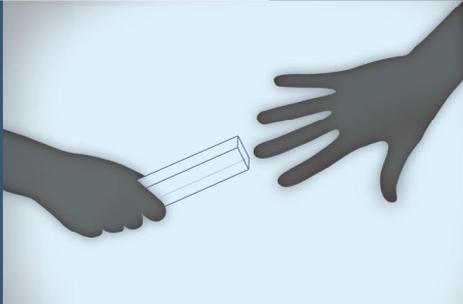


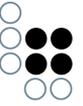
add meaning to your data



## White Paper K-Infinity

**intelligent views Wissenstechnologie:  
Wissenssteuerung durch Wissensnetze**





## 1 Einleitung

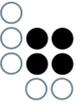
Information spielt heutzutage eine wichtigere Rolle als jemals zuvor. Gleichwohl erhält Information ihren entscheidenden Wert erst durch ein intelligentes Informationsmanagement. Informationen werden erst dann wirklich zugänglich, wenn sie organisiert, strukturiert, abrufbar und navigierbar sind und damit zu Wissen werden.

Ein Wissenspotenzial ist erst richtig nutzbar, wenn die einzelnen Informationen sinnvoll miteinander verknüpft sind und dieses Netz, das durch die Verknüpfungen entsteht, auch dargestellt wird. Die nachfolgenden Ausführungen sollen verdeutlichen, warum **intelligent views** in der faszinierenden Technologie der Wissensnetze einen effizienten Zugang zu einem gegebenen Wissenspotential sieht und wie sich diese Erkenntnis in der Produktpalette niederschlägt.

Wissensnetze bieten einen großen Nutzen, sowohl für den Informationskonsumenten als auch für den Informationsanbieter. Dieser Nutzen wird in Kapitel 3 und 4 verdeutlicht. Kapitel 5 fasst die Stärken der Philosophie zusammen, die den Wissensnetzen und ihrer Modellierung zugrunde liegt. Die sich anschließenden drei Kapitel erläutern zunächst allgemein die Anforderungen, die an Werkzeuge zum Aufbau und zur Nutzung eines Wissensnetzes zu richten sind, um dann ihre Umsetzung in die Produktpalette **K-Infinity** von **intelligent views** vorzustellen. Ergänzt wird diese Übersicht durch einige Bemerkungen zur Konfiguration dieser Werkzeuge. Schließlich findet man am Ende noch ein Glossar mit wichtigen Begriffen rund um Wissensnetze und Wissensmanagement.

## 2 Wissensnetze

Die Wissenstechnologie von **intelligent views** baut die Strukturierung des Wissens durch Wissensnetze auf. Wissensnetze versetzen **Informationsanbieter** in die Lage, ihren Kunden Informationsdienste und Publikationen einer neuen Qualität anbieten zu können. Informationsanbieter können so ihr Know-How optimal nutzen und anreichern. Unsere Technologie ist so angelegt, dass mit Hilfe von Wissensnetzen beliebige Inhalte wie Verlagsinhalte, Unternehmenswissen, Kundenwissen etc. indexiert werden können. Ein Wissensnetz kann für den **Informationsnutzer** verstanden werden als eine äußerst effiziente, redaktionell gepflegte Zugriffsstruktur. Im Unterschied zu anderen Ansätzen des Wissensmanagements ist dabei der Index vollkommen von den Inhalten entkoppelt und stellt einen Wert an sich dar.



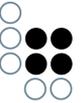
### 3 Nutzen für Informations- suchende

Die wichtigsten Fähigkeiten einer so entstehenden, neuen Generation wissensbasierter Informationssysteme sind für Informationssucher:

- Semantische Suche
- Individuelle Sichten
- Neue Interaktionsformen

Die **Semantische Suche** ermöglicht es dem Nutzer, thematisch relevante Materialien zu finden, auch wenn die eingegebenen Suchbegriffe in diesen Materialien gar nicht vorkommen. Ein Beispiel: Ein Nutzer sucht nach Informationen über Frauen in leitenden Positionen in der deutschen Wirtschaft. Relevante Artikel werden typischerweise Formulierungen enthalten wie „Marianne Mustermann, Geschäftsführerin der Düsseldorfer XYZ GmbH ...“ – Formulierungen, in denen also weder die Wörter „Frau“ noch „Position“ noch „deutsch“ noch „Wirtschaft“ direkt vorkommen. Durch die **Normierung** des Vokabulars in einem Wissensnetz und die Möglichkeit, Unter-, Oberbegriffe und andere semantische Beziehungen in die Suche einzubeziehen, können jedoch trotzdem die relevanten Artikel unabhängig von der Formulierung, ja sogar unabhängig von der verwendeten Sprache gefunden werden. Ein so gearteter Suchprozess kann durchaus mit der menschlichen Fähigkeit verglichen werden, die Ähnlichkeit unterschiedlicher Dinge, in diesem Falle Dokumente, durch assoziative Verknüpfung zu erkennen. Diese Fähigkeit wird im Wissensnetz durch die Beziehungen der Informationen untereinander verankert. Dies geht weit über die statistischen Ähnlichkeitsmodelle bestehender Wissensmanagementsysteme hinaus und bietet die Möglichkeit, mit Hilfe differenzierten Wissens Ähnlichkeit flexibel zu definieren.

**Individuelle Sichten und neue Interaktionsformen** machen den Nutzer unabhängig von fest vordefinierten Zugriffsstrukturen. Je nach Interesse können die Materialien thematisch, chronologisch, geographisch usw. gruppiert und geordnet werden – prinzipiell können alle Eigenschaften der Inhalte als Ordnungsdimension eingesetzt werden. Ein derart multidimensionaler Zugang umfasst auch eine explorative Suche, in der das Suchziel des Benutzers zunächst nicht klar sein muss, sondern sich erst im Laufe der Interaktion durch die gewonnenen Informationen verdichten kann. Typischerweise ist die explorative Suche in Analysesituationen relevant. Basierend auf differenzierten Wissensstrukturen lassen sich auch vollkommen neuartige, sehr effiziente Interaktionsmodelle



ableiten, die automatische Visualisierungs- und Layoutverfahren verwenden. Diese Verfahren garantieren, dass sich der Wissensnutzer auf seine eigentliche Tätigkeit konzentrieren kann und nicht durch unnötige Nebenarbeiten, wie etwa Formatierung, behindert wird.

