliegen in den Urlaub.“ (Verb)
Logische Analyse	Wortbeziehungen	*„Den Mercedes, den ich gekauft habe, um meinen BMW zu ersetzen, ist super."*
Semantische Analyse	Kontext	„Er geht am Stock bis zur Bank und setzt sich.“
		„Er geht in der Bank in den ersten Stock.“
		„Wenn Robben hinter Robben robben, Robben robben Robben nach.“

Anwendungen aus dem Finanzsektor

Versicherungen:

- Schadenmanagement
- Underwriting
- Produktmanagement
- Customer Experience Management
- Legal Advisor
- Contract Certainty (Reconciliation)
- …….

Banken:

- Regulatorik, Risk Managment
- Customer Experience Management
- Legal Advisor
- Competitive Intelligence
- ……

Anwendungen aus dem Finanzsektor

Versicherungen:

- **Schadenmanagement**
- Underwriting
- **Produktmanagement**
- Customer Experience Management
- Legal Advisor
- Contract Certainty (Reconciliation)
- ……

Banken:

- Regulatorik, Risk Management
- Customer Experience Management
- Legal Advisor
- Competitive Intelligence
- ……

Use case:Schadenmanagement

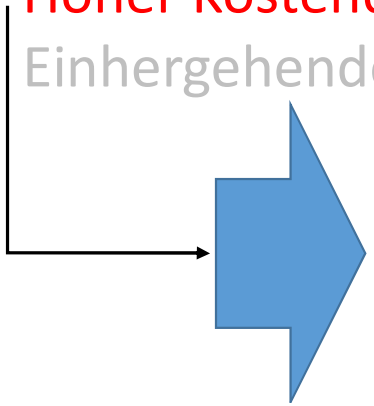
Schadenmanagement und Semantik ?????

- Hintergrund und Ausgangslage zum Prozeß: Schadenmanagement
- Was kann Semantik leisten ?
- Wie kann Semantik das Schadenmanagement optimieren ?
- Nutzen für den Versicherer ?

Schadenmanagement

Herausforderungen der Assekuranz - *Generalisiert*

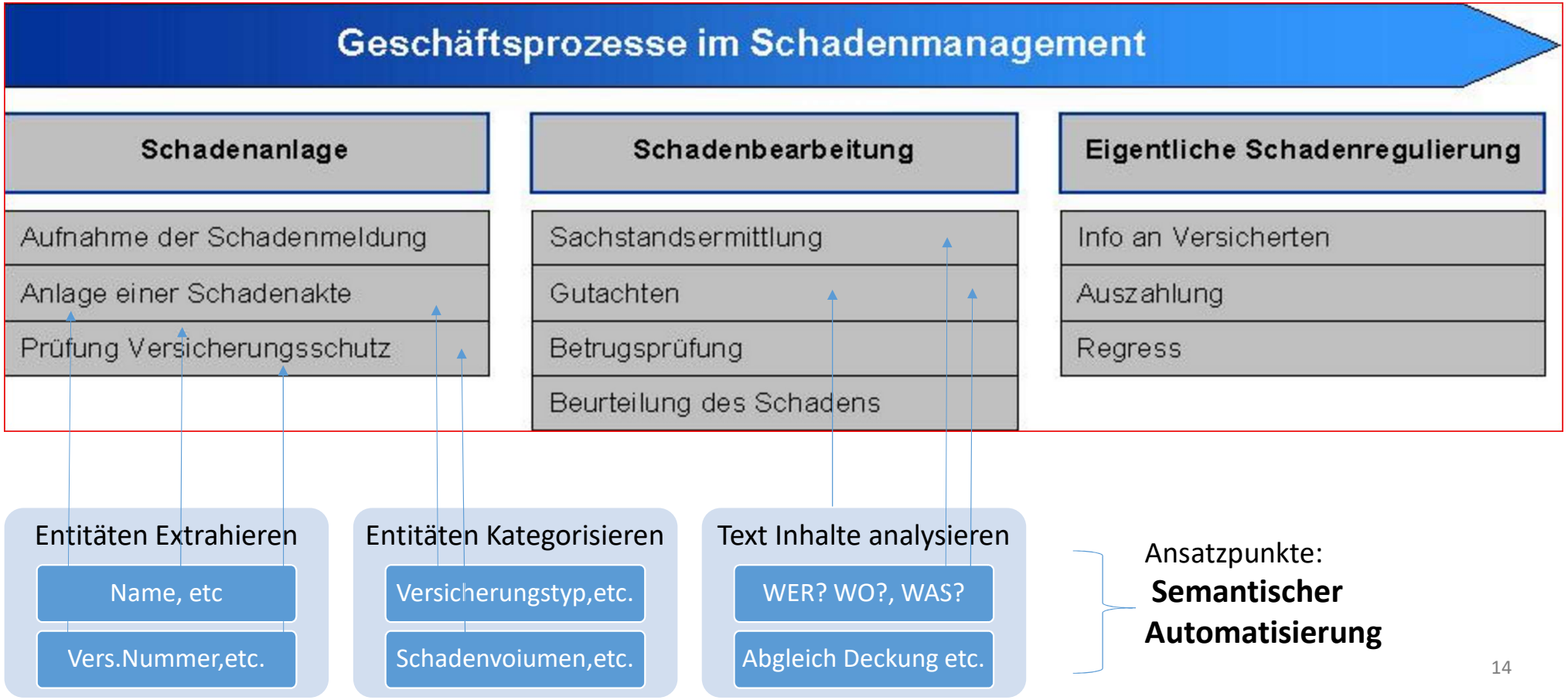
- Verdrängungswettbewerb initiiert u.a. durch Preiswettbewerb; Neue Marktteilnehmer: z.Bsp.Insurtecs
- Sinkende Kundenloyalität; Generation: Internet, Kundenzufriedenheit
- Neue Anforderungen an bedarfsspezifische Produkte und flexible Serviceprodukte
- Hoher Kostendruck (ca. 75% Schadenaufwände an den Gesamtkosten)
- Einhergehende sinkende Kapitalmarkterträge



