Missbrauchserkennung

mit p.REX.fd

Beschrieb von Vorgehen und Funktionen

Version 1.3

Juni 2006

1 ZWECK DIESES DOKUMENTS 2

2 DIE AUSGANGSLAGE UND AUFGABENSTELLUNG IN DER MISSBRAUCHSERKENNUNG 2

3 DER PROZESS DER MISSBRAUCHSERKENNUNG 4

4 DAS SYSTEM P.REX.FD ZUR MISSBRAUCHSERKENNUNG 5

4.1 DIE BASIS 5
4.2 DIE ERKENNUNGSMETHODEN 7
4.2.1 Die Textanalyse und das Namechecking 7
4.2.2 Die Mustererkennung 7
4.2.3 Die Linkanalyse 10
4.2.4 Die Erstellung von regelbasierten Filtern 11
4.3 DIE ORGANISATIONSLÖSUNG VON P.REX.FD 12
4.4 DER LAUFENDE LERNPROZESS IN DER ANWENDUNG VON P.REX.FD 13
4.5 DER EINSTIEG IN DIE MISSBRAUCHSERKENNUNG 14

5 DIE TECHNISCHE INTEGRATION 15

© Prospero AG, 2001-2006

1 Zweck dieses Dokuments

Die gezielte und intelligente Analyse von Daten ist ein wirksames Mittel zur Aufdeckung von Missbrauch in vielen Bereichen. Dieses Dokument gibt eine generelle Darstellung der Aufgabenstellung in der Missbrauchserkennung und beschreibt die Module und Funktionalitäten der Lösung p.REX.fd zum Aufbau eines umfassenden Schutzschilds zur Aufdeckung und Verhinderung von Missbrauch.

2 Die Ausgangslage und Aufgabenstellung in der Missbrauchserkennung

Generell hat Missbrauch immer mit Handlungen von Menschen zu tun. In der Informationsgesellschaft entstehen durch Handlungen von Menschen in vielen Fällen Daten, insbesondere bei der Ausübung von geschäftlichen Tätigkeiten. Missbrauch kann dabei von einer einzelnen Person oder von Gruppen von Personen betrieben werden. Die Beispiele sind vielzählig:

- Falschbuchungen von Rückgaben im Detailhandel
- Missbrauch von Kundeninformationen (Insiderhandel, Frontrunning)
- Unechte Rabattbuchungen in Reisebüros
- Doppelbuchungen von Lieferantenrechnungen
- Gefälschte Rechnungsstellungen im Einkauf
- Verrechnung von nicht erbrachten Serviceleistungen
- Missbrauch mit Kreditkarten
- Gefälschte Kundendaten und Identitäten
- Fälschung von Dokumenten
- Missbrauch von Serviceleistungen (z.B. Telefonservices)
- Etc.

Bei der Aufgabe der Aufdeckung von Missbrauch können grundsätzlich zwei Fälle unterschieden werden:

- Bekannte Missbrauchsszenarien

  Diese Szenarien kennzeichnen sich dadurch, dass die Abläufe des Missbrauchs bekannt sind und diese beschrieben werden können. Im Kreditkartengeschäft z.B. existieren durch die langjährigen Erfahrungen in der Missbrauchsbekämpfung eine Vielzahl von definierten

© Prospero AG, 2001-2006

Szenarien von Missbrauch. Zur Verhinderung dieser Missbräuche werden entsprechende Profile definiert, welche die Missbrauchsituationen in den Daten proaktiv erkennen. Ein einfaches Beispiel ist der Gebrauch der gleichen Kreditkarte innerhalb weniger Minuten an geografisch weit auseinander liegenden Automaten, der durch eine festgelegte Regel im System unterbunden wird.

Bekannte Missbrauchsszenarien können mit regelbasierten Filtern erkannt werden.

- Unbekannte Missbräuche

  Diese Situationen von Missbrauch sind ungleich schwieriger zu erkennen, da sie nicht vorab bekannt, sondern bislang unerkannt und verborgen sind. Die Anforderungen an ein leistungsstarkes Erkennungssystem gehen hier über ein regelbasiertes System hinaus. Das System muss Auffälligkeiten in Daten selbständig zuverlässig erkennen können. Die wesentlichen Herausforderungen sind:

  - Das selbständige Erkennen von generellen Auffälligkeiten und Ungewöhnlichkeiten ohne spezifizierte Anhaltspunkte;
  - Das laufende Schärfen der Erkennung aufgrund von Feedback;
  - Die adaptive Anpassung an sich verändernde Gegebenheiten (z.B. kann sich ein augenblicklich auffälliges, missbrauchverdächtiges Verhalten einer Person auf geänderte Lebensumstände zurückzuführen sein und Normalität werden).

  Pauschal formuliert muss bei der Erkennung von unbekanntem Missbrauch ein System in der Lage sein, über eine selbständig lernende Mustererkennung Missbräuche zuverlässig aufzudecken, ohne dabei viele Falscherkennungen zu generieren, d.h. Fälle zu erkennen, die gar keine Missbräuche darstellen.

Die Hauptaufgaben in der Missbrauchserkennung sind:

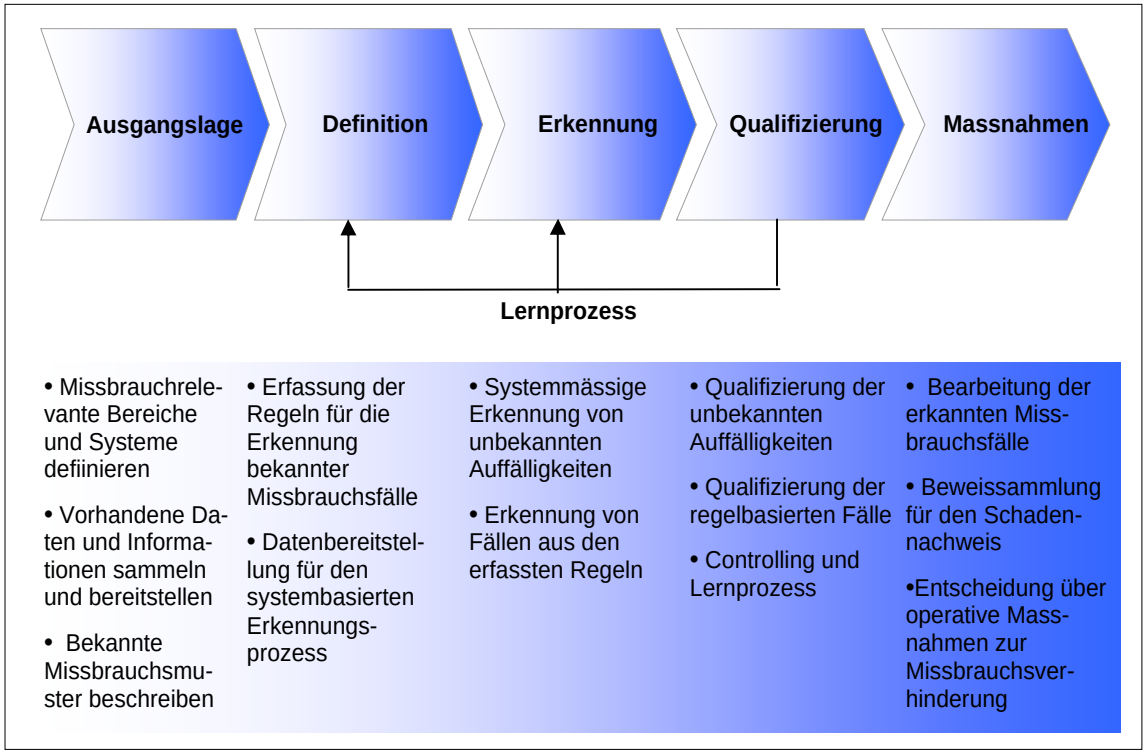
- Entwicklung von Algorithmen zur Suche, Auswertung und Speicherung von missbrauchsverdächtigen Daten;
- Identifikation von bisher unbekannten Zusammenhängen und Mustern;
- Erstellen und Beurteilen von Missbrauchsprofilen (ungewöhnliches Verhalten);
- Nutzung von bekannten Zusammenhängen für Vorhersagen;
- Feststellung von Vorgehensweisen bei erkannten Fällen.

Im Folgenden wird der Prozess der Missbrauchserkennung dargestellt und anschliessend die wesentlichen Ausprägungen und Funktionsweisen eines leistungsstarken Missbrauchserkennungssystem aufgezeigt.

© Prospero AG, 2001-2006

3 Der Prozess der Missbrauchserkennung

Grafisch kann der Prozess der Missbrauchserkennung wie folgt dargestellt werden:



Ausgangslage

In der Ausgangslage werden die missbrauchsrelevanten Bereiche und Systeme im Unternehmen identifiziert, die dort vorhandenen Daten und Informationen zusammengestellt und bereits bekannte Missbrauchsmuster und Fälle beschrieben.

Definition

In der Definitionsphase werden im System die bekannten Missbrauchsfälle mit der Definition von Regeln abgebildet und die Daten für den systembasierten Erkennungsprozess bereitgestellt.

Erkennung

Der Erkennungsprozess hat zwei Ziele:

- Das System findet unbekannte, auffällige Datenkonstellationen.
- Das System wendet die definierten Regeln zur Erkennung von bekannten Auffälligkeiten an und zeigt die erkannten Fälle.

Qualifizierung

In der Qualifizierungsphase werden die Fälle analysiert:

© Prospero AG, 2001-2006

- Die vom System gefundenen unbekannten Auffälligkeiten werden nach der Frage Missbrauchsfälle ja/nein qualifiziert und unterschieden. Diese Unterscheidung ist die Voraussetzung für den Lernprozess des Systems, dass die nicht missbrauchsrelevanten Auffälligkeiten in Zukunft nicht mehr gemeldet werden.
- Durch die Qualifizierung der durch die definierten Regeln erkannten Fälle können die Regeln angepasst und geschärft werden.

Massnahmen

Aus den Ergebnissen der Qualifizierungsphase werden erkannte Missbrauchsfälle weiter analysiert und Massnahmen zur präventiven Verhinderung definiert.