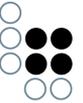
Nehmen wir z. B. an, in einem zehnbändigen populärwissenschaftlichen Nachschlagewerk zur Technik seien die Einträge nach Sachgebieten geordnet, also Drucktechnik unter Medien, Sehhilfen unter Optik etc. Ein alphabetischer Zugriff ist, wenn auch weniger bequem, über das Register möglich. Interessiert sich der Leser allerdings für die historische Dimension, beispielsweise den Zeitpunkt der Erfindung einer bestimmten Technik, so wird diese Information tief im Inneren der einzelnen Artikel versteckt und über alle zehn Bände verstreut sein. Eine wissensbasierte Erschließung liefert dem Leser hingegen auf seine Frage nach „Erfindungen der Renaissance“ den Buchdruck, die Brille, die Kompassnadel und die Gießereitechnik als Antwort. Zusätzlich wird ihm auf Wunsch die Möglichkeit geboten, die Geschichte dieser Erfindungen mittels einer grafischinteraktiven Zeitleiste chronologisch zu erschließen. Gerade innovative, grafische Interaktionsformen profitieren also von der Fülle der im Wissensnetz modellierten Informationen sowie entsprechenden Technologien zur automatischen Generierung interaktiver Diagramme für beliebige quantitative und qualitative Daten.

#### 4 Nutzen für Informationsanbieter

Auf der Seite der Inhaltsaufbereitung bieten die Wissensnetze dem Informationsanbieter eine ausgezeichnete Steuerung seiner Inhalte, die erheblich über den Horizont herkömmlicher Content Management Systeme hinausgeht. Neben der Zukunftssicherheit des Wissenskapitals, die durch eine strukturierte und standardisierte Beschreibung garantiert wird, liegen weitere Vorteile in

- der einfachen Extraktion von Spin-off-Produkten,
- der inhaltlichen Integration verteilter, heterogener Datenquellen und
- der automatischen Unterstützung bei der Indexierung.

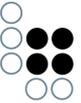
Per Wissensnetz strukturierte Inhalte ermöglichen auf einfache Weise die Erzeugung von **Spin-off-Produkten** und gewähren damit eine flexible und schnelle Reaktion auf veränderte Marktbedürfnisse. Eine vernetzte Fachencyklopädie „Kunst“ könnte beispielsweise Ausgangspunkt spezieller Ausgaben werden, etwa zu einzelnen Künstlern,



Stilrichtungen und Ländern, oder die Inhalte können anderen Informationsanbietern als Hintergrundmaterial für eher kurzlebige, aktuelle Publikationen lizenziert werden. Es ist sogar denkbar, dass standardisierte Wissensnetze als eigenständige Produkte weiterverkauft werden. In einem unternehmensinternen Informationsmarkt könnte das Wissensnetz die Rolle eines Unternehmensgedächtnis einnehmen. In diesem Zusammenhang bedeutet Spin-off-Produkt dann, dass eine „Stabsstelle Wissensmanagement“ aus dem ganzen Unternehmen die Expertise zu einem bestimmten Thema sammelt und diese aufgabenspezifisch einer Projektgruppe „als Produkt“ zur Verfügung stellt.

Die **inhaltliche Integration verteilter heterogener Quellen**, relevant etwa für Portal-, Community-, Intra- und Extranetanwendungen, kann durch Einsatz einer Wissensbasis effizient unterstützt werden, indem hier in einer Art Interlingua die verschiedenen Datenquellen beschrieben und sowohl untereinander als auch mit sinnvollen Benutzeranfragen in Beziehung gesetzt werden. Schließlich hat es sich in der Vergangenheit erwiesen, dass die Zukunftssicherheit der Inhalte nicht durch technische Formate garantiert werden kann, sondern nur durch eine möglichst explizite Beschreibung. Diese Integration wird durch den Einsatz einer effizienten Software-Busarchitektur verstärkt, an welche die verteilten Quellen samt Wissensnetz durch entsprechende Adapter angeschlossen werden können. Auf die Erstellung einer neuen, integrativen Gesamtdatenbank kann somit verzichtet werden – dieser Ansatz ist also sehr pragmatisch und kostengünstig. Zudem erlaubt er ein komfortables Integrationsmodell, denn die Wissensmodellierung kann mit der Anzahl und Heterogenität der Quellen mitwachsen. Schließlich basiert die Bustechnologie auf dem Internetprotokoll http in Verbindung mit dem XML-Standard, so dass auch externe Quellen weltweit eingebunden werden können.

Das aufgebaute Wissensnetz kann als Basis für **voll- oder halbautomatische Indexierungsprozesse** verwendet werden. Dies ist insbesondere dann relevant, wenn große Dokumentenbestände inhaltlich erschlossen werden sollen. Ohne Wissensnetz sind mit der Indexierung in der Regel hohe Kosten verbunden, weil herkömmliche Dokumentenmanagement- und Redaktionssysteme eine vollautomatische Erschließung von unstrukturierten Inhalten nicht anbieten.



## 5 Die Stärken der wissensbasierten Modellierung

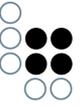
Die Vorteile der wissensnetzbasierter Modellierung sind verbesserte Strukturierungs- und Zugriffsfunktionen. Diese ergeben sich aus den Prinzipien der

- Objektidentität,
  - Trennung von Begriff und Benennung,
  - typisierten Beziehungen,
  - intelligenten Suchverfahren
- und der
- einfachen Pflege des Wissens.