Ansatz zur BusinessProzess Optimierung der Wertschöpfungskette:

Schadenmanagement

Predictive Analytics auf Basis unstrukturierter Daten



Ansatzpunkte:
Semantischer Automatisierung

Kommunikationsdokumente der Beteiligten

- **Schadensmeldung über verschiedene „Kanäle“**
  - App, Portal, e-mail, mail, Fax..
- Werkstattberichte, Rechnungen
- Arztberichte, Arztabrechnungen
- Gutachten
- Open Source Intelligence Daten (www, Geoinformationen,etc.)
- Partnerverträge, Policen
- …….

Daten und Informationen liegen **unstrukturiert** vor !!

Datenaufkommen verdoppelt sich ca. alle 2 Jahre !!

Kommunikationsdokument: **Schadenmeldung**

Kraftfahrzeug-Schadenanzeige

Versicherungsgesellschaft:
Capitol Versicherung

ART DER VERSICHERUNG

☒ Haftpflicht	☒ Kasko	☐ Insassenunfall	☐ Rechtsschutz	☐ Sonstige
Polizzennummer 3456	Polizzennummer 7890	Polizzennummer	Polizzennummer	Polizzennummer

VERSICHERUNGSNEHMER

Vor- und Zuname / Firma	Karl Mustermann
PLZ/Ort/Straße/Hausnummer	4567 Irgendwo Fritzstrasse 25
Telefonnummer	0180-345637000

VERSICHERTES FAHRZEUG

Art (PKW, LKW)	Marke, Type	Erstzulassung	Fahrgestell-Nr.	Kennzeichen
PKW	VW, Käfer	01.01.1999	VW1234567	K-YY 3456

LENKER

Vor- und Zuname	Frieda Mustermann
Geburtsdatum	31.12.1976
PLZ/Ort/Straße/Hausnummer	4567 Irgendwo Fritzstrasse 25
Telefonnummer	0181-5432100

Genaue Sachverhaltsdarstellung

Es war dunkel als der Fahrzeuglenker in die Platanenstrasse einbog. Plötzlich kam eine schwarze Katze von rechts auf die Fahrbahn in selbstmörderischer Absicht gelaufen. Diesem Katzenwunsch nicht zu entsprechen lenkte der Fahrer den Wagen nach links. Nach verlassen der Fahrbahn am linken Rand der Strasse durchbrach das Fahrzeug eine ca. einen halben Meter hohe Hecke. Hinter der Hecke befand sich ein Graben (50cm tief) der das Fahrzeug zum Stehen brachte. Die vorderer Stoßstange sowie der linke Kotflügel waren danach von der übrigen Karroserie getrennt. Das Fahrzeug konnte nicht mehr bewegt werden und die Katze war auch tot.

