

Seminararbeit

Andre Borck – Matr. Nr. 6066951

Christoph Laroque – Matr. Nr. 6067435

*Konzeption und Implementation einer Rich-Text-Analyse zur Generierung von
Wissensstrukturen aus unstrukturierten Informationsobjekten im
K-Discovery Framework*

Gliederung

1.	Einleitung und Zielsetzung	1
1.1.	Einleitung	1
1.2.	Zielsetzung	1
1.3.	Projekttablauf.....	2
1.4.	Struktur der Ausarbeitung	2
2.	Hintergrund: K-Discovery.....	3
2.1.	Knowledge Management.....	3
2.2.	K-Discovery	3
2.3.	Groupware-based Topic Map Engine	4
3.	Hintergrund Textanalyse.....	5
3.1.	Allgemein	5
3.2.	Textanalyse mit Malaga	5
4.	Motivation	7
5.	Konzeption.....	8
5.1.	Gestaltung Ablaufprozess	8
5.2.	Dokumentenselektion	9
5.3.	Textanalyse	10
5.4.	Ergebnisauswertung.....	11
6.	Implementierung.....	14
6.1.	Gestaltung Ablaufprozess.....	14
6.2.	Dokumentenselektion	15
6.3.	Textanalyse	16
6.4.	Ergebnisauswertung.....	17
7.	Zusammenfassung und Ausblick.....	20
	Anhang	



1. Einleitung und Zielsetzung

1.1. Einleitung

Heutige Industrieunternehmen nutzen in hohem Maße die Möglichkeiten der modernen Informationstechnologie. Das Bilden von weltweiten Netzwerken und der steigende elektronische Datenaustausch lässt die zu bewältigende Menge digitaler Informationen stetig wachsen. Die Unternehmen versuchen, bereits vorhandene Informationen zu strukturieren, um den Zugriff auf internes Wissen effizienter zu organisieren. Für diesen Prozess werden Mechanismen benötigt, die semiautomatisch aus einer Menge von unstrukturierten Objekten Strukturen erkennen und generieren.

Nur dann bleibt die Informationsflut beherrschbar, Informationen können aufgabenspezifisch verknüpft und gelenkt werden, und ein effizienteres Arbeiten kann realisiert werden.

Der Prototyp der Groupware-based Topic Map Engine des K-Discovery Projektes bietet durch die Implementierung von Topic Maps für die Groupware Plattform Lotus

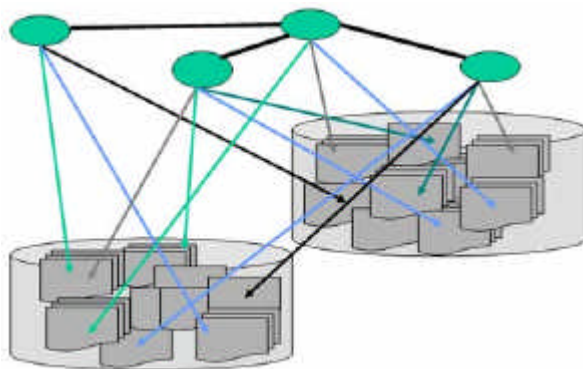


Abbildung 1 Topic Map

Notes/Domino solche Strukturen.

Über eine Metaebene aus miteinander verknüpften Schlüsselwörtern (Topics) kann assoziativ navigiert und damit inhaltlich relevante Dokumente datenbankübergreifend gefunden werden. In der Vergangenheit wurde die

Topic Map durch festcodierte Regeln aus statischen Dokumentenfeldern extrahiert.

1.2. Zielsetzung

Aufgabe der Seminararbeit war die Konzeption und Implementierung einer Textanalysefunktion zur Identifikation von Schlüsselwörtern aus Richt-Text Feldern der Dokumente. Dazu musste zunächst der Prozess der Textanalyse ausgeplant, verfeinert und implementiert werden.

Nach Auswahl der Content-Datenbank soll die Menge der zu analysierenden Dokumente anhand von Filterkriterien eingegrenzt werden können. Als Ergebnis der automatischen Textanalyse müssen die gefundenen Schlüsselworte manuell nachbearbeitet werden können, und als Topics in die Topic Map eingepflegt werden.



K-Discovery:

Konzeption und Implementierung einer Textanalyse

Zu den gefundenen Topics können Assoziationen generiert werden. Erst diese Verknüpfung bringt den realen Mehrwert der assoziativen Navigation.

1.3. Projektablauf

Im Knowledge-Management Projekt „Topic Maps & Ontologien“ wurden bis zum Sommersemester 2003 die Textanalysefunktionalitäten für ein einzelnes Dokument implementiert. Nach einer Recherche der zur Verfügung stehenden Textanalysetools wurde mit der Software Malaga (Universität Erlangen) eine Lösung integriert, die eine Extraktion von Schlüsselwörtern ermöglicht. Eine Anbindung der gefundenen Topics in die Topic Map durch Assoziationen konnte nicht mehr berücksichtigt werden.

In der Semesterarbeit wurde der Prozess der Textanalyse um die Auswahl von mehreren Dokumenten erweitert. Zusätzlich zu der Identifikation der Topics wurde das Generieren von Assoziationen implementiert.

Dazu wurde der bestehende Textanalyseprozess in allen Teilen erweitert. Die Abänderungen wurden zunächst in der Konzeptionsphase ausgeplant, und im weiteren Verlauf umgesetzt. Diese Dokumentation bildet zusammen mit der Präsentation der Ergebnisse den Abschluss der Seminararbeit.

1.4. Struktur der Ausarbeitung

Die Struktur der Ausarbeitung orientiert sich ab Kapitel 4 am zeitlichen Ablauf der Projektphasen. Ausgehend von einer Beschreibung der Motivation werden die Ergebnisse der Konzeptionsphase und in einem weiteren Kapitel die Implementierungsphase beschrieben. Zunächst sollen jedoch in den folgenden zwei Kapiteln (Kapitel 2 & 3) die inhaltlichen Grundlagen der Arbeit erläutert werden. Es werden dort die wesentlichen Begriffe eingeführt, die zum weiteren Verständnis erforderlich sind.

Das folgende Kapitel behandelt die Grundlagen des Knowledge-Management und des K-Discovery Projektes am Groupware Competence Center der Universität Paderborn. Kapitel 3 zeigt die Grundlagen der Textanalyse auf und erklärt den Kernprozess der Textanalyse, der in der Seminararbeit Verwendung findet.



2. Hintergrund: K-Discovery

Neben allgemeinen Aussagen zum Aufgabengebiet des Knowledge Management werden nachfolgend Arbeit und Ziel des K-Discovery Projektes beschrieben, in dessen Kontext diese Seminararbeit entstand.

Basis der Implementierungen ist der Prototyp der Groupware-based Topic Map Engine (GTME), die im dritten Abschnitt dieses Kapitels näher beschrieben werden soll.

2.1. Knowledge Management

Das Prinzip des Knowledge Management kann nicht als etwas grundlegend Neues angesehen werden. Generiertes Wissen wurde und wird bereits in verschiedensten Formen zwischen Personen oder Organisationen ausgetauscht. Da aber die Bedeutung des "Produktionsfaktors" Wissen für den Erfolg von Unternehmen in hohem Maße zugenommen hat, steigt das Bedürfnis, Wissen im Unternehmen systematisch zu managen. Die Integration der Informationstechnologie in die Unternehmensprozesse ermöglicht es, Wissen als Information zu speichern, auszutauschen und gemeinsam zu nutzen.