### Objektidentität

Das Prinzip der Objektidentität ist einer der Faktoren, die für die Funktion des Wissensnetzes als Zugriffsstruktur entscheidend sind. Beispielsweise kommt ein Ort wie London, als Geburtsort verschiedener Personen oder als Sitz einer internationalen Organisation etc. in der Wissensbasis vor. Das Prinzip der Objektidentität beruht darauf, dass ein solcher Ort nicht einfach nur als Eigenschaft bei diesen Personen und Organisationen eingetragen wird: In einer objektorientierten Wissensmodellierung wird vielmehr ein eindeutiges Informationsobjekt vom Typ **Ort** angelegt, auf das die Personen und Organisationen verweisen. Dieses eindeutige Informationsobjekt akkumuliert alle Informationen, die in der Wissensbasis aus den verschiedensten Richtungen zu London eingetragen werden und macht sie an einer zentralen Stelle zugreifbar.

Das Prinzip der Objektidentität ist einer der wichtigen Unterschiede zwischen einem wissensbasierten, objektorientierten Ansatz und einer SGML/XML-Strukturierung von Inhalten mit Hilfe von Attributen; letztere kann allerdings eine wichtige Vorstufe darstellen. Eine praktische Implikation der Objektidentität in der **intelligent views**-Wissenstechnologie ist, dass grundlegende Verwaltungsmechanismen bereits von der objektorientierten Entwicklungsumgebung bereitgestellt werden und die Entwicklung eines Informationsdienstes im wesentlichen aus der Umsetzung der Applikationslogik besteht.



## Trennung von Begriff und Benennung

Durch die Modellierung allgemeinen begrifflichen Wissens und die damit einhergehende Trennung von Begriff und Name werden Sprachvarianten oder Synonyme auf einen Begriff zurückgeführt. Dadurch wird auch die Indexierung und somit das Wissen über das Textmaterial **normiert**. Diese Normierung ermöglicht, Begriffe unabhängig von ihrer Formulierung im Textmaterial wiederzufinden.

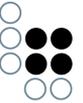
## Typisierte Beziehungen

Die Modellierung eines Wissensnetzes nutzt verschiedene Typen von Beziehungen zwischen den Objekten: Die Unter-/Oberbegriffsrelation, die Teile-/Ganzes-Relation und die Instanzbeziehung sind Beispiele für allgemeine Beziehungen; dazu kommen sachgebietsspezifische Beziehungen wie **geboren-in** zwischen Personen und Orten oder **Sitz-von** zwischen Orten und Organisationen. Eine Differenzierung in verschiedene Typen von Beziehungen erlaubt zum einen, selektiv Informationen zu entnehmen und sich auf die Aspekte zu konzentrieren, die für den Benutzer aktuell von Interesse sind. Sie ermöglicht es dem System aber auch, eigene Schlüsse zu ziehen. Beispielsweise aus der Information, dass ein **Lenkrad** ein Teil von **Auto** ist und **Sportwagen** ein Unterbegriff von **Auto**, kann geschlossen werden, dass auch ein Sportwagen ein Lenkrad hat, wobei diese Beziehung nicht nochmal explizit bei **Sportwagen** verankert werden muß.

## Intelligente Suchen

Die Suche in einem Wissensnetz mit seinen Objekten und Relationen bzw. in den im Netz aufgehängten Dokumenten unterscheidet sich wesentlich von einer Volltextsuche. Volltextsuche kann immer nur aus der Nähe zweier Wörter (genauer: Wortformen) in einer gegebenen Textmenge auf die inhaltliche Nähe zweier Begriffe schließen. Eine solche statistische Häufigkeitsanalyse liefert letztlich immer nur Assoziationen verschiedener Begriffe; welcher Art die Beziehung ist, bleibt unklar und kann nur vermutet werden. Eine Volltextsuche kennt eben nur eine Relation: das Wort A taucht zusammen mit dem Wort B auf.

Das damit verbundene Problem liegt auf der Hand: Häufig benutzte idiomatische Wendungen wie z. B. „König“ und „Kunde“ rücken in unmittelbare Nähe, obwohl diese Nähe allein durch ein gängiges Wortbild und nicht durch eine allgemeine inhaltliche Beziehung begründet ist. Objekte



inhaltlicher Relationen, beispielsweise Über- und Unterbegriffe oder Synonyme, kommen hingegen typischerweise gerade nicht in einem Satz vor, weisen also im Sinne der Volltextsuche keine starke Verbindung auf. Natürlich kann man die statistische Analyse der Volltextsuche auf sehr große Umgebungen anwenden, um beispielsweise Synonyme mit einer größeren Chance zu erfassen, allerdings erhöht sich dann auch die Fehlerquote.

### **Pflege eines Wissensnetzes**

Die Struktur des Wissensnetzes bringt es mit sich, dass ein Großteil seiner Objekte, nämlich die der höheren Hierarchieebenen, über lange Zeit aktuell sind und nur minimale Pflege benötigen. Unabhängig von der Anzahl der im Netz abgelegten Begriffe verfügt man also immer über ein stabiles Grundgerüst.

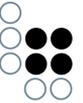
Treten nun aber Bedeutungsänderungen auf, vor allem auf der Ebene der Individuen, lassen diese sich in einem Wissensnetz mit wenig Aufwand einpflegen; mit einem einzigen Eingriff macht man Gerhard Schröder beispielsweise zum Ex-Bundeskanzler und eine andere Person zu seiner Nachfolgerin oder seinem Nachfolger. Alle mit dem Schlagwort „Schröder“ ausgezeichneten Dokumente werden von nun an über das Objekt „Ex-Bundeskanzler“ bzw. seine Synonyme im Wissensnetz gefunden. Die oben erwähnten statistischen Methoden tun sich mit solchen Änderungen hingegen wesentlich schwerer: ein Benutzer einer Volltextsuche muss warten, bis der Name der neuen Kanzlerin bzw. des neuen Kanzlers häufig genug in Verbindung mit dem Begriff „Bundeskanzler“ in den analysierten Texten auftritt, bevor seine Suche das richtige Ergebnis liefert. Eine Trennschärfe gibt es nicht.

## **6 Wissen steuern und Wissen nutzen – aber wie?**

Welche Werkzeuge zum Aufbau eines Wissensnetzes im obigen Sinn werden nun gebraucht?