EXTRAKTIONSELEMENTE

Beispiel: ICD10 Code Analyse

- Die ICD dient weltweit zur Verschlüsselung von Diagnosen
- Viele etablierte Prozesse basieren auf der ICD-10-GM z.Bsp.:
  - **Abrechnung**, Qualitätssicherung, Statistik und Epidimiologie

Verletzungen der Schulter und des Oberarmes (S40-S49)

Inkl.: Verletzung:
- Achselhöhle
- Oberarm
- Schulter
- Skapularregion

Exkl.: Beidseitige Beteiligung von Schulter und Oberarm (T00-T07)
Erfrierungen (T33-T35)
Insektenbiss oder -stich, giftig (T63.4)
Verbrennungen und Verätzungen (T20-T32)
Verletzungen:
- Arm, Höhe nicht näher bezeichnet (T10-T11)
- Ellenbogen (S50-S59)

Erfassung des Sachverhaltes: ICD plus schriftliche Ergänzungen des Arztberichts

✓ Regelbasierte semantische Verfahren

✓ Prüfung auf Plausibilität (Differentialdiagnostik!)

✓ Ganzheitliche Kostenbetrachtung

hohes Datenpotential !!!

Wie werden aus unstrukturierten Daten strukturierte Informationen ?

- Mit Expert System Software: die Texte liest wie sie der Mensch versteht.
- Mit SemantischerDatenanalyse oder auch Textanalyse, kognitive Analyse, Machine Learning Funktione , Natural Language Processing (NLP)
  - Automatische Extraktion von Begriffen, Namen, Orten, Organisationen aus den Belegen
  - Ontologien, Thesauri, Taxonomien legen die Begriffe fest.
    - Medizinische, Technische, Geografische Inhalte, etc.

Daten und Informationen liegen danach **strukturiert** vor !!

Use case: Produktmanagement/Entwicklung

Produktmanagement und Semantik

- Hintergrund und Ausgangslage
- Was kann Semantik leisten ?
- Wie kann Semantik das Produktmanagement unterstützen ?
- Nutzen für den Versicherer ?

Predictive Analytics auf Basis unstrukturierter Daten

Adobe Systems

Mitte September 2013 wurde Adobe das Ziel von Hackern. Circa 38 Millionen Datensätze von Adobe-Kunden wurden im Zuge des Cyberangriffs gestohlen - darunter die Kreditkarteninformationen von knapp drei Millionen registrierter Kunden. Die Hacker die hinter dem Angriff standen, wurden nicht gefasst.

Target Corporation

Die Target Corporation gehört zu den größten Einzelhandels-Unternehmen der USA. Ende des Jahres 2013 musste Target einen Cyberangriff eingestehen, bei dem rund 70 Millionen Datensätze mit persönlichen Informationen der Kundschaft gestohlen wurden. Weitaus schwerer wog jedoch, dass unter diesen auch 40 Millionen Datensätze waren, die Kreditkarteninformationen und sogar die zugehörigen PIN-Codes enthielten. Für

Sony Playstation Network

Im April 2011 ging bei vielen Playstation-Besitzern rund um den Globus nichts mehr. Der Grund: ein Cyberangriff auf das digitale Serviceportal Playstation Network (PSN). Neben einer Ausfallzeit des PSN von knapp vier Wochen (!) wurden bei der Cyberattacke jedoch auch die Daten (Kreditkarteninformationen und persönliche Daten) von rund 77 Millionen PSN-Abonennten gestohlen.