4 Das System p.REX.fd zur Missbrauchserkennung

4.1 Die Basis

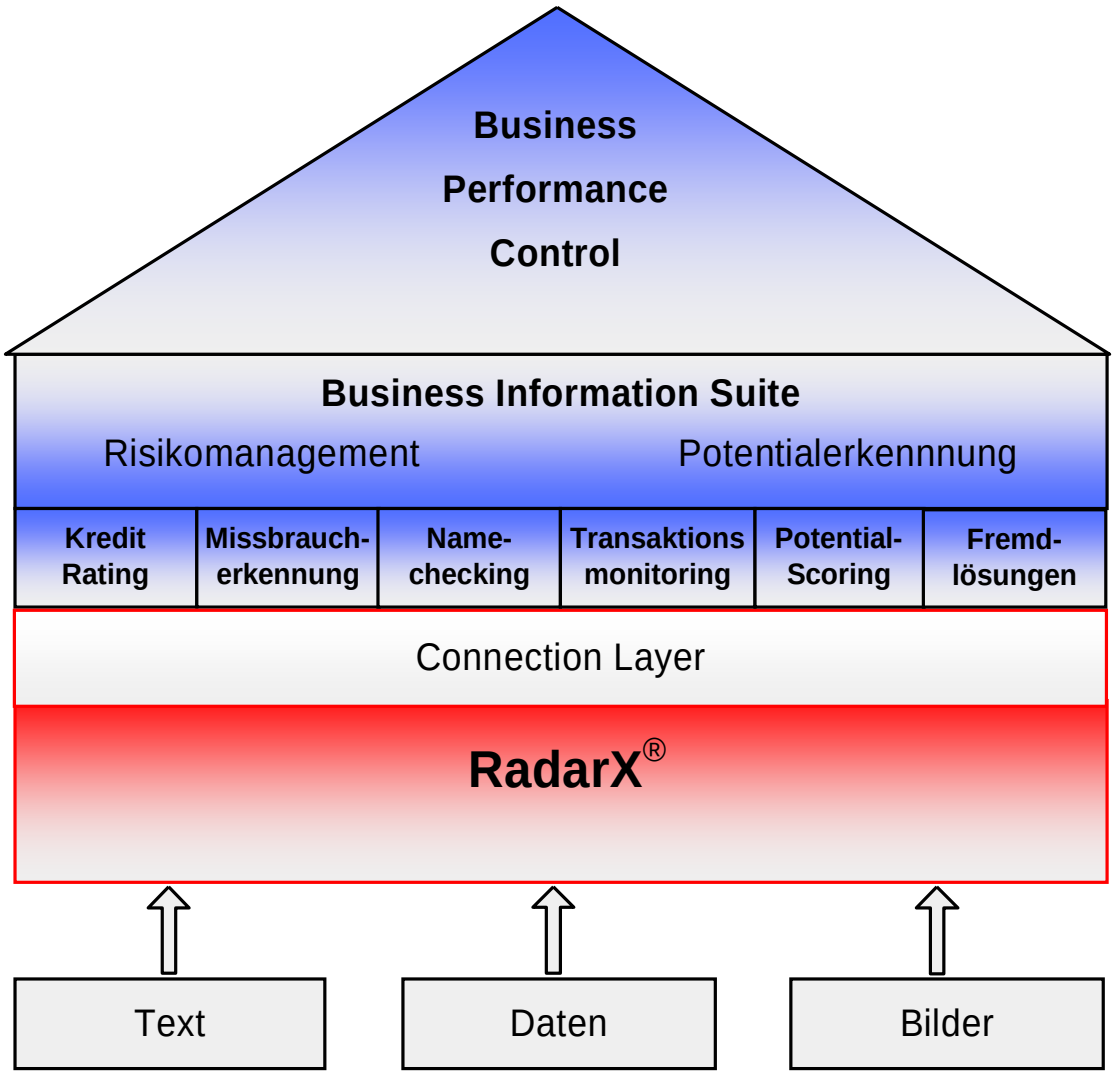
Die analytische Aufgabe in der Missbrauchserkennung ist generell:

Es geht darum, aus vorhandenen Informationen die relevanten, d.h. die Missbrauchsfälle zu erkennen. Die Rahmenbedingungen sind hohe Komplexität, verborgene Zusammenhänge und unbekannte Verhaltensmuster.

Für diese Aufgabe stossen die traditionellen betriebswirtschaftlichen Methoden wie z.B. statistische Verfahren an ihre Grenzen, die Lösung wurde durch die Übertragung von Wissen über das Verhalten von natürlichen Systemen gefunden: Ein Grundprinzip der Natur ist die Mustererkennung, mit ihr werden komplexe Tatbestände und Zusammenhänge umfassend wahrgenommen. Im Modul RadarX® wurden diese Prinzipien in einen systembasierten, autonomen Prozess der Erkenntnisgewinnung umgesetzt. Benchmarktests aus unterschiedlichsten Bereichen haben gezeigt, dass mit dem Cognitive Decisioning® signifkant bessere Erkennungsraten erzielt werden als mit anderen Verfahren – unabhängig davon, ob die Informationen in Daten, Text oder Bildern vorliegen. Dies ist das entscheidende Merkmal eines leistungsstarken Systems zur Missbrauchserkennung: die Maximierung der Erkennung der relevanten Missbrauchsfälle bei gleichzeitiger Minimierung der Falscherkennungen. In der Praxis werden mit dem Cognitive Decisioning® die Erkennungsraten in der Regel um ein Mehrfaches gesteigert – mit direkt positivem Einfluss auf die Prozessicherheit und die Kosten.

© Prospero AG, 2001-2006

p.REX.fd ist ein Modul aus dem Prospero Knowledge Haus:



Das Prospero Knowledge Haus ist eine Suite von Lösungen zur Quantifizierung von Chancen und Risiken in Businessprozessen. Das Herzstück ist der RadarX®, welcher aus unterschiedlichen Informationsquellen die relevanten Zusammenhänge und Erkenntnisse findet, wie z.B. für die Erkennung von Missbrauch. Mit dem Bio-Intelligence® - Verfahren ist dieser Prozess weitgehend automatisiert. Der Anwender erhält als Resultate direkt anwendbare Scores für Chancen und Risiken, z.B. die Wahrscheinlichkeit, dass eine Kassentransaktion im Detailhandel auf Missbrauch beruht. Der RadarX® beinhaltet eine Reihe von Modulen, mit welchen die individuellen Ausgangslagen von Informationen und Prozessen umfassend abgedeckt werden. Diese Erkennungsmethoden und Module werden im Folgenden dargestellt.

