

Ontologien im Trend Mining

Lars Wißler and Olga Streibel

Freie Universität Berlin
Department of Mathematics and Computer Science
Corporate Semantic Web and Networked Information Systems
Group



GEFÖRDERT VOM



Ontologien für wissensbasierte Trendanalysen

Lars Wißler, Olga Streibel
Freie Universität Berlin
Department of Mathematics and Computer Science
Corporate Semantic Web and
Networked Information Systems Group
Königin-Luise-Str. 24-26
14195 Berlin, Germany
jahftw@googlemail.com
streibel@inf.fu-berlin.de

Technical Report TR-B-12-07

12. Dezember 2012

Zusammenfassung

Der Report behandelt das größtenteils unerforschte Feld der ontologiegestützten, wissensbasierten Trenderkennung durch Textanalyse mit dem Schwerpunkt der Entwicklung von Trendontologien. Die Schwierigkeiten der Trenderkennung durch Textanalyse liegen in der mehrdeutigen Semantik natürlicher Sprachen, ihren verschiedenen Formen und Besonderheiten, sowie ihrer Dynamik, wodurch Probleme in der eindeutigen und statischen Formalisierung des in der Trenderkennung benutzten Wissens entstehen. Durch Ontologien können Sprachkomponenten identifiziert und in Relation zu einander verarbeitet und ausgewertet werden. Auf Grund von verschiedenen Sprachen und den Besonderheiten in ihrer Nutzung hinsichtlich Anwender und Anwendungsbereich, wie auch unterschiedlichem Trendverhalten verschiedener Bereiche, werden allerdings je nach Anwendung verschiedene spezialisierte Trendontologien benötigt. Um die modulare Entwicklung und Nutzung dieser verschiedenen Ontologien zu ermöglichen, wird eine standardisierende Basis in Form einer Metaontologie für die Trenderkennung benötigt. Die Entwicklung dieser Metaontologie ist der zentrale Aspekt des Reports.

Abstract

The report deals with the largely unexplored field of ontology-driven, knowledge based trend detection by means of text mining, focusing on the development of trend ontologies. The difficulties of trend detection with text mining lie in the ambiguous semantics of natural languages and their various forms, characteristics and dynamics. Due to this it is difficult to formalize knowledge used in trend detection unambiguously and statically. Using ontologies, language components can be identified and subsequently processed and analyzed regarding their relations to each other. However, due to different languages and specific usages depending on user and application fields, as well as specific trend behavior in certain application fields, trend ontologies specialized for the intended application are needed. In order to allow the modular development and usage of these different ontologies a standardizing base for trend ontologies is needed. This base can be realized as meta ontology and its development is the central aspect of the report.

Inhaltsverzeichnis

1	Einleitung	3
1.1	Was sind Ontologien	4
1.2	Ziel der Ontologie	5
1.3	Relevante Ontologien	7
1.3.1	Simple Knowledge Organization System - SKOS	7
1.3.2	OWL-Time	9
2	Entwicklung der Metaontologie	12
2.1	TREMA Ontologien	13
2.2	Metaontologie	15
2.2.1	Konzeptionelle Anforderungen	15
2.2.2	Allgemeine Beschreibung	16
2.2.3	Klassen und Relationen	18
2.2.4	Funktionalität	20
3	Anwendung der Metaontologie	22
3.1	Anwendung für Aktienanalysen	22
3.1.1	Klassen	23
3.1.2	Themenbereiche	23
4	Experimente und Evaluierung	25
4.1	Evaluierung hinsichtlich Kompetenzfragen	25
4.2	Vergleich von Prognose und Kursentwicklung	30
4.2.1	Auswahl der Indikatoren	30
4.2.2	Berechnung der Kennzahlen	30
4.2.3	Vergleich der Kursentwicklung	31
5	Fazit und Ausblick	35
5.1	Möglichkeiten	35
5.2	Probleme	36
5.3	Fazit	36
A	Ontology Specification Requirements Document	40
A.1	Zweck	40
A.2	Anwendungsbereich	40
A.3	Grad der Formalisierung	40
A.4	Zielgruppen	40
A.5	Vorgesehene Nutzung	41
A.6	Kompetenzfragen	41