Der Begriff Knowledge Management steht stellvertretend für eine Strategie, welche die effizientere Nutzung von Wissen in Unternehmen zum Ziel hat. Knowledge Management umfasst also die Entwicklung, Unterstützung, Überwachung und Verbesserung der Strategien, Prozesse, Organisationsstrukturen und Technologien zur Wissens- und damit auch Informationsverarbeitung in einer Unternehmung.

Die Informationstechnologie übernimmt eine wesentliche Unterstützungsfunktion für die erfolgreiche Etablierung eines Wissensmanagements im Unternehmen. Darüber hinaus hängt der Erfolg des Knowledge Managements aber auch von einer geeigneten Gestaltung der Organisation und der Unternehmenskultur ab.

2.2. K-Discovery

Am Groupware Competence Center der Universität Paderborn beschäftigt sich insbesondere das K-Discovery Projekt mit der Umsetzung von Knowledge Management Methoden auf der Groupware Plattform Lotus Notes / Domino. Durch die Verwendung von Topic Maps wird eine Struktur genutzt, die ein assoziatives Navigieren über unabhängige Content Datenbanken ermöglicht, und so ein besseres Identifizieren relevanter Daten erreicht. Kernprozesse sind die Nutzung der Navigation über die Strukturen der Topic Map durch den Endanwender und das Anlegen und Pflegen der Schlüsselwörter (Topics) und Verknüpfungen (Associations), aus denen die Metaebene



K-Discovery:

Konzeption und Implementierung einer Textanalyse

erzeugt wird. Den jeweiligen Schlüsselwörtern muss neben einer zuzuordnenden Obermenge ein Typ zugewiesen werden. So können über die Einbindung von Dokumenten hinaus auch andere Datenformate wie Grafiken oder Verweise (Intra- oder Internet) in die Topic Map aufgenommen werden. Die Associations werden durch einstellbare Regeln automatisch generiert oder können manuell zwischen Topics angelegt werden.

Grundsätzlich geht das Konzept von der Existenz eines Knowledge Workers aus, dessen Hauptaufgabe die Wartung und Pflege der Topic Map ist.

Dem Endanwender soll zwar die verbesserte Struktur der Topic Map nützen, für ihn soll sich jedoch kein zusätzlicher Aufwand ergeben. Durch die automatische Textanalyse werden für den Knowledge Worker Mechanismen erstellt, die eine schnellere Klassifikation der neu entstandenen Dokumente ermöglicht, ohne deren Inhalt genau kennen zu müssen.

2.3. Groupware-based Topic Map Engine

Als prototypische Umsetzung des K-Discovery-Projektes ist die Groupware-based Topic Map Engine (im Folgenden GTME) entstanden. Die Anwenderoberfläche wird hierbei durch ein Servlet realisiert, den „Semantic Surfer“. Für die Konfigurationsarbeit, die Wartung und die Datenpflege der Topic Map steht innerhalb von Lotus Notes eine eigene Oberfläche zur Verfügung, in die auch die Textanalyse integriert wurde.

Topics wie Associations werden dabei in einer angebundenen Oracle-Datenbank hinterlegt. Zu jedem Topic werden die mit ihm verknüpften Dokumente gesichert.

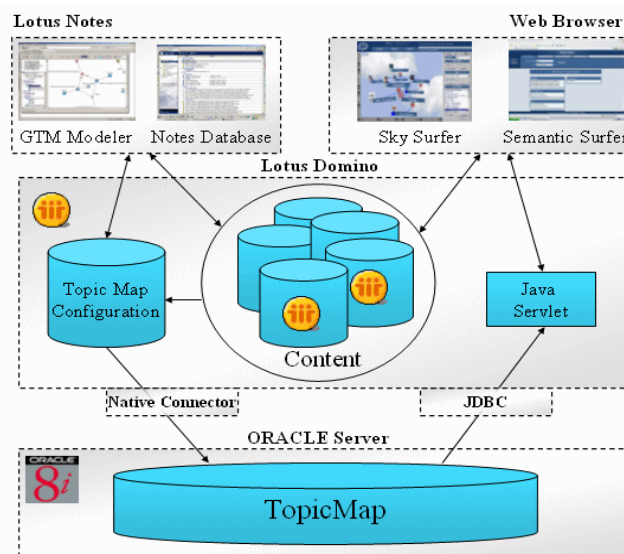


Abbildung 2 Struktur GTME



3. Hintergrund Textanalyse

In diesem Kapitel wird die Aufgabe der Textanalyse im Rahmen der Seminararbeit angerissen, insbesondere wird das in der Seminararbeit verwendete Verfahren beschrieben.

3.1. Allgemein

Grundlegende Aufgabe der Textanalyse ist das automatische Erkennen von Dokumenteninhalten. Verschiedene Verfahren mit jeweils unterschiedlichen Ergebnistypen stehen zur Verfügung. Neben einer Kurzzusammenfassung oder der Ausgabe von Schlüsselwörtern ist auch die Klassifikation von mehreren Dokumenten zu Dokumentengruppen ein Aufgabengebiet der Textanalyseverfahren.

Die wesentliche Aufgabe im Rahmen dieser Seminararbeit ist die Extraktion von Schlüsselwörtern (Key Entries) aus einem unstrukturierten Text. Eine Vielzahl am Markt erhältlicher Software-Produkte bietet diese Funktionalität. Sie nutzen teilweise über die Rückführung der Worte auf ihre Stammform hinaus neuronale Netze oder verwandte Techniken, um Verbindungen ähnlicher Wörter zu identifizieren. Damit soll die Güte der Ergebnisse gegenüber reinen Häufigkeitsverfahren verbessert werden.

Der Großteil der vorhandenen Software-Produkte funktioniert jedoch ausschließlich als Client-Software und stand deshalb nicht zur Integration in die Groupware Umgebung der GTME zur Verfügung.

Mit dem Software Tool „Malaga“ der Universität Erlangen, das zu wissenschaftlichen Zwecken zur freien Verfügung steht, wurde ein Produkt ausgewählt, das lediglich die Rückführung der verwendeten Wörter auf die jeweilige Stammform beherrscht. Es kann als externe Applikation in einem Konsolenfenster serverseitig eingesetzt werden. Die deutsche Grammatikbibliothek DMM Malagas ist ebenfalls zu wissenschaftlichen Zwecken frei verfügbar. Auf dieser Basis wurde eine eigene Textanalyse implementiert, auf die im folgenden Abschnitt näher eingegangen werden soll.

3.2. Textanalyse mit Malaga

Prinzipiell wurde die Entwicklung auf ein Auszählverfahren beschränkt. Nach einer einmaligen Konfiguration der Applikationspfade von Malaga durch den Knowledge Worker steht die Textanalyse in der GTME zur Verfügung.

Innerhalb des angestoßenen Textanalyseprozesses werden zunächst alle Texte aus den Richt-Text-Feldern der selektierten Dokumente eingelesen. Der so erzeugte Gesamttext



K-Discovery:

Konzeption und Implementierung einer Textanalyse

wird um Sonderzeichen bereinigt und von Malaga auf die Stammformen hin untersucht. Die gefundenen Substantive werden in die Groupware Umgebung eingelesen und ihre Häufigkeit ausgezählt. Dabei wird für jedes Dokument das Vorkommen der insgesamt häufigsten Schlüsselwörter gespeichert, um später eine Rückführung der selektierten Schlüsselwörter auf die betroffenen Dokumente zu ermöglichen. Das begrenzt die Speicherung von Verknüpfungen zu einem Topic auf die Menge derjenigen Dokumente, in denen das Topic auch real vorkommt.