© Prospero AG, 2001-2006

4.2 Die Erkennungsmethoden

4.2.1 Die Textanalyse und das Namechecking

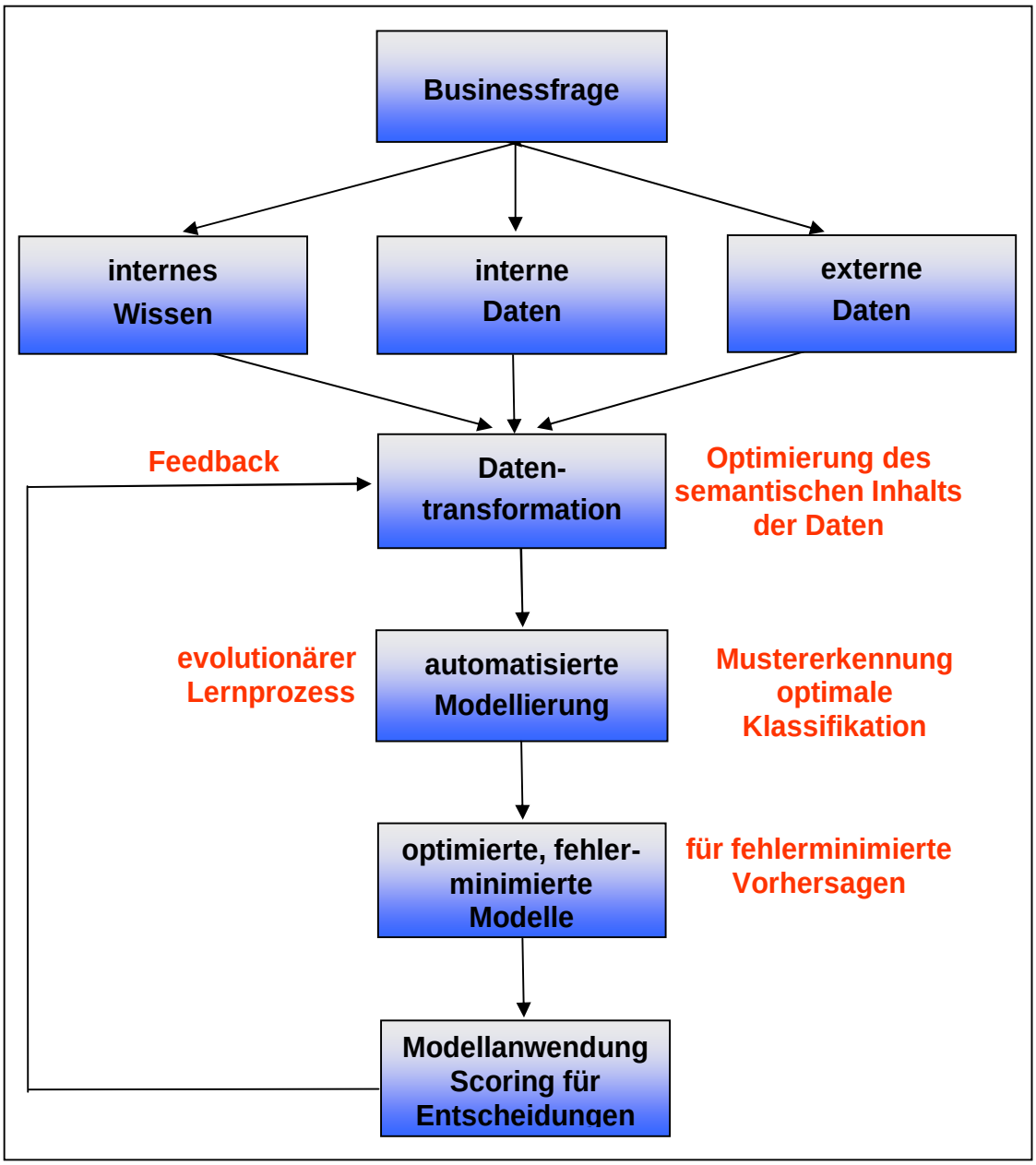
Das Modul Textanalyse erkennt Ähnlichkeiten und Muster in Informationen, welche in Form von Text vorliegen. Damit können z.B. gezielte Informationen oder zusammenhängende Dokumente gefunden werden. Speziell für die Missbrauchserkennung dient das Modul Namechecking, mit welchem Personendaten in beliebigen Kontexten geprüft und analysiert werden können.

4.2.2 Die Mustererkennung

Die Mustererkennung ist die Basis der Kognition und Wahrnehmung. Mit dieser umfassenden und ganzheitlichen Betrachtungsmethode können in komplexen Situationen unbekanntes Verhalten, verdeckte Zusammenhänge und verborgene Abhängigkeiten aufgedeckt werden. Ein Kernmodul des RadarX® ist der p.Modeler, welcher automatisiert und selbständig die relevanten Muster aus Daten erkennt.

Das generelle Vorgehen zeigt die folgende Grafik:

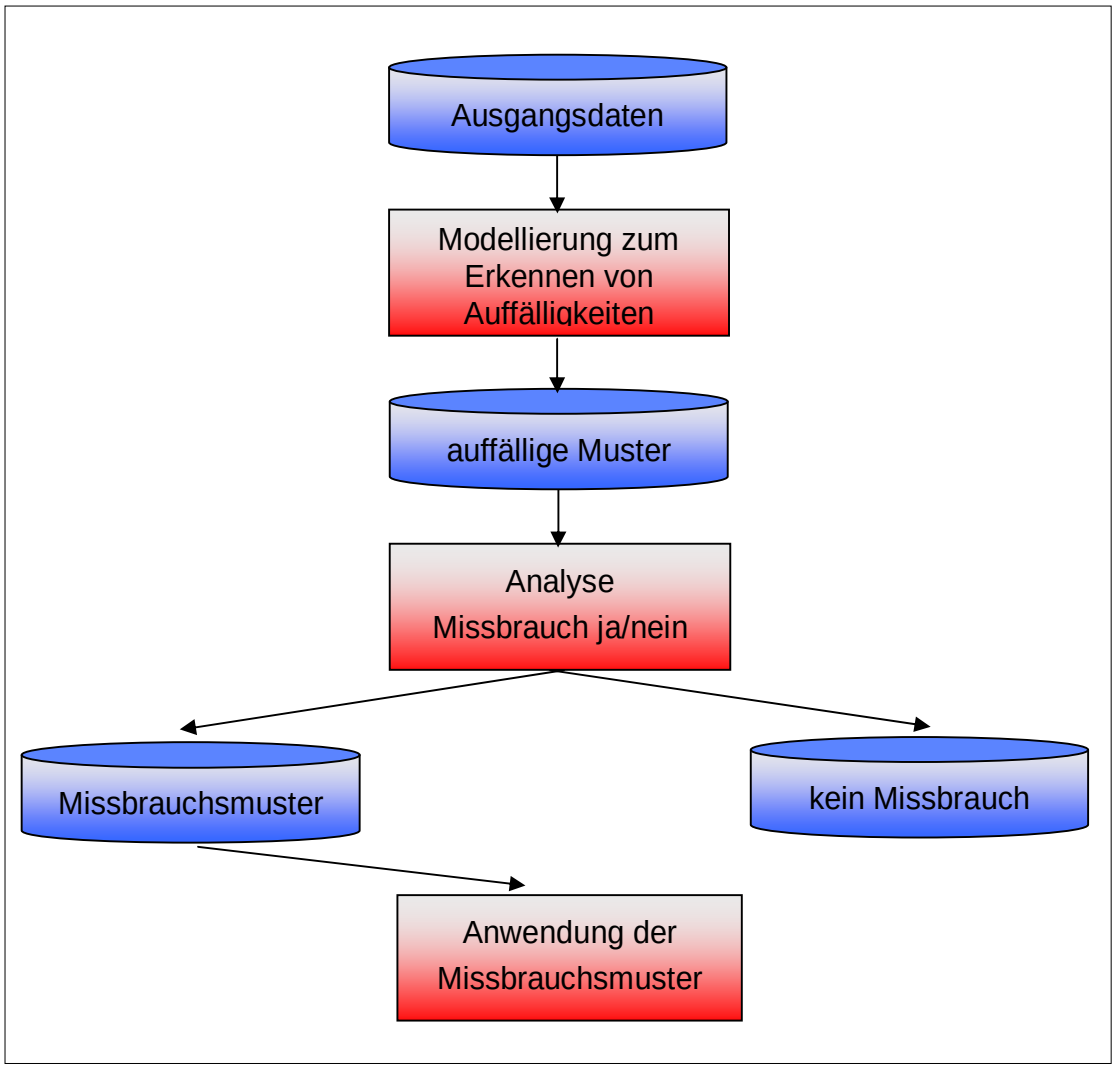
© Prospero AG, 2001-2006



Ausgangspunkt ist die Businessfrage, z.B. die Erkennung von Missbrauch. Danach werden die verfügbaren Informationen zur Aufdeckung von Missbrauch festgelegt. Die operativen Daten aus den zu analysierenden Prozessen werden dabei fallweise mit dem internen Wissen der Mitarbeiter oder externen Daten (z.B. Branchenkennzahlen, demografische Daten etc.) ergänzt. Die Daten werden bereitgestellt und dem Modeler übergeben. Dieser findet in einem automatisierten, evolutionären Lernprozess die aus der Vielzahl der möglichen Modelle (bei 30 Inputfaktoren sind ca. 536 Mio. unterschiedliche Modelle möglich) diejenigen, welche die missbrauchsverdächtigen Fälle am zuverlässigsten erkennen können.

© Prospero AG, 2001-2006

Wird aufgrund von fehlenden Anhaltspunkten über möglichen Missbrauch in einer ersten Phase nach generellen, unspezifizierten Auffälligkeiten gesucht, so sind die vom System gefundenen Resultate nach der Frage der Missbrauchsrelevanz zu analysieren:

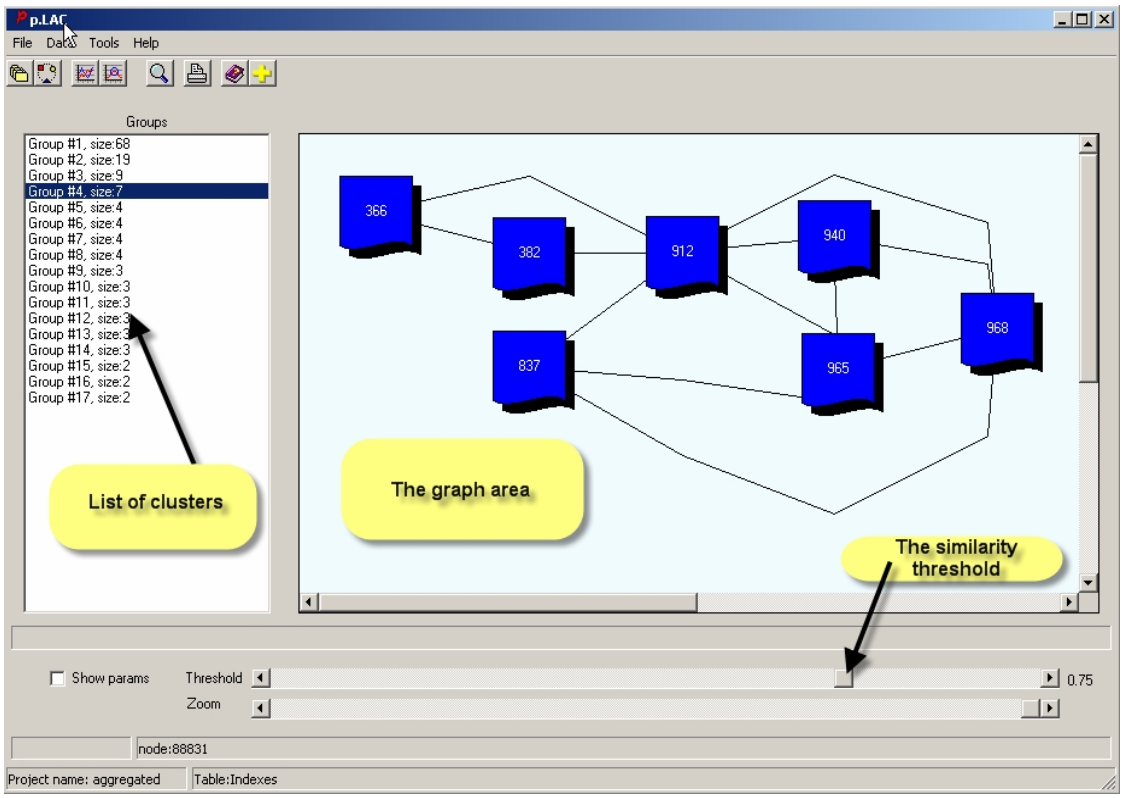


Dieser Ansatz ist in der Missbrauchserkennung speziell wichtig, da es wesentlich um die Aufdeckung von unbekanntem, bislang verborgenem Verhalten geht.