A.7 Glossar der Begriffe	43
B Kompetenzfragen und korrespondierende Abfrage	44
C Aktienkennzahlen und Berechnung	47

Kapitel 1

Einleitung

Trend: "A general direction in which something is developing or changing." [5]

Durch das massive Fortschreiten der Digitalisierung von Daten und das Anhäufen gewaltiger Datenmengen gewinnt die automatisierte maschinelle Extraktion von Wissen insgesamt stetig an Bedeutung. Das manuelle Durchsehen derartiger Mengen von Informationen ist zeit- und arbeitsintensiv. Eben diese Menge an Informationen birgt aber gleichzeitig die Gelegenheit für neue Anwendungen in der automatischen Wissensextraktion, wie zum Beispiel die automatische Trenderkennung. Im Zuge der Globalisierung ist es heute wichtiger denn je mit aktuellen Entwicklungen oder Trends Schritt zu halten. Gleichzeitig werden Entwicklungen vielfältiger und ihr Fortschreiten schneller.

Das Erkennen von aktuellen Entwicklungen und die Prognosen von zukünftige Entwicklungen ist traditionell hauptsächlich die Aufgabe von Marktforschungsinstituten, Wirtschaftsforschungseinrichtungen und statistischen Ämtern. Diese sind jedoch bei extremen Impulsen und den damit einhergehenden schnellen Veränderungen häufig Überfordert. Auf Grund der Geschwindigkeit, mit der sich die negativen Impulse der letzten Wirtschaftskrise über die Welt verbreiteten, haben sich Prognosen als schwierig erwiesen.¹ Die traditionelle Konjunkturforschung wie auch die amtliche Statistik waren mit diesen Entwicklungen überfordert, weshalb mehr als sonst auf weiche Indikatorsysteme wie Stimmungsumfragen zurückgegriffen wurde.[1] Informationsquellen sind in der Regel für die Nutzung und vor allem das Verständnis durch Menschen gedacht und dementsprechend aufbereitet; sei es ein Artikel, ein Blogeintrag oder auch ein Buch. Dementsprechend schwer ist die maschinelle Verarbeitung solcher Quellen. Der Mehrwert einer automatischen Trenderkennung auf Basis von Textanalyse ist jedoch groß, denn manuell lassen sich die verfügbaren Mengen an Daten kaum bearbeiten und die Auswertung von Kennzahlen reicht häufig nicht aus.

Diese Problematik tritt besonders in Zeiten schneller Entwicklungen hervor, da in diesen Zeiten die Menge an relevanten Informationen größer und der Zeitrahmen knapper ist. Hinzu kommt das gerade aktuelle Informationen hauptsächlich in Einträgen und Artikeln zu finden sind. Weiterhin ist Wissen über

¹“Schon zu normalen Zeiten sind Prognosen für das Folgejahr wenig treffsicher, wenn sie früh im Jahr gemacht werden.” [9].“Außerdem gibt es in der gegenwärtigen Wirtschaftskrise besondere Prognoseprobleme. Auch können Stimmungen die Realität beeinflussen und Prognosen deshalb Krisen verstärken.” [19]; [20].

aufkommende Trends in vielerlei Hinsicht interessant und hilfreich, sei es in der Marktforschung, um Absatzzahlen von Produkten durch Anpassung an aktuelle Trends zu erhöhen oder um menschliche Verhaltensweisen durch Einbettung in aktuelle Trends vorherzusagen. Trends weisen auf wahrscheinliche zukünftige Entwicklungen hin und Wissen über die Zukunft ist häufig äußerst wertvoll. Gerade beim Börsenhandel sind zuverlässige Aussagen über Markthaltung und Stimmung von Investoren höchst interessant und lukrativ, da Aktienkurse eng mit der Stimmung von Marktteilnehmern verknüpft sind.