- Redaktionswerkzeuge, mit denen Informationsanbieter das Wissen strukturieren und damit steuern können und
- Wissensnutzungs-Werkzeuge mit denen Informationssuchende das Wissen wirklich nutzen können

Im folgenden wird beschrieben, welchen Anforderungen sie genügen müssen.



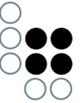
## Anforderungen an die Wissensredaktionswerkzeuge

Ein Wissensnetz sollte sich während des Aufbaus immer in einem konsistenten Zustand befinden. Dazu sollten neben einem vorgegebenen Schema von Objekttypen und Relationen auch Hilfs- und Überprüfungs-funktionen eingebaut sein (s.u.), die die **Konsistenz** für die definierten Objekte nach festgelegten Regeln prüfen.

Ein typisches Problem in diesem Zusammenhang ist die **Änderung eines beliebigen Objekts im Netz**, von der in der Regel viele andere Objekte ebenfalls betroffen sind. Angenommen, eine Anwendung verlangt, Personen bestimmte Rollen (Berufe, Funktionen etc.) zuzuordnen; dann sollte die Löschung einer Rolle automatisch eine Aktualisierung aller Personen nach sich ziehen, denen diese Rolle zugewiesen war. Als Beispiel für ein noch komplexeres Regelwissen mag folgendes Szenario dienen: Eine Patientendatenbank verwaltet Krankheitsdaten und Behandlungsmethoden der einzelnen Patienten, darunter Allergien, Befunde, verabreichte Medikamente, Wirkstoffe etc. Für einen speziellen Wirkstoff stellt sich nun heraus, dass er zu starken allergischen Reaktionen führen kann. Dem Informationssystem ist bekannt, dass dieser Stoff in einer Reihe von Stoffverbindungen und Medikamenten enthalten ist. Dann sollte die Patientendatenbank schnellstmöglich so aktualisiert werden, dass alle Patienten, die die Medikamente mit dem Wirkstoff oder (Wirkstoffen aus den Verbindungen) als Therapiemaßnahme erhalten, auch automatisch gefunden und die behandelnden Ärzte benachrichtigt werden. Durch einen Verfolgungsmechanismus (**Inferenzmechanismus**) über die Beziehungen der Medikamente mit ihren Inhaltsstoffen und ihren Bestandteilen oder den Synonymen der Stoffverbindungen kann dies gewährleistet werden.

Der unterliegende Inferenzmechanismus kann so integraler Bestandteil der Konsistenzkontrolle werden.

Dies sind nur zwei Beispiele für Regelmechanismen. Andere Kontrollen achten z. B. darauf, dass sich in hierarchische Relationen wie Unter-/Oberbegriff keine Zyklen einschleichen, oder halten das Netz von Redundanzen frei, die später Widersprüche verursachen könnten. Das Wissenssystem sollte diese Regeln implementiert haben und bereits beim Aufbau und Editieren des Netzes dafür sorgen, Eintragungen, die zu solchen Inkonsistenzen führen, zu vermeiden bzw. zumindest darauf hinzuweisen. Eine **automatische (oder halbautomatische) Kontrolle der**



**Konsistenz** ist besonders wichtig, wenn mehrere Redakteure am selben Wissensnetz arbeiten. Technisch bedeutet Konsistenzkontrolle, dass lokale Änderungen des Wissensnetzes durchaus globale Auswirkungen haben können. Zu deren automatischer Überwachung müssen deshalb regelbasierte, geschachtelte Mechanismen zur Verwaltung gekapselter Bearbeitungseinheiten (Transaktionsmechanismen) implementiert sein. Im Falle des Fehlschlagens einer Transaktion werden Rollback-Operationen angestoßen, die das Netz auf jeden Fall in den letzten konsistenten Zustand zurückführen.

Darüber hinaus ist es wichtig, dem editierenden Redakteur einen guten **Überblick über die Implikationen seiner Operationen**, den bisherigen Stand der Modellierung sowie seine aktuellen Optionen zu geben. So sollte dem Redakteur zu jedem Zeitpunkt klar sein, welche Verknüpfungen und Attribute für die Objekte, die er bearbeitet, überhaupt möglich sind und wie viele Verknüpfungen vom selben Typ sich eintragen lassen (Personen können mehrere Rollen haben, aber nur ein Geburtsdatum etc.)

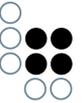
Schließlich müssen die Wissensredaktionswerkzeuge es dem Redakteur erlauben, seine **Arbeit effektiv zu organisieren**, sie sollten also verschiedene Filter, Suchmöglichkeiten und Darstellungsstile sowie eine Verwaltung der Objekte in eigens dafür vorgesehenen Ordnern enthalten.

Weiterhin müssen die erstellten Daten mit anderen Werkzeugen **austauschbar** sein. Das modellierte Wissensschema sollte in einem Standardformat (als SGML- bzw. XML-DTD) als plug-in in die Redaktionswerkzeuge importiert werden können. Das Wissensmodell ist dann operationalisierbar, d. h., das Redaktionssystem unterstützt die in der DTD definierten Datenverknüpfungs- und Konsistenzregeln.

### **Anforderungen an die Werkzeuge zur Wissensnutzung**

Um das in Form eines Wissensnetzes abgelegte Wissen tatsächlich auch nutzen zu können, sind neben sinnvollen Redaktionswerkzeugen zum Aufbau und zur Pflege des Netzes auch geeignete, an die spezielle Struktur und Leistungsfähigkeit der Netze angepasste Module zwingend erforderlich. Die Anforderungen an die Instrumente zur Wissensnutzung ergeben sich also direkt aus den Eigenschaften der Wissensnetze:

- Die Suchanfragen an das Netz sollen die im Netz angelegten **semantischen Beziehungen voll nutzen** können.