Heartland Payment Systems

Noch heute gilt der 2008 erfolgte Cyberangriff auf das US-Unternehmen Heartland Payment Systems als einer der größten Hacks aller Zeiten wenn es um Kreditkartenbetrug geht. Heartland ist einer der weltweit größten Anbieter für elektronische Zahlungsabwicklung. Im Zuge des Hacks wurden rund 130.000.000 Kreditkarten-Informationen gestohlen. Der Schaden für Heartland belief sich auf mehr als 110 Millionen Dollar, die zum größten

Cyber Crime und Risiko:

Fakten:

- Lloyd's schätzt aktuell ca.400 Milliarden € jährliche Verluste aufgrund von Cyber Attacken und nur ein kleiner Teil davon ist versichert.
- Der jährliche "Verizon Data Breach Report" zählte für das Jahr 2014 weltweit 80.000 Cyberangriffe. Gegenüber dem Jahr 2013 bedeutete das einen Anstieg von 55 Prozent.

Cyber Risk Versicherungen:

- Risiken kennen, Risiken bewerten, Verträge gestalten
- Datenbasis ??? Meldepflicht?
- Neue Risiken schwer vorhersehbar.? (Telematik? IoT, etc)

Predictive Analytics auf Basis unstrukturierter Daten

Informationsbasis für Cyber Crime im www als unstrukturierte Daten Daten :

- Zugriff auf Daten über semantische Crawler (web pages, social data, documents, texts, email, pdf, etc.) from internal and external sources.
- Kategorisierung und Extraktion auf Cyber Risk Ontologien, Taxonomien
- Überführung der unstrukturierten Information in quantifizierbare Kategorien

Wissensgenerierung und Aufbau einer Datenbank als Grundlage für Risikoabschätzungen nach Objektekategorisierung:

- Welcher breach ? Welches Ziel ? Welcher Schaden ? Welchen Ursprung ?
- Risiken updaten und Produktmanagement in ihrem Angebot zu unterstützen.

Predictive Analytics auf Basis unstrukturierter Daten

Zusammenfassung:

- Unstrukturierten Daten und Wissenspotentiale nehmen dramatisch zu.
- Dieses Wissen kann/muß automatisiert extrahiert werden.
- Semantische Softwarelösungen sind der Schlüssel schnell und zielgerichtet Entscheidungsunterstützung zu bieten.
- Cognitive Computing optimiert Standardprozesse wenn Informationen als Texten, Dokumenten vorliegen und quantitative Information vorliegen sollen.
- Liefert ergänzendes Datenmaterial für den Predictive Analytics Ansatz
- Ergänzt mit extrahiertem Wissen gerade bei geringem Datenbestand die „Tabelleinformation “.

Was Expert System bereitstellt

Wir entwickeln und implementieren Software, die:

- Die Bedeutung von geschriebener Sprache versteht, so wie sie von Menschen gelesen und verstanden wird
- Automatisierte Verfahren zur Extraktion, Kategorisierung von Textinhalten zur Wissensbildung im Fachbereich unterstützt
- Bereitstellung facettierter Suchkriterien zur Optimierung von Suchmaschinen.
- Eine Reihe leistungsstarker Text Analysetools beinhaltet
- Bereitstellung von Standard Taxonomien zur Kategorisierung
- Kunden Thesauri integriert und den life-cycle der Thesauri unterstützt.
- Lösungen für Claim Management, CRM, Risk Management, etc. und Industrieübergreifend.

Predictive Analytics auf Basis unstrukturierter Daten

Expert System ?