Die Schwierigkeit einer auf Texten basierenden automatischen Trenderkennung liegt in der maschinellen Extraktion des Wissens aus Texten, die für Menschen als LeserInnen gedacht sind. Es ist schwer mittels Computer natürliche, also menschliche Sprachen so auszuwerten, dass sich der Inhalt der Nachricht nicht ändert. Man denke dabei bspw. an das Übersetzungsprogramm von Google.² Somit stellt sich die Frage, wie sich eine maschinelle Trenderkennung umsetzen lässt und welche Erkenntnisse sich aus den Ergebnissen erschließen. Der in dieser Arbeit verfolgte Lösungsansatz beschäftigt sich mit der Nutzung von Ontologien, um Computern ein für die Trenderkennung ausreichendes Verständnis der menschlichen Sprache zu ermöglichen. Hierbei wird ein Trend als ein aufkommendes oder zurückgehendes Thema im Web betrachtet und als solches behandelt. Aus den Entwicklungen dessen, was im Internet vorrangig thematisiert wird, lassen sich aussagekräftige Informationen über den aktuellen Stand und mögliche zukünftige Entwicklungen gewinnen.

“Internetdaten liefern eine interessante, bisher praktisch ungenutzte Datengrundlage.³ Sie sind rasch und umfangreich verfügbar und reagieren flexibel auf Änderungen der Rahmenbedingungen.” [2]

Ausgehend von der Konzipierung von Trendontologien in [13], entwickelt an der Freien Universität Berlin im Rahmen des TREMA-Projektes⁴, wird eine Metaontologie als Wissensbasis für die automatisierte Trenderkennung entwickelt, die als standardisierende Basis für speziellere, domänenspezifische Trendontologien fungiert. Im weiteren Verlauf dieses Kapitels wird der Begriff der Ontologie sowie das Ziel der Metaontologie erläutert und anschließend relevante Ontologien analysiert und diskutiert. Kapitel 2 erläutert die entwickelte Metaontologie und Kapitel 3 ihre Anwendung auf deutschsprachige Aktienanalysen. Anschließend werden in Kapitel 4 die Ergebnisse der Evaluierung dargestellt, um in Kapitel 5 ein Fazit aus den gewonnen Erkenntnissen zu ziehen und einen Ausblick auf weiterführende Möglichkeiten zu gewähren.

1.1 Was sind Ontologien

“...in its most prevalent use in AI, an ontology refers to an engineering artifact, constituted by specific vocabulary used to describe a certain reality, plus a set explicit assumptions regarding the intended meaning of the vocabulary words. This set of assumptions has usually the form of a first order logical theory, where

²<http://translate.google.de> (8/2012)

³Diese Datenquelle wurde bereits im Zusammenhang mit den amerikanischen Präsidentschaftswahlen in [3] beziehungsweise mit Grippeepidemien in [6] ausgewertet.

⁴TREMA: Trend Mining, Analysis and Fusion of Multimodal Data

vocabulary words appear as unary or binary predicate names, respectively called concepts and relations. In the simplest case, an ontology describes a hierarchy of concepts related by subsumption relationships; in more sophisticated cases, suitable axioms are added in order to express other relationships between concepts and to constrain their intended interpretation.” [7]

Ontologien in der Informatik sind explizite formale Repräsentationen und Konzeptualisierungen von Wissen in Form eines Graphs von Konzepten und ihren Beziehungen. Sie dienen der Strukturierung und Austauschbarkeit von Wissen und sind als eindeutige Wissensmodelle maschinell verarbeitbar. Es ist immer wieder zu beobachten, dass Verständigungsprobleme zwischen Menschen, zwischen Computern oder auch zwischen Menschen und Computern auftreten. Beispiele reichen hier von missverstandenen Anforderungen bei Aufträgen, über nicht kompatible Programme bis hin zu von Nutzern verschuldete Fehler bei der Programmausführung. Durch Ontologien können Annahmen über Wissensbereiche explizit dargestellt und ein einheitliches Verständnis dessen, was Informationen sind, festgelegt werden. Typischerweise bestehen Ontologien aus Klassen (Konzepten), Relationen (Beziehungen) zwischen Klassen, Instanzen (Individuen) von Klassen sowie Einschränkungen. Zu den zurzeit geläufigsten Sprachen zur Modellierung von Ontologien gehören das Resource Description Framework (Schema)⁵ (RDF/RDFS) und dessen Erweiterung Web Ontology Language⁶ (OWL). Jede Klasse, jedes Individuum und jede Relation ist durch einen Bezeichner (URI) eindeutig benannt, sodass jeder Bestandteil einer Ontologie durch ihren Namen eindeutig identifizierbar ist.