Neben der Darstellung der absoluten Häufigkeit werden dem Knowledge Worker unterschiedliche Kennzahlen angezeigt, anhand derer er die Qualität des Textanalyseverfahrens subjektiv einordnen kann.

Empirische Tests haben durchweg zufrieden stellende Ergebnisse erbracht, so dass das Verfahren zur weiteren Verwendung festgelegt wurde.



4. Motivation

Im Rahmen des Gesamtprojektes der Integration einer Textanalyse galt es durch diese Seminararbeit insbesondere die Implementation der Associations in den Textanalyseprozess zu realisieren. Darüber hinaus sollte die Auswahl mehrerer Dokumente zur Textanalyse ergänzt werden. Im Ergebnis steht damit eine leistungsfähige Textanalyse für die GTME zur Verfügung, die vorhandene Funktionalitäten optimal ergänzt.

Bereits im Vorfeld wurde innerhalb der Projektarbeit „Topic Maps & Ontologien“ ein grundlegender Textanalyseprozess entworfen, der Basis der Entwicklungen war. Er musste im Rahmen der Konzeptionsphase den neuen Anforderungen angepasst werden. Neben der Auswahl von Filterkriterien zum Erstellen einer Dokumentenliste musste der Prozess zur Integration von Associations neu entwickelt werden.

Eine Einarbeitung in die grundlegende Thematik und den Prototypen entfiel durch den Zusammenhang mit dem vorhergehenden Projekt.

Dennoch mussten die einzelnen Bereiche des Textanalyseprozesses neu erarbeitet werden, und den neuen Anforderungen angepasst werden. Aus einer Vielzahl von verschiedenen Ideen mussten die Besten zu einem Gesamtprozess zusammengesetzt werden. Nach dessen Konzeption sollte er innerhalb der Seminararbeit umgesetzt werden.

Die kommenden beiden Kapitel orientieren sich an dem gängigen Prozess der Software-Entwicklung.

Das nächste Kapitel wird auf die Besonderheiten der Konzeptionsphase eingehen. Im Anschluss daran wird sich ein Kapitel mit der Umsetzung des Gesamtprozesses beschäftigen.



5. Konzeption

In diesem Kapitel werden die Ergebnisse der Konzeptionsphase zusammengetragen. Nach der grundsätzlichen Erläuterung des Gesamtprozesses werden die wichtigsten überarbeiteten Einzelprozesse erläutert.

Neben der Möglichkeit zur Auswahl mehrerer Dokumente musste der Kernprozess der Textanalyse überarbeitet werden, um Texte aus mehreren Dokumenten bearbeiten zu können. Die Auswertung der Ergebnisse wurde um das Einfügen von Assoziationen erweitert.

5.1. Gestaltung Ablaufprozess

Der Prozess der Textanalyse folgt einem Schema, das in der vorhergehenden Projektarbeit entworfen wurde und das in seinen Grundstrukturen übernommen werden konnte. Der Knowledge Worker, der die Textanalyse durchführt, wird durch eine Reihe von Dialogen geleitet, die die Übersichtlichkeit des Gesamtprozesses fördern.



Abbildung 3 Gesamtprozess Textanalyse

Nach der Anpassung an die neuen Anforderungen gestaltet sich der Gesamtprozess dabei wie nachfolgend beschrieben:

Im ersten Schritt wird eine Dokumentenliste generiert, auf der die Textanalyse arbeitet. Nach der Auswahl der Content-Datenbank und eines Views kann eine Liste von Dokumenten erzeugt werden. Durch verschiedene Filterkriterien kann diese Liste eingegrenzt werden. Durch Auswahl eines Startdatums wird die Dokumentenliste auf nur neu erstellte oder veränderte Dokumente verkleinert. Weitere Vorselektionen können durch Auswahl einer Mindestanzahl an Wörtern in dem Rich-Text Feld oder die Angabe eines im Dokument enthaltenen Stichwortes getroffen werden. Durch die Angabe einer Formel, die auf den View angewendet wird, kann die Liste zusätzlich verkürzt werden. Nach wie vor ist die Einzelauswahl eines Dokumentes möglich.

Für alle aufgelisteten Dokumente wurde zusätzlich eine Schnellansicht implementiert, anhand derer sich der Knowledge Worker bei Unkenntnis der einzelnen Dokumente orientieren kann. Alle Dokumente der Liste werden zudem in ein Auswahlfeld



K-Discovery:

Konzeption und Implementierung einer Textanalyse

geschrieben. Die zu analysierenden Dokumente können hier markiert und für die Textanalyse übernommen werden.

Im Folgedialog wird der Kernprozess der Textanalyse (siehe auch Kapitel 3.2) durchgeführt. Der Anwender wird über den Status informiert. Wurde die Schlüsselwortextraktion beendet, können aus den gefundenen Schlüsselwörtern (Key Entries) diejenigen selektiert werden, die als Topics in die Datenbank eingepflegt werden sollen. Manuell können der Liste weitere Key Entries hinzugefügt werden. Als ein Teilergebnis des nächsten Dialogs werden die Schlüsselwörter als Topics in der Datenbank hinterlegt. Dazu müssen jedem neuen Topic jeweils Topic-Type und Occurrence-Role (Art des Vorkommens) zugewiesen werden. Sind Topics bereits in der Datenbank vorhanden, können diese um Dokumente ergänzt werden, und die Attribute ggf. angepasst werden. In einem letzten Schritt wird durch das Einfügen von Assoziationen in die Datenbank die Vernetzung generiert, die der Endanwender im Semantic Surfer nutzt.

Dabei werden Associations zwischen zwei Topics manuell eingefügt. Neben der Auswahl vorhandener Regeln können neue Assoziationen angelegt und verwendet werden.

5.2. Dokumentenselektion

Die Auswahl der Dokumente zur Textanalyse musste den gesteigerten Erfordernissen angepasst werden. Aufgrund der Anforderung mehrere Dokumente auszuwählen musste der Selektionsprozess modifiziert werden. Eine Aneinanderreihung einzelner Selektionen von Dokumenten erschien in Bezug auf Effizienz, Effektivität wie auch Benutzerfreundlichkeit wenig zweckmäßig.

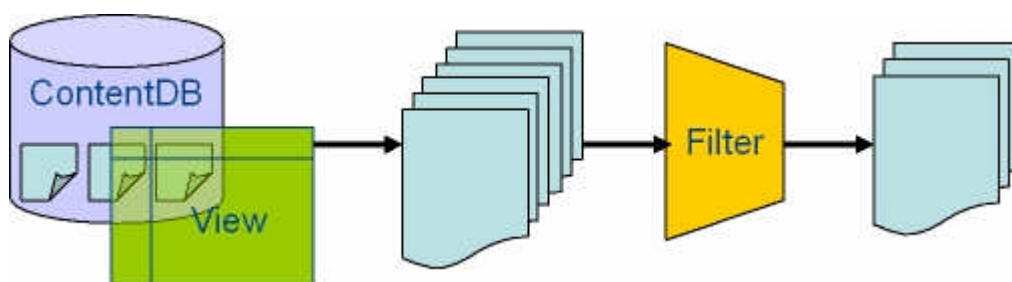


Abbildung 4 Generierung Dokumentenliste

In der Konzeptionsphase wurde nach Absprache festgelegt, automatisch auf Basis des selektierten Views eine Auswahlliste von Dokumenten zu generieren, die sich durch verschiedene Filterkriterien eingrenzen lassen soll. Die entstehende Liste soll durch eine



K-Discovery:

Konzeption und Implementierung einer Textanalyse

dynamische generierte Tabelle, eine Art Kopie des eigentlichen Views, visualisiert werden. Wird kein Filter verwendet werden alle Dokumente des selektierten Views in die Auswahl übernommen. Die bereits implementierte Funktionalität zur Auswahl eines einzelnen Dokumentes aus dem View bleibt zusätzlich erhalten.