© Prospero AG, 2001-2006

4.2.3 Die Linkanalyse

Eine spezielle Herausforderung in der Missbrauchserkennung ergibt sich, wenn einzelne Objekte (z.B. Transaktionen) für sich eigentlich unauffällig sind, im Zusammenhang betrachtet jedoch die Missbrauchssituation erkennbar wird. Mit dem Modul Linkanalyse werden auffällige, zusammenhängende Konstellationen in Daten erkannt.

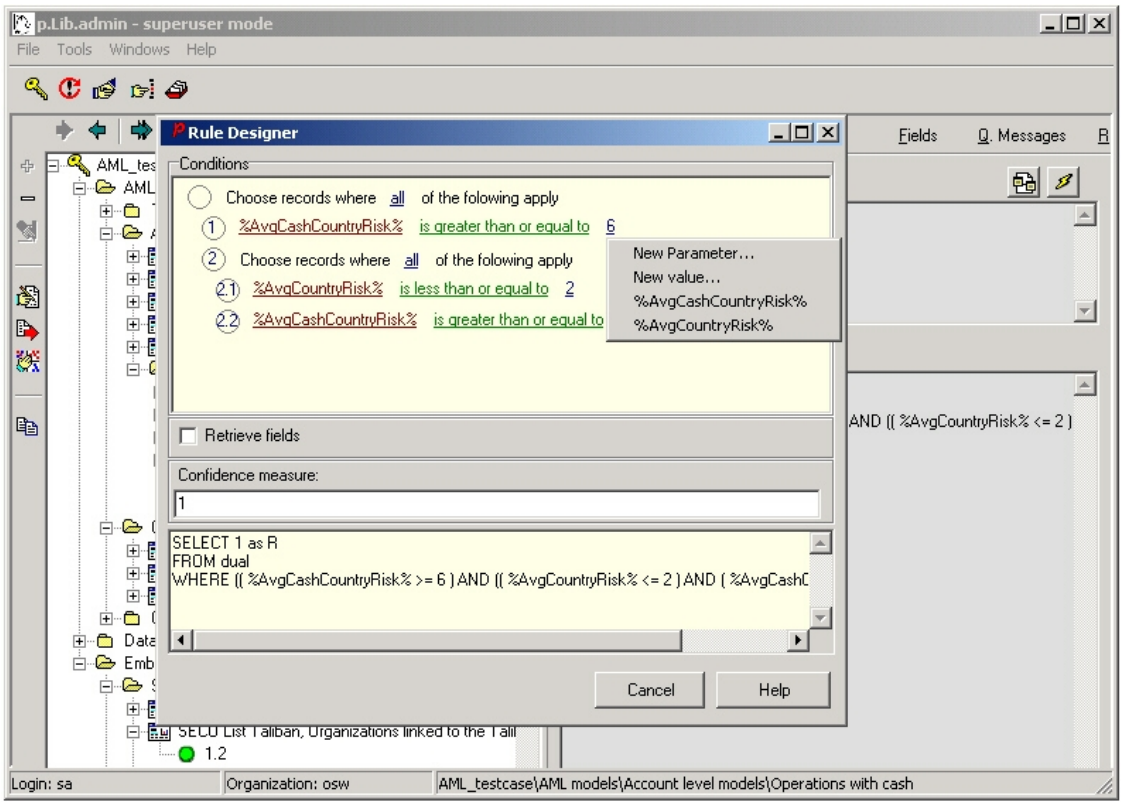


Das obige Bild zeigt in der Übersicht anonymisiert die Resultate einer Linkanalyse von Zahlungstransaktionen. Das System hat auffällige, zusammenhängende Transaktionen erkannt und diese können jetzt einfach und schnell auf die Frage von Missbrauch analysiert werden.

© Prospero AG, 2001-2006

4.2.4 Die Erstellung von regelbasierten Filtern

Diese Erkennungsmethodik deckt den einfachsten Fall der Missbrauchserkennung ab. Mit regelbasierten Filtern werden bereits bekannte Missbrauchsmuster abgebildet.