Technisch betrachtet ist eine Ontologie in RDF eine Sammlung von Tripeln der Form (Subjekt, Prädikat, Objekt), wobei Subjekt und Objekt Klassen oder Instanzen sind und Prädikat Relationen. Anschaulich kann man sich eine Ontologie folgendermaßen vorstellen: Jeder Bestandteil setzt sich aus zwei Haken zusammen, die durch eine Schnur verbunden sind. Dabei ist jeder Haken und jede Schnur durch einen Bezeichner eindeutig benannt. Es können nun beliebige Haken-Schnur-Haken-Kombinationen hinzugefügt werden. Spricht man von Ontologien, heißen Haken Klassen oder Instanzen und Schnüre Relationen; ein Bestandteil also eine Haken-Schnur-Haken-Kombination ist ein Tripel. Alle Haken bzw. Schnüre mit demselben Bezeichner bilden eine bestimmte Klasse bzw. eine Relation.

1.2 Ziel der Ontologie

Der Zweck einer Metaontologie liegt in den besonderen Anforderungen an Trendontologien begründet. Da ein Großteil an Informationen in Texten gesammelt ist, die für Menschen als LeserInnen gedacht sind, ist die automatische Wissensextraktion dadurch erschwert, dass die natürliche Sprache in ihrer semantischen Bedeutung zum einen höchst komplex und zum anderen nicht eindeutig ist. Dies liegt daran, dass die Satzstellung relevant und in verschiedenen Sprachen verschieden ist; Kontext kann demselben Satz eine völlig andere Bedeutung geben. Manche Wörter gewinnen im Zusammenspiel mit anderen eine neue Bedeutung,

⁵<http://www.w3.org/RDF/> (6/2011)

⁶www.w3.org/TR/owl-ref/

wobei eindeutige Regeln zur Bestimmung derselben fehlen.

“...the analysis of open ended questions still requires human involvement since it is based on the opinion analysis – text analysis where the steps based on categorization, generalization, and interpretation of information are mostly conducted manually.” [13]

Die Probleme aktueller Ansätze für die automatische Trenderkennung hängen vor allem mit den Schwierigkeiten zusammen, Trends in Texten (Kundenmeinung, Artikel, Nachrichten, etc.) automatisch zu erkennen. Die Hauptaufgaben solcher automatischen Trenderkennungssysteme sind das Erkennen und Auswerten relevanter Aussagen und die Einordnung dieser Aussagen anhand von domänenspezifischen Abhängigkeiten und Kategorien. Die Idee von Trendontologien als Wissensbasen zur Definition von Konzepten und Beziehungen für die textbasierte Trenderkennung ist naheliegend, denn in ihnen können Konzepte eindeutig definiert und zugehörige Begriffe bzw. Begriffskombinationen festgelegt werden. Auf diese relativ beliebig definierbaren Konzepte können nun Funktionen, also bspw. Abhängigkeiten oder Wirkungen angewandt werden.[13]

Im Rahmen des TREMA-Projektes an der Freien Universität Berlin wurden Trendontologien konzipiert. Es wurden auf Basis dreier Wissensmodelle unter aktiver Beteiligung von Expertinnen und Experten prototypische Trendontologien entwickelt. Das Ziel war eine leichtgewichtige Wissensbasis zu erstellen, um die Ergebnisse statistischer Methoden in Echtzeit zu verbessern. Die drei erstellten Modelle gliedern sich stufenweise in keyword-/konzeptbasierte Trendontologie, feldbasierte Trendontologie und zeitlich unabhängige Trendontologie, siehe 2.1 and [13].