Um die Benutzerfreundlichkeit zu steigern, muss die automatisch generierte Auswahlliste manuell abänderbar sein und die Ansicht der aufgeführten Listendokumente zur Verfügung stehen.

Im Ergebnis ist die generierte Tabelle von der eigentlichen Auswahl abgekapselt. Sie dient der Entscheidungsunterstützung bei der Auswahl. Alle in der Tabelle aufgelisteten Dokumente werden ins Auswahlfeld geschrieben und müssen dort für die Textanalyse übernommen werden.

Jedes Dokument wird über seine eindeutige ID identifiziert. Im Rahmen des Kernprozesses werden die Texte der Richt-Text Felder aus jedem ausgewählten Dokument extrahiert.

5.3. Textanalyse

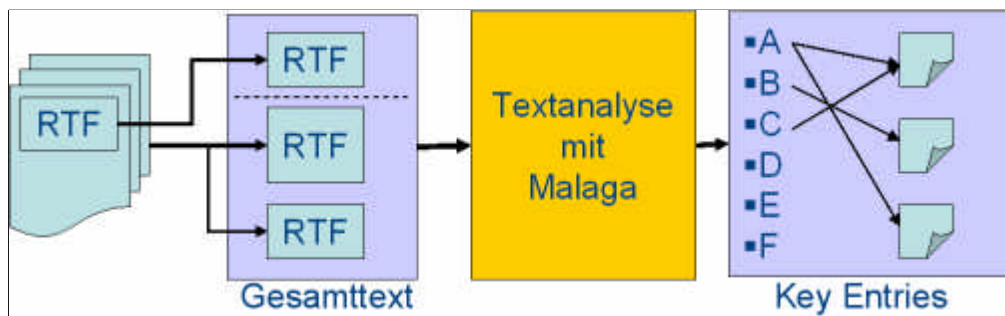


Abbildung 5 Textanalyseprozess mit Malaga

Die einzelnen Texte werden während der Textanalyse zu einem Gesamtext zusammengefasst und von Malaga auf die jeweiligen Stammformen zurückgeführt. Die Zugehörigkeit der Textteile zu den einzelnen Dokumenten wird gespeichert, um im nächsten Schritt das Ableiten der „betroffenen“ Dokumente bei Auswahl eines Key Entries zu ermöglichen.

Die Zusammenfassung in einem Gesamtext erlaubt ein Minimieren der externen Aufrufe auf einen schreibenden und einen lesenden Zugriff pro Textanalyseprozess.

Aus dem Ergebnis der Rückführung auf die Stammformen werden alle identifizierten Substantive eingelesen und nach der Häufigkeit ihres Auftretens ausgezählt. Eine nach Häufigkeit sortierte Liste gefundener Schlüsselwörtern wird an die Benutzeroberfläche



K-Discovery:

Konzeption und Implementierung einer Textanalyse

ausgegeben. Deren Länge kann dabei vorher konfiguriert werden. Die einzelnen Schlüsselwörter werden prozentual zueinander in Beziehung gesetzt. Das ermöglicht dem Knowledge Worker Abstufungen zwischen den gefundenen Schlüsselwörtern zu erkennen. Zur Gesamteinordnung des Ergebnisses wird zusätzlich die Häufigkeit des am meisten gefundenen Schlüsselwortes angegeben, und mit der Gesamtzahl aller im Text befindlichen Substantive verglichen.

5.4. Ergebnisauswertung

Aus der Liste der gefundenen Schlüsselwörter (Key Entries) kann eine Menge von Topics übernommen werden. Die Auswahl kann um manuell erstellte Key Entries ergänzt werden:

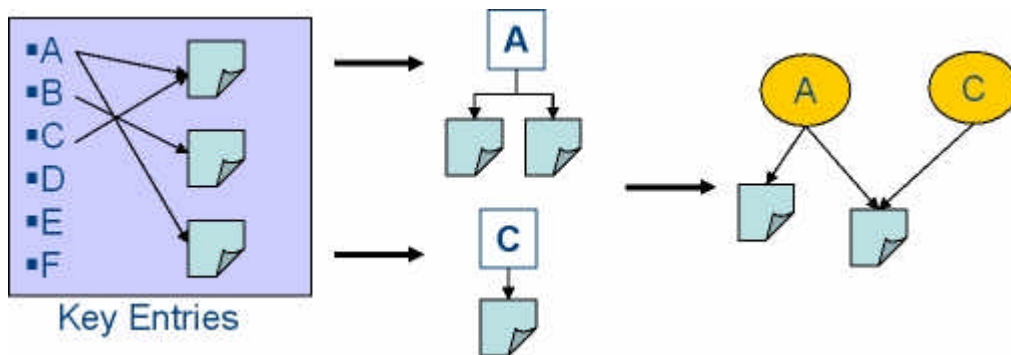


Abbildung 6 Übernahme Key Entries zu Topics

Im nächsten Dialog müssen die neu gefundenen Topics parametrisiert werden, um in die Datenbank übernommen werden zu können. Für jedes neue Topic muss Topic Type sowie Occurrence Role ausgewählt werden. Die Parameter der Topicvorschläge können dabei zunächst auf bereits in der Datenbank vorhandene Werte vorbelegt werden. Hierfür wird die Datenbank nach gleichnamigen Topics durchsucht. Wird ein übereinstimmender Eintrag gefunden, werden dessen Topic Type und Occurrence Role für das neue Topic vorgeschlagen. Diese Funktion soll die Wahl der passenden Parameter für die Topics erleichtern.

Nach diesem Topic Typing genannten Prozess können Associations für jedes der neuen Topics erzeugt werden. Hierfür lassen sich Regeln definieren, wie neue Associations gebildet werden sollen. Diese Regeln (im weiteren Association Orders genannt) ähneln den bisherigen Association Rules in der GTME.

Beim Anlegen einer Association Rule werden zunächst zwei Topic Rules angegeben, die durch ein bestimmendes Association Type miteinander verknüpft werden. Da bei



K-Discovery:

Konzeption und Implementierung einer Textanalyse

Topic Rules letztendlich bestimmt wird, welche Dokumente einem definierten Topic Type zugeordnet werden sollen, werden eigentlich zwei Topic Types durch ein Association Type einander zugeordnet.

Das Erstellen einer Association Order lehnt sich daran an. Für jede Association Order müssen zwei Topic Types und ein Association Type definiert sein. Es ist also zunächst eines der neuen Topics auszuwählen, um eine Association Order zu erstellen. Das Topic Type auf der linken Seite ist bereits durch das Topic Typing festgelegt worden. Nun müssen Association Type und das Topic Type der rechten Seite ausgewählt werden. Es kann ein beliebiges Association Type aus der Konfigurationsdatenbank gewählt werden. Wahlweise kann man sich Association Types aus Associations, die bereits in der Datenbank vorhanden sind, vorschlagen lassen oder neue Association Types über die Standardmaske der GTME anlegen. Abschließend muss das Topic Type der rechten Seite selektiert werden. Die Vorbelegung der Association Type kann übernommen oder durch ein beliebiges Topic Type der GTME ersetzt werden.