- Verschiedene **Sichten auf das Netz** sollen dazu dienen, das komplexe und vielschichtige Informationsgefüge des Wissensnetzes durch die Bedürfnisse der konkreten Anwendung zu filtern.
- Der Benutzer muss die Möglichkeit haben, das Netz von einem selbst gewählten Startpunkt aus zu „**erforschen**“, indem er sich von der Umgebung des gerade gewählten Begriffes zum nächsten Schritt inspirieren lässt.
- In bestimmten Anwendungen ist es wünschenswert, das Wissensnetz in einer **vertrauten Form** (Listen, Kataloge) präsentieren zu können.
- Außerdem sollten sich die aus dem Wissensnetz abgerufenen Informationen u. U. in eine **vorgegebene Umgebung und ein feststehendes Layout integrieren** lassen.
- Schließlich verlangen die unterschiedlichen Informationstypen, die bei einer Suchanfrage an das Netz als Antwort zu erwarten sind, auch **unterschiedliche Darstellungsformen**: ein Unternehmensvergleich liefert evtl. eine Reihe von Wirtschaftsdaten, die sich vorteilhaft in einem Diagramm oder einer Tabelle anordnen lassen, eine andere Suchanfrage führt gezielt zu einem einzigen Text, der an den Benutzer weitergereicht wird.

## 7 K-Infinity

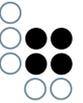
**K-Infinity** ist die Antwort von **intelligent views** auf das oben skizzierte Anforderungsprofil zu Aufbau, Pflege und Nutzung eines Wissensnetzes. Es bietet sowohl dem Redakteur vielfältige Unterstützung für die objektorientierte Wissensmodellierung als auch dem Nutzer eine breite Palette an Möglichkeiten zur Erschließung des vernetzten Wissens.

### Werkzeuge für Wissensanbieter

Folgende **redaktionelle Komponenten** gehören zu **K-Infinity**:

#### Schema-Editor

Der **Schema-Editor** bildet den Arbeitsplatz des Wissensingenieurs und erlaubt die Definition von Objekt- und Beziehungstypen bzw. der dazugehörigen operationalen Semantik. Objekt- bzw. Beziehungstypen sind i.A. anwendungsspezifisch. Für eine Anwendung im Bereich



Literatur dürften beispielsweise Informationsobjekte wie Schriftsteller, Werk, Entstehungsort und –zeit, Verlag sowie Beziehungen der Form „hat geschrieben“, „war befreundet mit“ oder „hat herausgegeben“ wichtig sein. Für eine Anwendung im Bereich Investment sind Branchen, Firmen, deren Performance-Daten sowie Anteilsbeziehungen zwischen Firmen von Bedeutung. Allgemeinlexika wiederum möchten eine übergreifende Sicht auf die Welt bieten und verlangen dementsprechend die Möglichkeit, zunächst vollkommen unabhängig erscheinende Sachverhalte zusammenzuführen zu können, etwa dass eine Firma als Mäzen die literarischen Werke eines Schriftstellers mit einem Preis bedacht hat.

Jede dieser Bedeutungen kann mit dem **Schema-Editor** definiert werden.

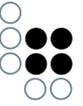
### **Knowledge-Builder**

Der **Knowledge-Builder** ist die Kernkomponente des Redaktionssystems **K-Infinity**. Mit seiner Hilfe können sowohl neue Wissensobjekte angelegt, gelöscht, umbenannt und modifiziert, als auch Beziehungen zwischen einem Objekt und anderen Objekten angelegt werden. Dies kann in zwei verschiedenen Sichten geschehen:

Der **Graph-Editor** eignet sich für eine Gesamtsicht auf das Beziehungsgeflecht von Objekten und bietet die Möglichkeit, dieses auf Basis des definierten Schemas zu erweitern und das Wissen mit Hilfe der vom Redaktionssystem unterstützten Konsistenzregeln zu kontrollieren. Zur Grundfunktionalität des **Graph-Editors** gehört ein interaktiver Netzwerklayoutalgorithmus zur Exploration des Wissensnetzes. Das im Wissensmodell angelegte Geflecht von Beziehungen kann genutzt werden, um geeignete semantische Filter zu implementieren, die die Komplexität des Explorationsprozesses auf das benötigte Mindestmaß reduzieren.

Der **Konzept-Editor** ist eine Datensicht, welche es erlaubt, jeweils ein Objekt mit seinen durch semantische Beziehungen verknüpften Nachbarobjekten in den Mittelpunkt zu rücken. Es ist eine Ergänzung des Graph-Editors, da diese Sicht es zulässt, die Beziehungen bzw. deren Attribute auf komfortable Weise im Detail zu erfassen und zu modifizieren.

Der Wissensredakteur benötigt neben den eigentlichen Werkzeugen zum Editieren des Netzes noch Organisationshilfen, die es ihm erlauben, seine tagtägliche Arbeit zu



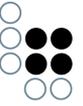
strukturieren. Diesem Zweck dient der **Organizer**, mit dem sich die Objekte des Wissensnetzes ordnen und gruppieren lassen, entweder von Hand oder durch die Ausnutzung vorgegebener Objekteigenschaften; so kann man sich z. B. alle Objekte, die vor einem bestimmten Termin angelegt wurden, oder alle Begriffsobjekte, die mehr als zehn Unterbegriffe auf sich vereinigen, in einem Arbeitsordner zusammenstellen lassen.

### **Knowledge-Accelerator**

Der **Knowledge-Accelerator** ist eine Ergänzung zum **Knowledge-Builder** und dient der verteilten (eingeschränkten) Pflege des Wissensnetzes. Damit mehrere, auch auf verschiedene Standorte verteilte Nutzer gleichzeitig das Wissensnetz pflegen können, ist der **Knowledge-Accelerator** ein browserbasiertes Werkzeug, mit dem über das Internet Individuen angelegt, geändert und gelöscht werden können.