Trendontologien haben unschärfere Strukturen als bspw. Ontologien für chemische Strukturen. Wenn Expertinnen oder Experten spezifische Begriffe verwenden, die in der allgemeinen Verwendung der Sprache anders ausgedrückt werden, müssen beide Versionen berücksichtigt werden. Außerdem ändern sich trendrelevante Konzepte und vor allem ihre Bedeutung und Wertung mit der Zeit. Einige können obsolet werden, während andere eine neue Bedeutung bekommen oder durch andere ersetzt werden. Kontext bestimmt maßgeblich die Bedeutung von Ausdrücken; Bedeutungen sind also abhängig von lokalen Faktoren. Wissen im Bereich der Trenderkennung ist höchst veränderlich und nicht unbedingt eindeutig definierbar. Dies gilt auch für die Ontologie, die dieses Wissen abbildet und nutzt.[13]

Weiterhin ist das in Trendontologien zu speichernde Wissen in unterschiedlichen Bereichen, Ländern oder Sprachen, also unterschiedlichen Anwendungsfällen, verschieden. Zwischen den verschiedenen Anwendungen ändern sich nicht nur Wörter und Ausdrücke, sondern auch Regeln, seien sie semantischer, logischer oder systematischer Natur. Trendontologien müssen also veränderlich, einfach nutzbar und anpassungsfähig sein und vor allem muss eine Vielzahl verschiedener spezifischer Trendontologien je nach intendierter Anwendung entwickelt werden. Diese sollten modular nutzbar und entwickelbar wie auch wiederverwendbar sein, um den Entwicklungsaufwand zu minimieren. Um eine Modularisierung von Trendontologien zu ermöglichen und eine Kompatibilität hinsichtlich des Datenaustausches zu gewährleisten, wird eine standardisierende Basis für Trendontologien benötigt. Diese Basis wird durch eine Metaontologie beschrieben, die grundlegende Konzepte, Relationen und Einschränkungen

definiert. Dabei ist sie unabhängig von Domänen und Zeitparametern valide. Sie stellt die grundlegenden für die Trenderkennung benötigten Konzepte und Funktionalitäten bereit, die je nach Anwendung zu spezialisieren sind bzw. durch Einbindung domänenspezifischer Ontologien erweitert wird. Um die Rolle einer standardisierenden Basis auszufüllen, bindet die Metaontologie nach Möglichkeit bereits bestehende anerkannte Ontologien und Erkenntnisse ein.

Es werden aktuell einige Bemühungen in der Forschung zur automatischen Trenderkennung unternommen wie zum Beispiel in [1], [12] und [18], jedoch existieren bisher kaum Forschungsergebnisse zu Trendontologien. Arbeiten zu diesem Thema entstanden im Rahmen des bereits erwähnten TREMA-Projektes und an der Universität Karlsruhe wie ein Vorschlag zu Trenderkennung von Hellingner in [8]. Veröffentlichte Trendontologien existieren nach unserer Kenntnis bisher nicht und dieser Bereich stellt damit ein junges und potenziell interessantes Thema dar. Aufbauend auf der erwähnten Konzipierung von Trendontologien wird im Folgenden die Entwicklung, Anwendung und Evaluation der Metaontologie dokumentiert. Hierbei werden Ontologien als ein mögliches Mittel zur Trenderkennung betrachtet, das durchaus in Kombination mit anderen Verfahren benutzt werden können.

1.3 Relevante Ontologien

Um die intendierte Standardisierungsfunktion zu erfüllen muss die Metaontologie selbst auf einer möglichst allgemeinen Basis stehen und sollte soweit wie möglich einschlägige Ontologien verwenden bzw. auf ihnen aufbauen. Jede neue Modellierungsart einer Anforderung ohne erkennbaren Mehrwert erhöht die Informationsmengen, ohne den Informationsgehalt zu verändern. Dies sollte hinsichtlich der Anforderung einer basisbildenden Metaontologie mit hoher Kompatibilität zu schon bestehenden oder noch zu entwickelnden Ontologien nach Möglichkeit vermieden werden. Es existiert eine Vielzahl von verschiedenen Ontologien, die meisten von diesen beziehen sich auf äußerst spezialisierte Bereiche. SKOS und OWL-Time sind zwei allgemein und einfach gehaltene Ontologien zur Darstellung von semantischen und temporalen Beziehungen. Beide Formen der Beziehungssetzung sind für Trendontologien sehr relevant, weshalb sie hier näher betrachtet werden.