Da Associations nicht zwischen Topic Types sondern zwischen Topics gebildet werden (sonst wären alle Topics eines Topic Types miteinander verbunden), müssen für jede Association Order Regeln definiert werden, welche Topics assoziiert werden sollen.

Hierfür wurde sich ebenfalls an der Association Rule orientiert. Dort muss ausgewählt werden, ob die Association nur zu Topics desselben Dokuments oder auch zu anderen Dokumenten gebildet werden soll. Wird die Assoziationenbildung zu anderen Dokumenten erlaubt, muss eine Formel hinterlegt werden, in der die Zuordnungsbedingung definiert wird.

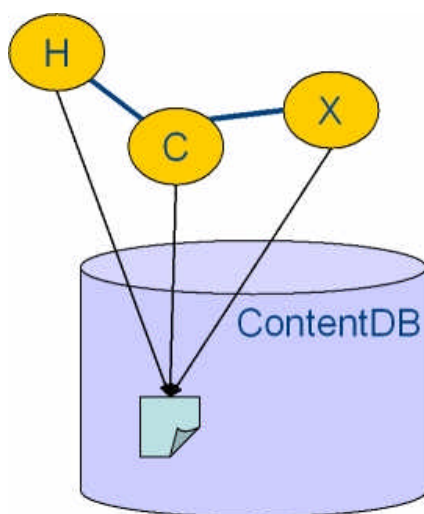


Abbildung 7 Association Order

Beim Aufbau der Association Order wurde ein analoger Weg eingeschlagen. Zunächst bieten sich die Optionen, Assoziationen entweder nur zum gleichen Dokument oder auch zu anderen Dokumenten zuzulassen. Bei der Assoziationenbildung zum gleichen Dokument kann man Assoziationen zu allen anderen Topics dieses Dokuments bilden oder ein bestimmtes Topic selektieren. Es werden dem Knowledge Worker dazu alle zu diesem Dokument verknüpften Topics aufgelistet. Sollten zu mehreren Topics Assoziationen erstellt werden, müssen mehrere Association Order angelegt werden.



K-Discovery:

Konzeption und Implementierung einer Textanalyse

Im anderen Fall werden Associations zu anderen Dokumenten gebildet.

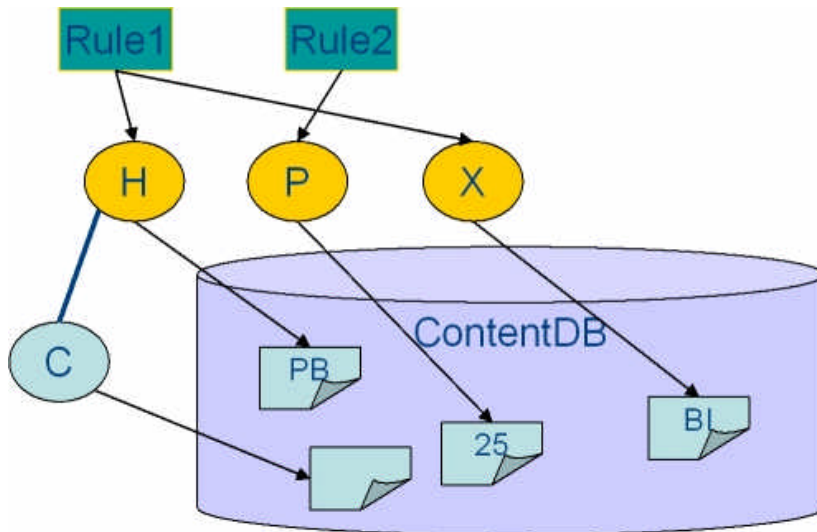


Abbildung 8 Association Order zu anderem Dokument

Hier muss zum Topic Type der rechten Seite eine Topic Rule aus der GTME gewählt werden, auf der die gebildeten Topics basieren. Die Suche nach der Kombination aus Topic Type und der dem Dokument zugeordneten Form, die aus dieser Topic Rule entstanden sind, berücksichtigt auch durch Malaga erstellte Topics. Nach der Auswahl der Topic Rule muss analog zur Association Rule eine Formel angegeben werden, die die Bedingung für die Assoziationenbildung definiert.

Diese Formel wird auf jedes zu der Topic Rule gehörende Dokument angewandt. Wird die Formel zum Wert ‚wahr‘ ausgewertet, so wird zu diesem Topic (bzw. dem zugehörigen Dokument) eine Assoziation erstellt.



6. Implementierung

Dieses Kapitel beschreibt die technische Umsetzung der Konzeptionsphase und die Integration in die GTME. Analog zur Gliederung des vorherigen Kapitels wird zunächst der Gesamtprozess betrachtet, danach werden die Einzelprozesse detaillierter betrachtet.

6.1. Gestaltung Ablaufprozess

In diesem Abschnitt wird eine Übersicht gegeben, wie die für die GUI neu erstellten Design-Elemente im Zusammenhang stehen.

Beim Aufruf der Topic Selection aus dem Default View oder des Action-Menüs, wird das Frameset „TopicSelection“ aufgerufen. Dieses Frameset bildet die Rahmenstruktur des Gesamtablaufs. Es beinhaltet auf der linken Seite eine Statusleiste, die zur Orientierung den aktuellen Schritt der gesamten Textanalyse anzeigt. Diese Statusleiste ist statisch in das Frameset eingebunden, wobei sich der rechte Frame über das Setzen von Variablen in der Session mit beliebigen Inhalten füllen lässt.

Hierzu wurden die Session-Variablen „MGSelectedType“ und „MGSelectedElement“ eingeführt, welche die Art des anzuzeigenden Design-Elements und das genaue Element festlegen.

Zu Beginn wird im rechten Frame die Page „TopicSelectionIntro“ angezeigt, die eine einführende Übersicht der Einzelschritte des Gesamtprozesses zeigt.

Der nächste Schritt dient der Selektion der Dokumente, die in die Textanalyse mit einbezogen werden sollen. Hierbei wird das Frameset „TopicSelectionProcess“ im in den rechten Frame geladen. Dieses besteht aus zwei Frames, von denen der obere mit der Form „TopicSelectionSelectDB“ und der untere mit der Page „TopicSelectionBackground“ vorbelegt ist. Die genaue Funktionsweise der Dokumentenselektion wird im Abschnitt ‚Dokumentenselektion‘ ausführlicher besprochen.

Nach der Dokumentenselektion folgt die Durchführung der Textanalyse. Hierfür wird die Form „TopicSelectionSearch“ ins Frameset geladen.

Nach der eigentlichen Textanalyse werden die Ergebnisse mit dem Frameset „TopicSelectionRes“ angezeigt, welches im oberen Frame die Form „TopicSelectionResult“ und im unteren wieder die Background Page anzeigt.

Im letzten Schritt wird die Form „TopicSelectionTopic“ angezeigt, mit der der Ablauf der Textanalyse endet.



K-Discovery:

Konzeption und Implementierung einer Textanalyse

6.2. Dokumentenselektion

Die Dokumentenselektion wird durch den Benutzer durch die Form „TopicSelectionSelectDB“ durchgeführt. Bei der Auswahl der Datenbank wird auf vorhandene Database Connections der Konfigurationsdatenbank zugegriffen. Von der ausgewählten Datenbank werden die Namen der Views exclusive der versteckten Views (mit ‚(...)‘) ausgelesen. Die Dokumentenmenge eines aus dieser Liste selektierten Views stellt die Basisdokumentenmenge dar. Entweder kann nun über eine Picklist ein einzelnes Dokument aus diesem View gewählt werden, oder es werden Filter definiert, die die Dokumentenmenge verkleinern können.