### **Das Markup-Tool**

Das **Markup-Tool** dient dazu, neue Inhalte zu indexieren. Die Begriffe, die dabei in den Texten gefunden werden, können schon im Wissensnetz vorliegen und so für die Verknüpfung der Texte mit dem Netz (ihre Einordnung in den Zusammenhang allgemeiner Begriffssysteme) sorgen. Es können aber auch neue Begriffe identifiziert und ins Wissensnetz eingeführt werden. Das Auffinden der relevanten Begriffe kann mindestens ein halbautomatischer Prozess sein, bei dem ein für das Textmaterial geeignetes Textanalysewerkzeug dem Redakteur Wörter zur Auszeichnung vorschlägt. Grundlage dieser Vorschläge ist – neben linguistischen Verfahren (z. B. Stemming, also das Berücksichtigen von Wortstämmen) und Spezialwörterbüchern (z. B. mit Personen- oder Ortsnamen) – auch das bisherige Auszeichnungsverhalten: wird beispielsweise dem Wort „Halle“ mehrmals der gleichnamige Ort und nicht das Gebäude als Objekt zugewiesen, schlägt das System fortan direkt den Ort vor. In vielen Fällen dient zudem das mit Hilfe der Redaktionswerkzeuge aufgebaute Wissensnetz selbst als Grundlage des Textanalyse-Prozesses. Die Auszeichnung eines Dokumentes mit Hilfe des **Markup-Tools** nutzt also neben den vollautomatischen, von der Volltextsuche her bekannten, Strategien zusätzlich das Wissensnetz.



## Plug-in-Wissensnetzschablonen

Plug-in-Wissensnetzschablonen sind von **intelligent views** oder unseren Partnern vorgefertigte Wissensnetze, die zusammen mit dem Redaktionswerkzeug ausgeliefert werden können. Sie erleichtern oft den Einstieg in die Modellierung eines Themas und dienen somit als konzeptuelle und technologische Grundlage zur Ableitung eines für den Kunden speziellen Wissensnetzes. Da die Wissensschablonen modular angelegt sind, ist es möglich, verschiedene Module miteinander zu kombinieren. Beispiele für solche Schablonen sind:

- Literatur und Kultur des 20. Jahrhunderts,
- Staaten und internationale Organisationen,
- Personen des öffentlichen Lebens und ihre Rollen und Funktionen in Politik und Wirtschaft,
- Allgemeine Ontologien.

## Werkzeuge zur Wissensnutzung

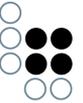
Die Anforderungen an die **Wissensnutzung** hat **K-Infinity** in die folgenden Komponenten umgesetzt:

### Semantic-Matcher

Der Semantic-Matcher vergleicht Dokumente auf Ähnlichkeit bezüglich des in der Anwendung modellierten semantischen Wissens. Auch in diesem Zusammenhang sind die oben erwähnten normierten Indexvektoren wichtig, denn eine Normierung des Indexvokabulars trägt wesentlich dazu bei, eine Ähnlichkeitsmetrik zu operationalisieren. Darüber hinaus kann das semantische Wissensnetz genutzt werden, um den Begriff Ähnlichkeit weiter zu fassen als Wortähnlichkeit; so sind sich zwei Dokumente, deren indexierte Begriffe einen geringen Abstand im semantischen Netz haben, inhaltlich ähnlich, obwohl u. U. auf Textebene keine oder nur wenige explizite Gemeinsamkeiten vorliegen.

### Semantic-Finder

Diese Komponente ist das „Herzstück“ der Werkzeuge zur Wissensnutzung, weil sie gezielt die Struktur des Wissensnetzes für die Formulierung einer Suche auf dem Netz ausnutzt. Der Benutzer hat die Möglichkeit, die im Netz angelegten Individuen über die ihnen zugeordneten



Kategorien und Relationen zu suchen. Beispielsweise könnte man in einem Literatur-Wissensnetz nach einem Werk einer bestimmten Person suchen, in der eine andere Person als Figur vorkommt. Der Vorteil dieser Suchmöglichkeit ist die exzellente Qualität der Antworten, da diese ja das Wissensnetz direkt widerspiegeln.

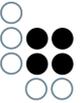
### **Net-Navigator**

Ein Wissensnetz ermöglicht neben automatischem Matching (Auffinden ähnlicher Inhalte) und intelligenten, gezielten Suchen auch die interaktive Exploration der Inhalte auf thematischer Ebene. Für diese Exploration stellt **intelligent views** den **Net-Navigator** zur Verfügung, der eine sehr allgemeine Form der Navigation auf beliebigen relationalen Strukturen bietet. Dem Benutzer wird jeweils ein übersichtlicher Ausschnitt des Wissensnetzes als Graph gezeigt, eine Navigationsoperation auf einem Objekt öffnet dessen Nachbarobjekte. So verschiebt sich stetig die Sicht auf das Netz, neue Bereiche kommen ins Blickfeld, während andere Bereiche aus dem sichtbaren Bereich verschwinden. Zur Qualität dieser Navigation trägt in erster Linie eine klare Darstellung der Nachbarschaftsbeziehungen bei einem gleichzeitig möglichst fließenden Übergang von einem Zustand zum anderen bei. Der **Net-Navigator** nutzt einen internen Optimierungsalgorithmus, um diese beiden Aspekte zufriedenstellend miteinander zu vereinbaren.

Analog zum Graph-Editor ist es auch beim Net-Navigator möglich, die durch die Anwendung festgelegten semantischen Relationen oder benannten Gruppen von Relationen als Filter auf das Wissen zu legen und somit die Navigationsmöglichkeiten einzuschränken.

### **Layout-Engine**

Dieses Modul hat die Aufgabe, mediale Inhalte (Texte, Grafiken, Tabellen, Bilder etc), die aus möglicherweise unterschiedlichen Wissensquellen stammen, entsprechend typografischer Regeln und stilistischer Vorgaben vollautomatisch in ein effektives und ansprechendes Seitenlayout zu transformieren. Diese Situation ist in vielen On-Demand-Situationen erforderlich, insbesondere dann, wenn Individualisierung des Zugriffs wichtig ist und schematisch gestaltete Layoutschablonen unerwünscht sind. Wichtig in diesem Kontext ist, dass die **Layout-Engine** die Corporate Identity des Anbieters und andere Vorgaben nicht antastet, gleichzeitig aber ein maximal individualisiertes Layout generiert, ohne dass der Nutzer eingreifen



müsste. Die Layout-Engine wird auch zur Gestaltung von Webseiten und Webportalen verwendet.