1.3.1 Simple Knowledge Organization System - SKOS

SKOS⁷ ist eine der bekannteren Ontologien im Bereich des Semantic Web. Als W3C-akkreditierte Ontologie zur Darstellung von Hierarchien und Verwandtschaften hat sie eine gewisse Allgemeingültigkeit und erfüllt damit die grundlegenden Voraussetzungen für die Metaontologie.

Allgemeine Beschreibung

SKOS ist eine Thesaurus Ontologie. Es gibt nur wenige Klassen: *Concept*, *ConceptCollection* mit der Unterklasse *OrderedCollection* und *ConceptScheme*. Konzepte sind das Herzstück des SKOS Schemas. Sie beschreiben Gedankeneinheiten, wie z.B. Ideen, Objekte, Bedeutungen. In Konzeptsammlungen können ver-

⁷<http://www.w3.org/2004/02/skos/core#> (5/2011)

wandte Begriffe zusammengefasst werden. Es können hierarchische und assoziative Relationen zwischen Konzepten dargestellt werden. Durch *hasTopConcept* oder *inScheme* können Konzepte in ein Konzeptschema eingebunden werden. Das Schema ist ursprünglich als Thesaurusband gedacht, kann aber auch für Ontologiebereiche verwendet werden, da SKOS recht lose definiert ist und viele Möglichkeiten offen lässt. SKOS hat nur wenige Beschränkungen, es definiert Leitlinien kaum Verbote.

Funktionalität

SKOS unterscheidet zwischen zwei Kategorien von semantischen Relationen hierarchische und assoziative. All diese Relationen haben *Concept* als Domäne, können also auf alle Klassen angewendet werden, die Konzepte sind. Hierarchische Relationen beschreiben Abstufungen in der Allgemeingültigkeit von Konzepten. So kann man ausdrücken, dass ein Konzept in einem gewissen Sinn allgemeiner (*broader*) oder weniger allgemein (*narrower*) ist. *broader* und *narrower* sind invers zueinander. Auf diese Weise werden die gespeicherten Aussagen reduziert, ohne dass die Ontologie Aussagekraft verliert. Durch die Einführung der *broader-/narrowerTransitive*-Relation können Reasoning-Tools aus verketteten Relationen weitreichende Schlussfolgerungen über die Hierarchie von Ontologien erreichen, die SKOS benutzen.

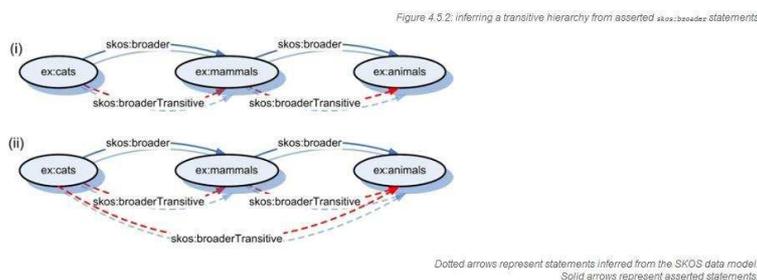


Abbildung 1.1: Transitive in SKOS [16]

Assoziative Relationen beschreiben Verwandtschaften zwischen Konzepten, die in keiner hierarchischen Relation stehen. Zu diesen Relationen gehören *relatedMatch*, *closeMatch* und *exactMatch* mit der Oberrelation *related*. Die verschiedenen Relationen geben verschiedene Verwandtschaftsgrade an. Während *relatedMatch* nur bedeutet, dass die Konzepte irgendwie in Beziehung stehen, folgt aus *exactMatch*, dass die beiden Konzepte tatsächlich Synonym verwendbar sind. Semantische und hierarchische Relationen können mit den Relationen *broadMatch/narrowMatch* u.A. kombiniert benutzt werden. Es gibt also eine enge Verbindung zwischen den semantischen und hierarchischen Relationen, da *broadMatch* bzw. *narrowMatch* Unterrelationen von *broader* bzw. *narrower* sind sowie beide Unterrelationen von *relatedMatch* und damit auch von *related* sind. Eine solche zusammengesetzte Relation enthält die relationalen Informationen der Teilrelationen.