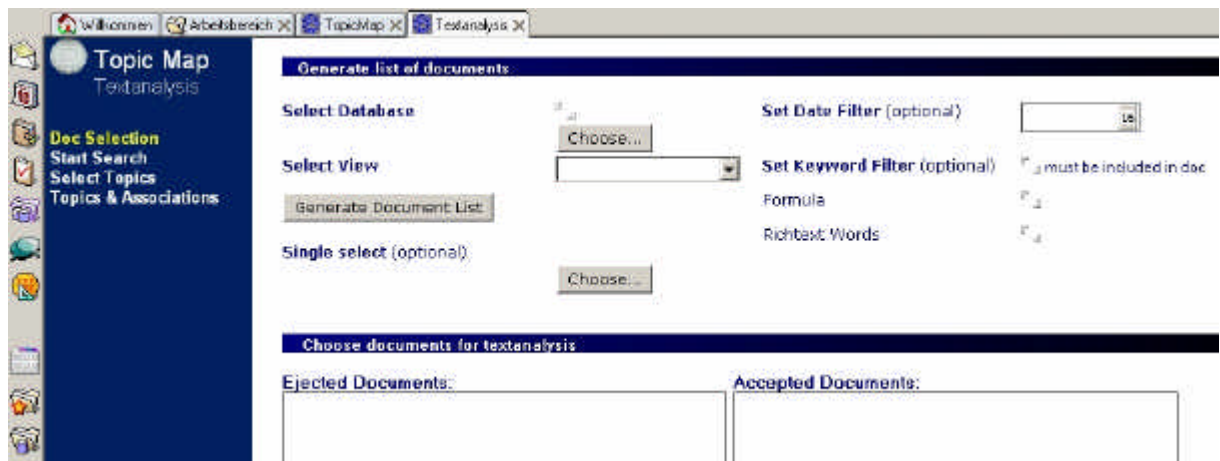


Abbildung 9 Screenshot Dokumentenliste generieren

Es existieren folgende Filter:

Datum

Ein Datum kann über einen Date/Time Auswahldialog gewählt werden oder per Direkteingabe bestimmt werden. Nach einer Überprüfung, ob es sich bei dem eingegebenen Wert um ein Datum handelt, wird für jedes Dokument überprüft, ob dessen Attribut „LastModified“ (Zeitpunkt der letzten Änderung) kleiner als das angegebene Datum ist. Damit werden alle Dokumente älteren Datums aus der Dokumentenmenge entfernt.

Volltextsuche

Im Backend wird die Dokumentenmenge mit Hilfe einer NotesViewEntryCollection verwaltet. Diese bietet die Möglichkeit einer Volltextsuche auf der Dokumentenmenge. Diese wird um alle Dokumente verringert, die nicht den gesuchten Begriff beinhalten.

Anzahl der Worte in den Rich-Text Feldern



K-Discovery:

Konzeption und Implementierung einer Textanalyse

Alle Einträge im Property Items der Dokumente werden getestet, ob sie vom Typ Rich-Text sind. Aus allen Felder dieses Typs gefunden wird der Text ausgelesen. Aus dem so entstandenen Gesamttext werden die Wörter ausgezählt. Wird die geforderte Mindestwortanzahl unterschritten, wird das Dokument aus der Dokumentenmenge entfernt.

Formelausdruck

Über die Definition einer Formel kann auf sehr individuelle Weise die Dokumentenmenge eingegrenzt werden. Die Formel wird mit der Evaluate-Anweisung auf jedes Dokument angewandt. Ergibt die Formel den Wert ‚falsch‘ wird das Dokument aus der Menge herausgenommen.

Nachdem die Filtereinstellungen vom Knowledge Worker gesetzt wurden, beginnt der Selektionsprozess. Dabei werden die Datenbank-, View- und Filterinformationen über ein Profildokument an den Server übertragen. Hier wird die entsprechende Datenbank in der Methode createSelectionTable() der Klasse MalagaAction geladen und der View geöffnet. Die ausgelesene NotesViewEntryCollection wird durch die Filter verringert, so dass eine Dokumentenmenge von maximal 50 Einträgen übrig bleibt. Für diese Dokumentenmenge wird eine Tabelle in einem Rich-Text Feld angelegt, in die die ersten 6 Spalten des Views eingelesen werden. Die genannten Beschränkungen sind nötig, damit die angelegte Tabelle nicht zu groß und unübersichtlich wird.

Die eindeutigen IDs (UNIDs) der Restmenge und das neu erstellte Rich-Text Feld werden per Profildokument wieder an den Client übertragen, so dass dieses lediglich in das Anzeige-Dokument („TopicSelectionDocTable“) kopiert werden braucht.

6.3. Textanalyse

In diesem Abschnitt wird die Erweiterung der Textanalysefunktion besprochen. Die grundlegende Funktionsweise der Malaga Textanalyse ist nachzulesen in der Projektdokumentation des Projektes „Topic Maps und Ontologien“.

Der Aufruf der Malaga-Engine wurde in der vorherigen Version über die Klasse SendKeys verwirklicht, die die Eingaben direkt in die Malaga-Eingabeaufforderung gesendet hat. Jetzt wird Malaga als Filter für die Eingabeaufforderung benutzt, was zum Vorteil hat, das die Textanalyse auch funktioniert, wenn niemand am Server eingeloggt ist und Domino als Dienst läuft. Während der Analyse durch Malaga wird die Funktion canOpenFile() der Klasse MalagaAction benutzt, um festzustellen, ob Malaga mit dem Schreibvorgang auf die Ausgabedatei fertig ist. Dieser Funktion wird ein Timeout-Wert



K-Discovery:

Konzeption und Implementierung einer Textanalyse

übergeben, nach dessen Überschreitung die Funktion abgebrochen wird. Dies beendet den Agent im Fehlerfall nach einer definierten Zeit.

In der bisherigen Textanalyse konnte nur ein Dokument gleichzeitig analysiert werden. Nun wird eine Dokumentenmenge analysiert, wobei die Einzeltex te aller Rich-Text Feldern der Dokumentenliste zusammengenommen werden, um eine Meta-Textanalyse durchzuführen. Hierzu werden die UNIDs aller Dokumente aus dem Selektionsprozess an den Server übertragen. In der Methode `extractKeywords()` der Klasse `MalagaAction` werden nun alle Dokumente geöffnet und der Text in einem String (Begrenzung auf 2GB) gesammelt. Diese Textmenge wird wie bisher mit der Methode `rmNonAlphabeticals()` der Klasse `MalagaAction` von unerwünschten Zeichen bereinigt, so dass nur eine Wortliste übrig bleibt. Neu bei diesem Prozess ist, dass für jedes Dokument die Position im Gesamttext gespeichert wird. Bei der Schlüsselwortermittlung können später sowohl die Schlüsselwörter für den Gesamttext generiert werden, als auch die Häufigkeit der gewonnenen Schlüsselwörter für jedes Dokument hinterlegt werden. Die Schlüsselwörter und deren Vorkommenshäufigkeit in den einzelnen Dokumenten der Gesamtliste werden zurück an den Client gesendet. Nach der Auswahl der gewünschten Topics, werden nur die Dokumente, die über einen Vergleich mit der hinterlegten Worthäufigkeitsmatrix auch tatsächlich das Schlüsselwort enthalten, im nächsten Schritt in das Topic Typing einbezogen.