### Visualisation-Engine

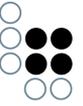
Die **Visualisation-Engine** ist ein Visualisierungswerkzeug, das die Aufgabe hat, objekt-relationale Daten automatisch zu klassifizieren und diese Klassifikation als Eingabe für die vollautomatische Konstruktion von interaktiven Diagrammen zu verwenden. Dem liegt die Idee zugrunde, grafische Darstellungen und Interaktionen gezielt als Explorationswerkzeuge nutzen zu können und so zu einem visuellen Knowledge-Mining-Prozess zu gelangen. Entscheidend ist eine intuitive Unterstützung des Wissensnutzers in seiner explorativen Tätigkeit durch die generierten grafischen Darstellungen, z. B. dadurch, dass Datenvergleichsoperationen mit Hilfe von grafischen Darstellungen selbsterklärend werden und damit einfach zu verstehen sind, oder dass Datenaggregations-Operationen visuell nachempfunden werden können und dadurch ein konsistenter Explorationsmechanismus gewährleistet wird.

## 8 Werkzeuge zur Konfiguration von Anwendungen

Es liegt auf der Hand, dass den Möglichkeiten, die in einem Wissensnetz und seiner intelligenten Nutzung stecken, nicht mit einer vorgegebenen Konfiguration der entsprechenden Werkzeuge gedient ist. Die Informationsdienste, die die Kunden von **intelligent views** auf ihren Wissensnetzen aufbauen, stammen aus den verschiedensten Bereichen; das zugrundeliegende Geschäftsmodell, zusammen mit den jeweiligen Informationsquellen wird jeder Anwendung einen ganz eigenen Charakter verleihen. Darüber hinaus ist die Unverwechselbarkeit eines Produktes z. B. im Publishing-Sektor ein entscheidender Faktor für den Erfolg. Besondere Aufmerksamkeit schenkt **intelligent views** daher den Möglichkeiten des Informationsanbieters, die Wissensnutzungskomponenten seiner Anwendung anzupassen. Dazu gehören

- die Auswahl der Such- und Matching-Strategien  
sowie
- die Auswahl von Navigationstyp und Stilvorgaben für die Interface-Elemente.

Das Layoutmodul bestimmt das Gesamterscheinungsbild der Anwendung, hier ist der Konfigurationsbedarf am höchsten und dementsprechend die Konfigurations-



möglichkeiten am komplexesten. Zur Verfügung gestellt werden:

- die grafisch-interaktive Definition von Formaten,
- die grafisch-interaktive Definition von Regionen,
- die Angabe von Referenzpunkten und Ausdehnungsverhalten sowie
- die Zuordnungen von Quellelementen zu Formaten und Regionen.

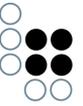
Im Visualisierungsmodul schließlich umfassen die Konfigurationsmöglichkeiten

- die Bestimmung der grafischen Ressourcen und
- die Auswahl der Interaktionsmöglichkeiten und des Anpassungsverhaltens.

## 9 Zusammenfassung

Wissensnetze haben sich als erfolgreicher Ansatz erwiesen, einen dynamisch wachsenden Wissenspool zu kontrollieren. Dem Wissensnutzer wird die Möglichkeit gegeben, die dem Wissenspool innewohnende Struktur, also seine Schlüsselbegriffe und –themen sowie deren semantische Beziehungen untereinander, dazu zu verwenden, sowohl intelligente Suchanfragen an den Wissenspool richten zu können, als auch eine schrittweise Erschließung desselben vorzunehmen. Diese Möglichkeiten gehen weit über die traditionellen Verfahren hinaus, gesuchte Begriffe und Themen über ihre Verwendung in einer gegebenen Dokumentenmenge aufzufinden.

**intelligent views** hat mit **K-Infinity** eine Reihe von effizienten, einfach zu benutzenden Werkzeugen entwickelt, mit denen Wissensnetze einerseits aufgebaut und gepflegt werden und die das Wissen andererseits einem Nutzer in verschiedenen Ansichten und mit intelligenten Suchstrategien erschließbar machen.



## Glossar

### Ähnlichkeit

Eigenschaft von Objekten (z.B. Begriffe, Sätze, Texte), die bestimmte Merkmale gemeinsam haben, also eine Verbindung aufweisen. **Semantische** Ähnlichkeit liegt dann vor, wenn die Objekte ähnliche Bedeutung haben (z.B. Hund und Katze). **Syntaktisch** ähnliche Wörter beinhalten übereinstimmende Zeichen.

### assoziative Verknüpfung

Verbindung zweier Objekte, die auf ihrer Einbettung in einen gemeinsamen Kontext beruht, z. B. die Verbindung zwischen „Wüste“ und „Kamel“.

### Client

Rechner (oder Softwareeinheit), der zur Erfüllung seiner Aufgaben auf Daten und Dienste eines zentralen Rechners (oder Softwareeinheit), den → Server, zugreift.

### Content-Management-System

CMS, Softwaresystem zur Erzeugung und Pflege von Webinhalten. Typische CMS unterstützen den redaktionellen Workflow, regeln Zugriffsrechte, bieten Archivfunktionen, binden externe Datenquellen ein und unterstützen Webstandards (z.B. XML).

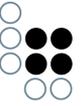
### Dokumenten-Management-System

DMS, ein System, mit dem der gesamte Prozess einer Dokumentenproduktion gesteuert und verwaltet werden kann. Es stellt in der Regel Funktionen für die Erstellung, Überarbeitung, Freigabe, Versionspflege, Nutzung durch Index- oder Volltextsuche, Langzeitarchivierung etc. von Dokumenten zur Verfügung. Die Schnittstelle zu anderen Desktop-Anwendungen wird durch den ODMA (Open Document Management Alliance oder Open Document Management API)-Standard gewährleistet.

### DTD → SGML

### Eltern-Kind-Relation

in einem Wissensnetz die Relation eines Kind-Objektes zu anderen Objekten (Eltern-Objekten), aus denen es inhaltlich (Vererbung der Eltern- auf Kindeigenschaften) oder existenziell (Kinder-Objekte werden durch die Menge der Eltern-Objekte erzeugt) hervorgeht.