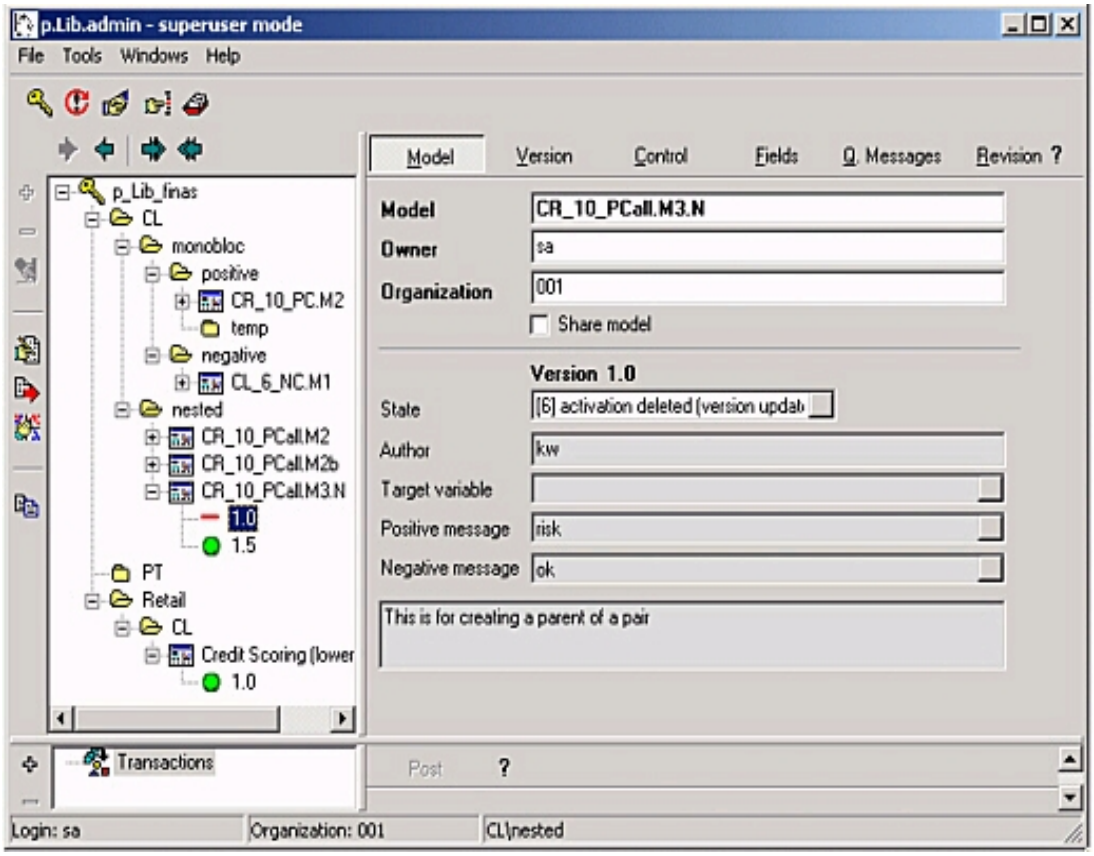


Je nach Prozesssituation werden die Missbrauchsfälle dann präventiv verhindert oder nach ihrer Entstehung sofort aufgedeckt.

Mit dem p.RuleDesigner können sowohl einfache, schwellwertbasierte Regeln wie auch komplexe Regelwerke mit Abhängigkeiten von sich verändernden Variablen und/oder Tabelleninhalten in einer einheitlichen Oberfläche anwenderfreundlich definiert und übersichtlich verwaltet werden. In den Regelwerken werden bekannte Zusammenhänge abgebildet. Diese können in der Anwendung kombiniert mit den fehlerminimierten Modellen zur Erkennung von unbekannten Mustern und Zusammenhängen eingesetzt werden. Damit kann eine auf die jeweilige individuelle Situation abgestimmte, mehrstufige Risikostrategie zur Missbrauchserkennung realisiert werden.