Patentierter und mehrfach ausgezeichneter Technologieanbieter für Semantik

3 der weltweit führenden Öl- und Gas Unternehmen sind unsere Kunden

90% unserer Kunden setzen unsere Software in unterschiedlichen Anwendungen ein

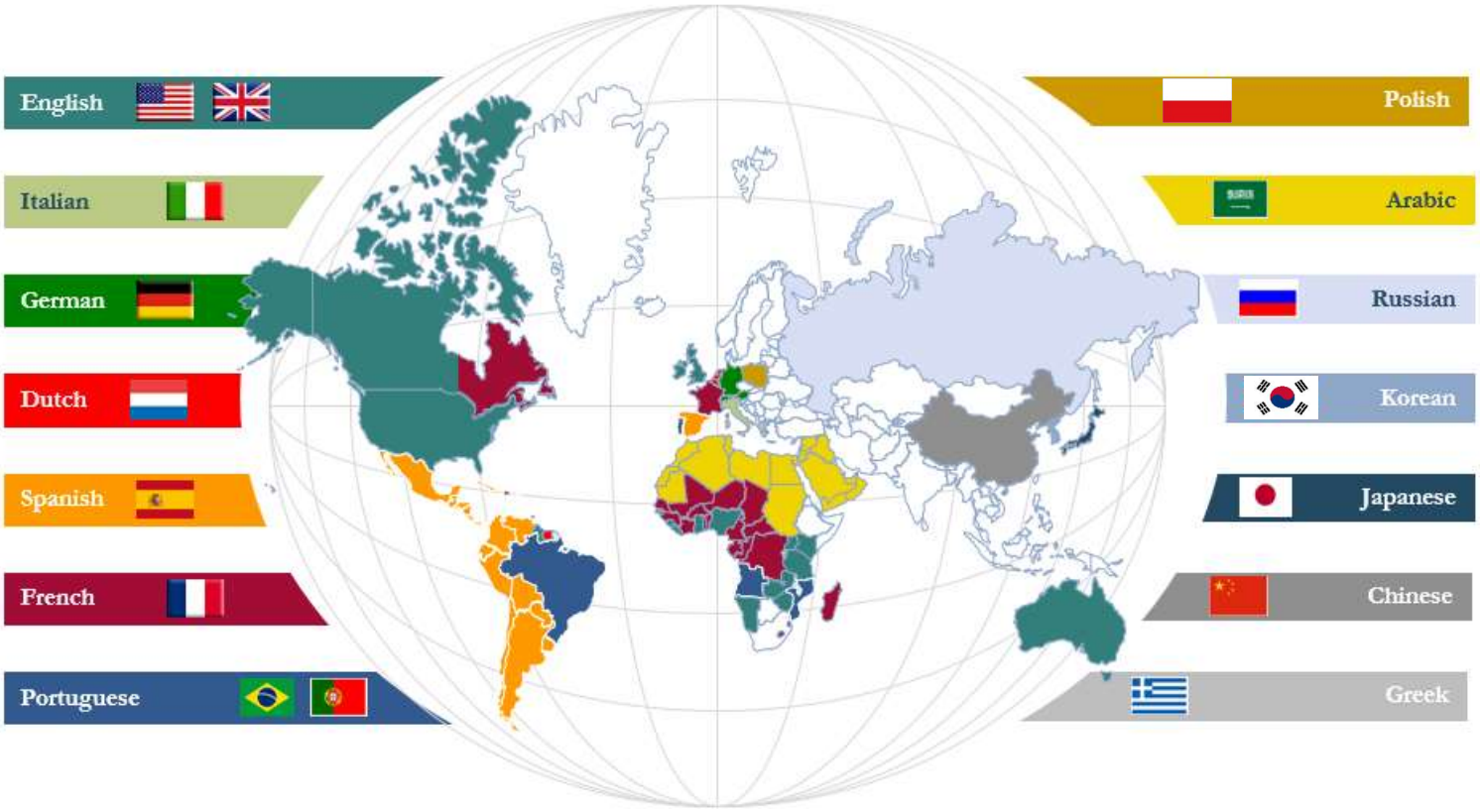
Weltweit präsent und an der Mailänder Börse gelistet

Führende US und Europäische Behörden arbeiten mit uns zusammen

weltweite Versicherungs- und Bankkunden benutzen unsere Software

Bewertete Technologie: FORRESTER® IDC Analyze the Future Gartner

Unterstützte Sprachen



Expert System unterstützt gegenwärtig 14 Sprachen.

Auszug aus der Referenzliste

EXPERT SYSTEM
SEMANTIC INTELLIGENCE

GOVERNMENT

Life Sciences

MEDIA & PUBLISHING

ENERGY, OIL & GAS

FINANCE

Danke !

jlackner@expertsystem.com