### **explorative Suche**

Suchstrategie, bei der Teilergebnisse das Suchverhalten oder den Lösungsansatz beeinflussen. Beispiel: WWW-Surfen. In der Regel ist das Suchziel zu Beginn noch nicht vollständig festgelegt.

### **Faktorisierung**

der Vorgang des Zerlegens und Neustrukturierens eines Objektes in zusammengehörige Einheiten (funktional und inhaltlich). So kann man beispielsweise aus typischen Datensätzen über Personen Name und Geburtsdaten durch eine Faktorisierung von der Anschrift trennen.

### **Graph → Wissensnetz.**

### **Individuum**

konkrete Ausprägung eines Objekts eines bestimmten Konzepts. So sind z. B. „Hänsel“ und „Gretel“ Individuen des Konzepts „Märchenfigur“.

### **Inferenz**

das Ableiten von neuen Fakten aus bekannten Fakten gemäß gegebener Regeln.

### **Knowledge-Mining**

„Graben nach Informationen“, gezielte Auswertung sehr großer Datenmengen mit dem Ziel, vorher nicht bekannte Zusammenhänge aufzudecken.

### **Mapping**

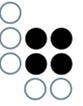
Abbildung zwischen Konzepten/Individuen aus zwei unterschiedlichen Wissensnetzen. Ein Konzept/Individuum aus Netz A wird auf einen oder mehrere Konzept/-Individuen aus Netz B abgebildet, die gleiche oder ähnliche Bedeutung haben. Da die Abbildung nicht unbedingt vollkommen eindeutig ist, muss man das Mapping evtl. mit einer Gewichtung versehen.

### **Markup**

Aufhängung eines Dokumentes in ein Wissensnetz durch Verknüpfung mit Konzepten des Wissensnetzes. Im einfachsten Fall wird das Dokument als ganzes in das Netz eingebunden, es können aber auch Teile des Dokumentes adressiert und mit verschiedenen Konzepten verknüpft werden.

### **Ontologie**

ein Regelwerk, das festlegt, welche Behauptungen innerhalb einer Faktenmenge gemacht werden können und was diese Behauptungen bedeuten.



### **Relation**

allgemein eine Beziehung zwischen Elementen einer Menge. Im Zusammenhang mit →Wissensnetzen sind damit Beziehungen zwischen zwei Wissensobjekten bestimmten Typs gemeint.

### **Relationstupel**

mathematische Repräsentation einer Relation, bestehend aus dem Bezeichner der Relation und den zueinander in Beziehung zu setzenden Objekten.

### **Retrieval, semantisches →semantisches Retrieval.**

#### **semantisches Retrieval**

Extraktion von Daten basierend auf semantischen Vorgaben und nicht, wie meist üblich, aufgrund von Wörtern, denen eine externe Semantik zugeordnet ist (Volltextsuche).

### **Server**

Rechner (oder Softwareeinheit), der es anderen, mit ihm verbundenen, Nutzern erlaubt, auf seine Daten und Programme zuzugreifen.

### **SGML**

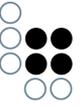
Abk. für Standard Generalized Markup Language, ein Standard, der festlegt, wie in elektronischen Dokumenten Strukturen und Inhalte beschrieben werden sollen. Die International Standards Organization (ISO) hat SGML 1986 zu einer Norm gemacht. Eine auf Webprogrammierung zugeschnittene Untermenge der sehr komplexen SGML ist →XML. Während sich SGML-Dokumente immer auf eine Dokument-Typ-Definition (DTD) beziehen müssen, gibt es im Rahmen von XML das Konzept der wohlgeformten Dokumente, die auch ohne eine DTD weitergegeben werden können.

### **Stemming**

bezeichnet das Bilden von Wortstämmen. Eine Volltextsuche, die Stemming beherrscht, kann nicht nur das eingegebene Wort, sondern auch dessen konjugierte oder deklinierte Form finden. Wird beispielsweise nach „finden“ gesucht, dann erkennt ein entsprechendes Such-Programm auch Fundstellen mit „fand“ und „gefunden“.

### **Topic Map**

ein strukturiertes Netzwerk von Beziehungen zwischen einer Sammlung von Informationsobjekten. Der



Standard ISO/IEC 13250:1999 beschreibt Topic Maps durch SGML- und HyTime-Strukturen; in diesem Sinn ist ein Topic Map also ein SGML (bzw. XML)-Dokument.

### **Vererbung**

Vorgang des Hinzufügens der Eigenschaften des Oberobjekts zum Unterobjekt (Erbe). Damit ist das Unterobjekt immer auch in Kontexten sichtbar, in dem Oberobjekte betrachtet werden. Die bessere Bezeichnung wäre also „Erbung“, da der Vorgang vom Erben ausgeht und nicht umgekehrt.

### **Wissensmanagement**

ganz allgemein der Versuch, die vorhandenen Informationen beispielsweise eines Unternehmens zu klassifizieren und zu organisieren. Im einfachsten Fall reicht dazu eine →Dokumenten-Management-System, ambitioniertere Ansätze wie die auf →Wissensnetzen basierende Technologie von **intelligent views** haben hingegen die Zielsetzung, ein vorhandenes Wissen unabhängig von einer Dokumentenbasis zu modellieren.

### **Wissensnetz**

semantisches Netz, in dem Konzepte über Eltern-Kind-Beziehungen und inhaltliche semantisches Netz, in dem Konzepte über Eltern-Kind-Beziehungen und inhaltliche Verknüpfungen einen großen Graphen aufspannen und in dem Individuen als reale Ausprägungen der abstrakten Konzepte Informationen tragen.

### **XML**

Abk. für e**X**tensible **M**arkup **L**anguage, eine verkürzte Version der →SGML, die sich vor dem Hintergrund entwickelt hat, dass sich SGML im Web nicht durchsetzen konnte. Die Erweiterbarkeit von XML bildet den Hauptunterschied zu HTML. HTML ist für den Normalbenutzer nicht erweiterbar. Über jede Änderung an der HTML-Syntax muss das W3C beraten. Eine Auszeichnungssyntax, die im Rahmen von XML definiert wird, kann dagegen ohne Probleme durch den Web-Designer erweitert werden.