© Prospero AG, 2001-2006

4.3 Die Organisationslösung von p.REX.fd

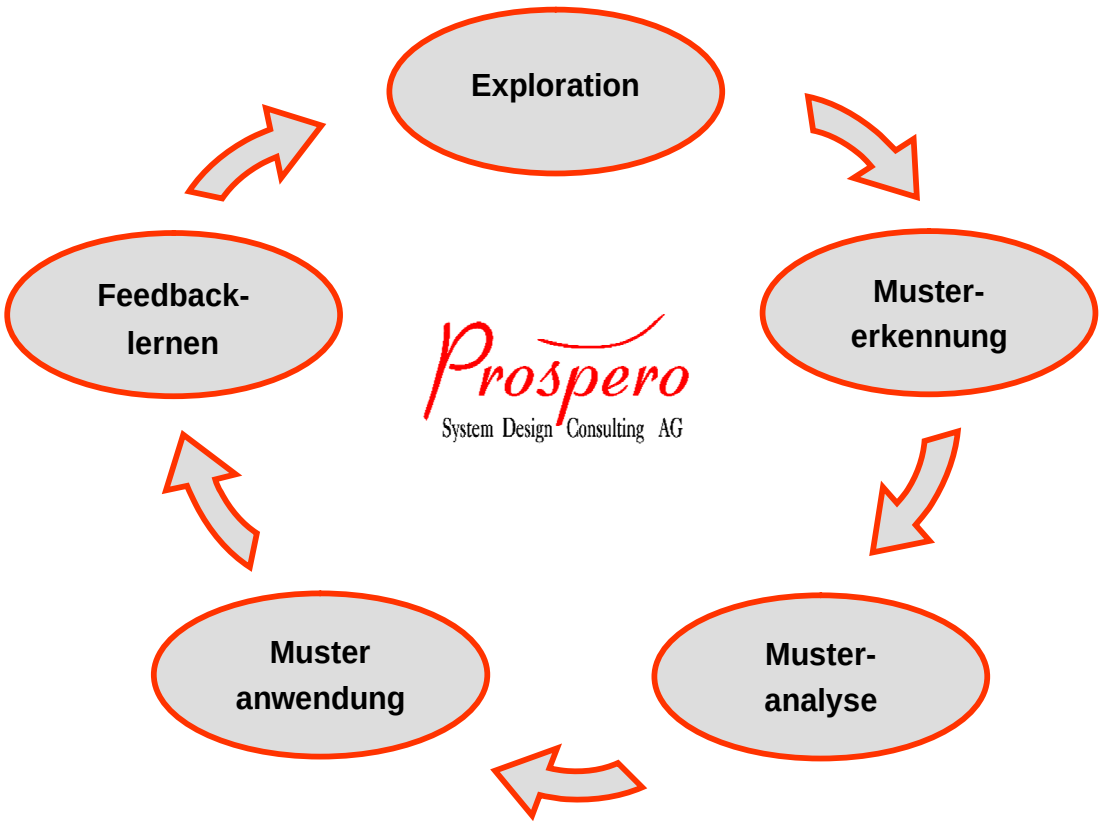


Eine leistungsstarke Lösung für die Missbrauchserkennung erfordert fallspezifisch den kombinierten Einsatz der in 4.2. dargestellten Erkennungsmethoden. Für die praktische Anwendung und den Nutzen ist deshalb neben der analytischen Leistungsstärke zum Finden der verlässlichen, relevanten Entscheidungsinformationen (zuverlässiges Scoring mit minimierten Fehlern, Erkennung der relevanten Faktoren, deren Wichtigkeit und Abhängigkeiten, Adaptionsfähigkeit auf sich verändernde Bedingungen etc.) die organisatorische Seite der Lösung entscheidend. p.REX.fd beinhaltet eine komplette Organisationslösung für die Abbildung aller Prozesse und Workflows, die für die Analyse von Daten und das Erstellen und Einsetzen von Modellen und Filtern erforderlich sind. Der Leitgedanke bei deren Realisierung war, den analytischen Prozess der Missbrauchserkennung analog wie andere Businessprozesse (z.B. ein Einkaufsprozess in der Beschaffung, ein Produktionsprozess oder ein Zahlungsablauf in der Buchhaltung) zu betrachten und in einer zentralen, bedienerfreundlichen Anwendung abzubilden. Unter vielen anderen Aspekten ist damit der Prozess der Missbrauchserkennung komplett nachvollziehbar, d.h. auditsicher. Dies gilt für die Modellbildung, -verwaltung und –anwendung gleichermassen. Organisatorisch ist die Missbrauchserkennung somit in den Businessprozess integriert. Neben dem betriebswirtschaftlichen Gesichtspunkt einer effizienten Handhabung und Pflege bekommt dieser Aspekt eine spezielle Wichtigkeit in allen Gebieten, in welchen die

© Prospero AG, 2001-2006

Gesetzgebung die Nachvollziehbarkeit fordert. Beispiele sind Basel 2 oder die Gesetze zur Geldwäschebekämpfung.

4.4 Der laufende Lernprozess in der Anwendung von p.REX.fd



In der Anwendung ist die Mustererkennung mit p.REX.fd ein laufender Lernprozess. Die musterbasierten Modell und die variablen Filter schärfen sich in der Anwendung in einem automatisierten Prozess. Diese Adaptionsfähigkeit des Systems ist eine wesentliche Voraussetzung für ein dauerhaft leistungsstarkes System, das sich ändernden Gegebenheiten anpasst.

© Prospero AG, 2001-2006

4.5 Der Einstieg in die Missbrauchserkennung

In der Praxis stehen beim Einstieg in das Thema Missbrauchserkennung oft Fragen des Ansatzpunktes und der verfügbaren Daten. Typische Ausgangsszenarien sind:

- Grosse Datenmengen mit umfangreichen Informationen, keine oder nur einzelne bekannte Missbrauchsszenarien und vage Vermutungen. Ziel: Klärung der vagen Vermutungen und Aufdecken von unbekannten Auffälligkeiten zur Missbrauchsanalyse, systematische Früherkennung der bekannten Missbrauchsszenarien. Zusätzlich soll erkannt werden, welche Daten wie relevant sind für die Missbrauchserkennung. Damit wird ein Prozess der laufenden Effektivitätssteigerung des Datenmanagements unterstützt.
- Grosse Datenmengen mit wenig Informationen pro Datensatz, keine oder nur einzelne bekannte Missbrauchsszenarien und vage Vermutungen. Ziel: Klärung des semantischen Gehalts der verfügbaren Informationen für die Erkennung der vagen Vermutungen und das Aufdecken von unbekannten Missbrauchsfällen, systematische Früherkennung der bekannten Missbrauchsszenarien. Zusätzlich sollen Hinweise für die Anreicherung der Daten zur Verbesserung der Ausgangslage gefunden werden.
- Wenig Daten mit umfangreichen Informationen pro Datensatz, keine oder nur einzelne bekannte Missbrauchsszenarien und vage Vermutungen. Ziel: Klärung, ob der Datenumfang ausreicht für eine zuverlässige Erkennung der vagen Vermutungen und das Aufdecken von unbekannten Missbrauchsfällen und für eine systematische Früherkennung der bekannten Missbrauchsszenarien.
- Wenig Daten mit wenig Informationen pro Datensatz, keine oder nur einzelne bekannte Missbrauchsszenarien und vage Vermutungen. Ziel: Klärung, ob die vorhandenen Daten ausreichen für eine zuverlässige Erkennung der vagen Vermutungen und das Aufdecken von unbekannten Missbrauchsfällen und für eine systematische Früherkennung der bekannten Missbrauchsszenarien.

Diese und andere Ausgangslagen der Missbrauchserkennung werden mit p.REX.fd in Verbindung mit dem im Kapitel 3 dargestellten Prozess der Missbrauchserkennung in einem proof of value – Ansatz von Prospero effizient und kostengünstig geklärt. Mit dem RAM – Ansatz (rapid modeling) steht ein Verfahren zur Verfügung, die Ausgangslage effizient zu qualifizieren.

© Prospero AG, 2001-2006

5 Die technische Integration

Die Gesamtarchitektur der Prosperolösungen ist für eine einfache Integration in bestehende Systemumgebungen ausgerichtet. Einige Hauptmerkmale dieser generischen Schnittstellenarchitektur sind, dass

- die proprietäre Datenhaltung in allen Systemen erfolgen kann, für die ODBC-Treiber oder eine ADO-Schnittstelle (Microsoft Standard) existieren, das gleiche gilt für den I/O von Betriebs- und Ergebnisdaten;
- die Datenhaltung, Berechnung und Anwendung von Scoringmodellen auf physikalisch getrennten Maschinen (Umgebungen) erfolgen kann;
- die Integration in ein DWH über die Möglichkeit der Einplanung als DB-Jobs erfolgen kann;
- die Programmteile via COM/COM+ und z.T. auch über CORBA angesprochen werden können.

Diese Architektur stellt sicher, dass Implementierungsprojekte auch in komplexen Umgebungen schlank und ohne grosse, ev. unbekannte Integrationsrisiken realisiert werden.

Weitere Informationen:
System Design Consulting Prospero AG
Gewerbeweg 15
FL – 9490 Vaduz
info@prospero.ch

Ihr Ansprechpartner:

Christian Schaefle

Tel: +41 79 431 82 32

Mail: c.schaefle@prospero.ch

© Prospero AG, 2001-2006.
Prospero und RadarX®, Bio-Intelligence®, p..REX.fd, p.CORD, p.SECCO etc. sind Trademarks der System Design Consulting Prospero AG. Andere Produkte oder Unternehmen, auf die hier referenziert wurde, sind Trademarks ihrer jeweiligen Eigentümer. Details dieser Beschreibung und der genannten Produkte können sich ohne besonderen Hinweis ändern.