6.4. Ergebnisauswertung

Der Schritt für das Topic Typing und die Festlegung der Association Orders wurde in einer komplexen Oberfläche realisiert, die sich in vielen Bereichen nachträglich konfigurieren lässt.

Für das Topic Typing dient ein Textfeld für die Anzeige der neuen Topics und deren Topic Type und Occurrence Role. Sowohl der Standardwert für Topic Type und Occurrence Role lassen sich definieren, wie auch die prä- und postfixe der beiden Werte. Der Standardwert für die erfolglose Datenbanksuche nach bereits existierenden Belegungen für die Topics lässt sich ändern. Diese Einstellungen lassen sich für die Topic Types über die Felder `preTT`, `afterTT`, `defaultTT`, `defaultDBTT` und für die Occurrence Roles über `preOT`, `afterOT`, `defaultOP`, `defaultDBOT` ändern.

Für die Konfiguration der Standardwerte der Association Orders sind folgende Felder verantwortlich:

`showDefaultAT`: Der angezeigte Wert vor der Wahl des Association Types

`showDefaultRTT`: Der angezeigte Wert vor der Wahl des rechten Topic Types



K-Discovery:

Konzeption und Implementierung einer Textanalyse

defaultAssocTypeSelect: Definition, ob die Assoziationenbildung zu einem Dokument oder zu mehreren Dokumenten als Standard zu setzen ist. Wert „1“ steht für die erste, Wert „2“ für die zweite Möglichkeit

showDefaultTTSelection: Hier wird festgelegt, was angezeigt werden soll, wenn Assoziationen zu allen Topics dieses Dokuments angelegt werden sollen, aber kein spezifisches Topic ausgewählt wurde.

showDefaultTR: Festlegung, was vor der Wahl einer TopicRule angezeigt werden soll.

showAssocTypeSelect1: Festlegung, was angezeigt werden soll, wenn Assoziationen nur zu einem Dokument eingestellt wurde.

showAssocTypeSelect2: Festlegung, was angezeigt werden soll, wenn Assoziationen zu anderen Dokumenten zugelassen sind.

showDefaultFormula: Festlegung, was vor der Wahl einer Formel angezeigt werden soll.

Beim Anlegen von Association Orders mit Aktivierung der Checkbox „Check Associations“ wird in der relationalen Datenbank nach bereits vorhandenen Associations für das linke Topic Type der Association Order gesucht. Technisch wird eine SQL-Query auf der Tabelle „Assoc“ ausgeführt, die alle Tupel zurückgibt, die dieses Topic Type beinhalten. Der Benutzer kann dann mittels einer Picklist auf die vorgeschlagenen Werte zugreifen.

Wird ein Topic zum selben Dokument ausgewählt, werden die Selektionsmöglichkeiten aus der Tabelle „Topic“ bezogen.

Durch die Betätigung des Buttons „Write all to RDBMS“ wird ein komplexer Vorgang aufgerufen, der die definierten Topic Types und Occurrence Roles für die Topics ins RDBMS schreibt. Darüber hinaus werden die Association Orders ausgeführt und entsprechende Associations im RDBMS gebildet.

Zuerst werden die Daten mittels eines Profildokuments an den Server gesendet. Hier wird in der Klasse MalagaAction lediglich die Methode write() aufgerufen, die sich eine Instanz der Klasse ManualInsert anlegt und sich selbst als Referenz übergibt. Nun wird von dieser neuen Instanz die Methode writeManualRDBMS() aufgerufen, die sich in zwei Abschnitte gliedert. Zuerst werden die neu angelegten Topics bearbeitet, welche über den Aufruf von writeRDBSMManualTopic() verarbeitet werden. Wichtig ist, dass alle eingefügten Topics in dem Array insertedTopics zwischengespeichert werden.

Im zweiten Abschnitt werden nacheinander die Association Orders (gespeichert im Array AssocList) ausgeführt. Die zuerst aufgerufene Methode writeRDBSMManualAssoc() zerlegt die aktuelle Association Order in ihre Komponenten und ruft danach mit den



K-Discovery:

Konzeption und Implementierung einer Textanalyse

entsprechenden Parametern entweder `insertAssocForDoc()` oder `insertAssocForTopicRule()` auf. Wenn sich die Association Order auf dasselbe Dokument bezieht, wird die erste Methode verwendet. Wird der Methode der Wert „*all*“ übergeben, werden zu allen Topics dieses Dokuments Assoziationen gebildet. Dazu wird die Methode `searchTopicIDsForDoc()` der Klasse `MalagaAction` benutzt, die alle Topics zu einem bestimmten Dokument zurückgibt.

Wird der Methode hingegen ein bestimmtes Topic übergeben, so wird folglich nur zu diesem einen Topic eine Association erstellt.

Mit der Methode `insertAssocForTopicRule()` werden Association Orders behandelt, bei denen eine Topic Rule ausgewählt wurde. Diese ermittelt aus dem Topic Rule Dokument die erforderlichen Informationen für die Auffindung der Topics zur übergebenen Topic Rule.

In jedem Falle werden bei der Bildung der Assoziationen nur Werte in die Tabellen `AssocDataA` und `AssocDataB` geschrieben. Die eigentlichen Assoziationen werden wie beim bisherigen Assoziationsbildungsprozess über einen Join bezüglich dieser beiden Relationen gebildet.

Mit dem Einfügen der neu erstellten Topics in das RDBMS und die Integration der neuen Associations über die Association Orders in die Topic Map ist der Gesamtprozess der Textanalyse abgeschlossen.

Für den Endanwender stehen ab diesem Zeitpunkt die neuen Topics und Associations in seiner Benutzeroberfläche „Semantic Surfer“ zur Verfügung.



7. Zusammenfassung und Ausblick

Wie die vorherigen Kapitel gezeigt haben wurde innerhalb der Seminararbeit die Arbeit des vorherigen Knowledge Management Projektes „Topic Maps & Ontologien“ fortgeführt und erweitert. Durch die Integration von Associations und der Möglichkeit der Nutzung von Dokumentenlisten zur Textanalyse wurde ein leistungsfähiger Baustein innerhalb der Groupware-based Topic Map Engine ergänzt, mit dessen Hilfe der Aufbau der Topic Map weiter unterstützt werden kann.

Neben der dynamischen Generierung von Topics aus den eigentlichen Inhalten der Dokumente (Rich-Text Felder) können diese durch Angabe von Associations in die Topic Map eingepflegt werden und stehen dem Endanwender zur Verfügung. Die Textanalyse ist Teil des Strukturaufbaus der Topic Map und wird von einem Knowledge Worker durchgeführt. Die Oberfläche wurde deshalb innerhalb der Lotus Notes Anwendung integriert und konnte unabhängig von der Oberfläche der Endanwender konzipiert und umgesetzt werden.

Zur Steigerung der Benutzerfreundlichkeit und besseren Übersicht des Gesamtprozesses wird der Knowledge Worker während der Textanalyse durch einzelne Dialoge geleitet.

In der Implementierung wurde bewusst darauf geachtet, einzelne Bausteine der Textanalyse möglichst leicht wartbar und austauschbar zu implementieren. In zukünftigen Projekten könnten also einzelne Bausteine des Gesamtprozesses verfeinert oder durch ergänzende Methoden ersetzt werden.

Eine Erweiterung um Unterfunktionalitäten ist somit grundsätzlich denkbar, war aber nicht mehr Bestandteil der Aufgabenstellung.

Mit der Integration der automatischen Textanalyse aus Rich-Text Feldern der Dokumente kann die Topic Map in ihrem inhaltlichen Aufbau verbessert werden, ohne das für den Endanwender ein Mehraufwand entstände. Ihm kommen jedoch die dadurch verbesserten Suchergebnisse zu Gute.



Anhang A: Abbildungsverzeichnis

Abbildung 1 Topic Map.....	1
Abbildung 2 Struktur GTME	4
Abbildung 3 Gesamtprozess Textanalyse	8
Abbildung 4 Generierung Dokumentenliste	9
Abbildung 5 Textanalyseprozess mit Malaga	10
Abbildung 6 Übernahme Key Entries zu Topics.....	11
Abbildung 8 Association Order zu anderem Dokument	13
Abbildung 9 Screenshot Dokumentenliste generieren	15



Anhang B – Auflistung modifizierter und hinzugefügter Design Elemente

Framesets:

- TopicSelection
Übergeordnetes Frameset, welches über den gesamten Textanalyseprozess verwendet wird.
- TopicSelectionProcess
Dient der Zweiteilung bei der Dokumentenselektion in Konfigurations- und Anzeigebereich.
- TopicSelectionRes
Dient wie TopicSelectionProcess der Zweiteilung des Anzeigebereichs im Abschnitt der Schlüsselwortauswahl.

Pages:

- TopicSelectionBackground
Wird benötigt als Hintergrund bei den Framesets TopicSelectionProcess und TopicSelectionRes
- TopicSelectionIntro
Dieses Element bildet die Einstiegsseite mit den Hinweisen für den Benutzer
- TopicSelectionNav
Diese Page wird den gesamten Ablauf über auf der linken Seite des Framesets TopicSelection angezeigt.

Forms:

Modifizierung von TopicType & AssociationType

Die Ereignisse PostSave und QueryClose wurden benutzt, um das neu erstellte Topic Type sofort in die Datenbank zu schreiben, so fern die Form aus dem Schritt Topic Typing der Textanalyse aufgerufen wurde.

- TopicSelectionResult
Form zur Anzeige der gefundenen Schlüsselwörter im Frameset TopicSelectionRes
- TopicSelectionSearch
Form für die Ausführung der Malaga Textanalyse
- TopicSelectionSelectDB
Form zur Konfiguration der Dokumentenselektion im Frameset TopicSelectionProcess



K-Discovery:

Konzeption und Implementierung einer Textanalyse

- TopicSelectionTopic
Mit dieser Form kann das Topic Typing und die Erstellung der Association Orders durchgeführt werden.
- (Malaga Configuration)
Form für die Konfiguration der Malaga-Engine. Kann aufgerufen werden über die Administration Action des ALL Views.
- (MalagaClient2ServerTemp) & (MalagaServer2ClientTemp)
Diese Forms dienen der Informationsübertragung zum Server/Client.
- (Malaga_Data_Doc)
Diese Form dient der Informationsübertragung zwischen den einzelnen Schritten des Textanalysprozesses.
- (TopicSelectionDocTable)
Diese Form dient lediglich der Anzeige der aufbereiteten Rich-Text Felder im Schritt der Dokumenten- und Schlüsselwortauswahl.

Views

- ALL (modifiziert)
Es wurden im Action Menu die Actions: Administration/Configure Malaga und Start Topic Selection hinzugefügt.

Agents

- MalagaServerAction
Dieser Agent führt alle Aufgaben, die auf dem Server durchgeführt werden müssen, aus. Hier wird immer eine Instanz der Klasse MalagaAction aus der Script Library TMMalagaAction instanziiert, und die entsprechende Methode aufgerufen. Zur Informationsübermittlung zwischen Client und Server werden die Forms (MalagaClient2ServerTemp) und (MalagaServer2ClientTemp) verwendet.

Script Libraries

- TMLog – Class ActivityLog (modifiziert)
Public Sub delete() (modifiziert)
Funktion überprüft nun, ob sich es beim aktuellen Log um ein Log Doc handelt und ruft die Funktion saveLogDocId() auf, damit die ID des Log Docs gespeichert wird.



K-Discovery:

Konzeption und Implementierung einer Textanalyse

Public Sub writeEndTime()

Schreibt nach Fertigstellung eines Malaga Logs die Endzeit in das Dokument.

Public Sub saveLogDocId()

Speichert nach Durchführung einer der Malaga Funktionen das Log Dokument.

Public Sub setMalagaLog()

Deklariert das aktuelle Log Objekt als Malaga Log.

Public Sub openLogById()

Öffnet das zuletzt geöffnete Malga Log Doc, dessen ID im (Malaga Configuration) Dokument gespeichert wurde.

- TMMalagaAction – Class MalagaAction As ActivityLog

Sub new(): Konstruktor

Sub updateTopicType()

Fügt ein neu erstelltes Topic Type sofort in das RDBMS ein.

Sub updateAssocType()

Fügt ein neu erstelltes Association Type sofort in das RDBMS ein.

Sub createSelectionTable()

Erstellt die Richt-Text Tabelle für die Dokumentenselektion.

Sub createResultTableString()

Erstellt die Rich-Text Tabelle für beim Auswählen der Schlüsselwörter.

Sub checkForNewLogDoc()

Überprüft, ob die maximal angegebene gröÙe für die MalagaLogs überschritten wurde.

Function getSQLConnection() As SQLOperations

Wird benutzt, um eine SQL-Connection zu erhalten.



K-Discovery:

Konzeption und Implementierung einer Textanalyse

Sub extractKeywords()

Extrahiert die Schlüsselworte aus den übergebenen Dokumenten der übergebenen Doc-Ids.

Sub canOpenFile()

Diese Funktion blockiert so lange, bis sich die angegebene Datei öffnen lässt.

Function rmNonAlphabeticals()

Entfernt nicht erwünschte Zeichen aus den Texten der Rich-Text Felder.

Sub searchDBTopics()

Sucht im RDBMS nach schon vorhandenen Parametern für die neuen Topics.

Sub write()

Führt mittels der Klasse Manual Insert den Schreibvorgang der neuen Elemente in das RDBMS aus.

Sub searchTTForDoc()

Gibt die Topic Types für ein bestimmtes Dokument zurück.

Function searchTopicIDsForDoc()

Gibt die Topics für ein bestimmtes Dokument zurück.

Sub searchAssocsForTT()

Sucht nach schon vorhandenen Associations für ein bestimmtes Topic Type.

Sub startMalagaLog()

Beginnt ein neues Malaga Log Doc.

Sub endMalagaLog()

Beendet ein Log Doc, wenn die zulässige Größe überschritten wurde.

Function checkLogDoc()

Überprüft die Gültigkeit des Log Docs.



K-Discovery:

Konzeption und Implementierung einer Textanalyse

- TMMalagaAction – Class ManualInsert As ActivityLog

Sub new()

Konstruktor

Function writeManualRDBMS()

Wird zu Beginn aufgerufen und steuert den restlichen Schreibprozess in das RDBMS.

Function writeRDBMSManualTopic()

Diese Funktion dient dem sofortigen Schreiben der neu definierten Topics.

Function getFormNameForDocID()

Wird benötigt, um festzustellen, welchen Formnamen ein Dokument hat.

Function writeRDBMSManualAssoc()

Koordiniert den Einfügeprozess der Associations.

Function insertAssocForDoc()

Erstellt Associations, die sich auf dasselbe Dokument beziehen.

Function insertAssocForTopicRule()

Erstellt Associations, die mit Hilfe einer Topic Rule definiert wurden.