SKOS stellt eine Bandbreite verschiedener Beschriftungsmöglichkeiten zur Verfügung: die Label. Durch sie können Konzepte in natürlicher Sprache beschrieben werden. Labels sind unterteilt in bevorzugte (*prefLabel*), alternative

(*altLabel*) und versteckte (*hiddenLabel*) Label. Letztere sind nur für textbasierte Index- und Suchsysteme sichtbar. Jedes Label kann mit einer Sprache (z.B.: animals@en) annotiert werden, um einfaches multilinguales Labeling zu erlauben.

Relevanz für Trendontologien

SKOS ist ein einfaches und weitreichend einsetzbares Schema, das generell zu einer allgemeinen standardisierenden Ontologie für die Trenderkennung passt. Alle assoziativen und hierarchischen Relationen in SKOS sind zwischen Konzepten. Um also die verschiedenen Teilontologien der Ebenen der Ontologie austauschbar zu halten und die volle Aussagekraft von SKOS zu behalten müssen nur alle Ebenen von *Concept* erben. Dadurch kann eine hohe Modularisierung erreicht werden, in der Assoziationen und Hierarchien allgemeingültig sind. Durch das Labeling wird die Dokumentation auch in verschiedenen Sprachen möglich und dadurch das Verständnis und der Umgang mit den Ontologien vereinfacht. In der Trendontologie könnten synonyme Konzepte auch durch verschiedene Label an einem Konzept definiert werden. Das spart je nach Implementierung Klassen oder Instanzen. Allerdings müsste das System, das die Ontologie verwendet, die Label durchsuchen, um ein Synonym zu finden. Wenn spezifischere Relationen benötigt werden, können diese teilweise als Unterrelationen von den in SKOS definierten gehandhabt werden. *ConceptSchemes* sind wenig sinnvoll, solange keine Relationen zwischen ihnen vorhanden sind, da sie keine weitere Funktionalität bergen und auch als separate Ontologien behandelt werden können. Es ist aber möglich über sie Beziehungen zwischen *Schemes* mit einer Relation *specializes* herzustellen oder unter anderem *broader* auf *Schemes* zu erweitern und so aus *Schemes* einen Spezialisierungsbaum aufzubauen.

1.3.2 OWL-Time

OWL-Time⁸ ist eine W3C akkreditierte Ontologie um zeitliche Inhalte von Webseiten und Web-Services darzustellen. Das zur Verfügung gestellte Vokabular kann topologische Relationen zwischen Zeitpunkten und Zeitabschnitten beschreiben, ebenso wie die Länge von Zeitabschnitten und die genaue Zeitangabe eines Zeitpunktes.[17]

Allgemeine Beschreibung

Grundlegend ist jede Zeiteinheit (*TemporalEntity*) ein Zeitpunkt (*Instant*) oder ein Zeitintervall (*Interval*) (siehe [17]). Jede *TemporalEntity* kann einen Anfang und ein Ende in Form von *Instants* haben. Ein *Interval* hat eine Dauer, *Instant* ist formal ein Zeitpunkt ohne Länge, kann aber nach Modellierung eine Dauer haben. *Instant* kann auch als sehr kurzes Intervall approximiert werden und *Properties hasBeginning* und *hasEnd* sind zulässig für alle *TemporalEntity*. Das heißt ein *Instant* kann als sehr kurzes Intervall betrachtet werden. Eine Zeitdauer ist durch die Klasse *DurationDescription* definiert und kann in beliebigen Einheiten von Sekunden bis Jahren ausgedrückt werden.

Um Zeitpunkte strikt von Intervallen zu trennen, gibt es die Klasse *ProperInterval* als Unterklasse von *Interval*, die disjunkt von Zeitpunkten ist. Diese wie-

⁸<http://www.w3.org/2006/time#> (5/2011)

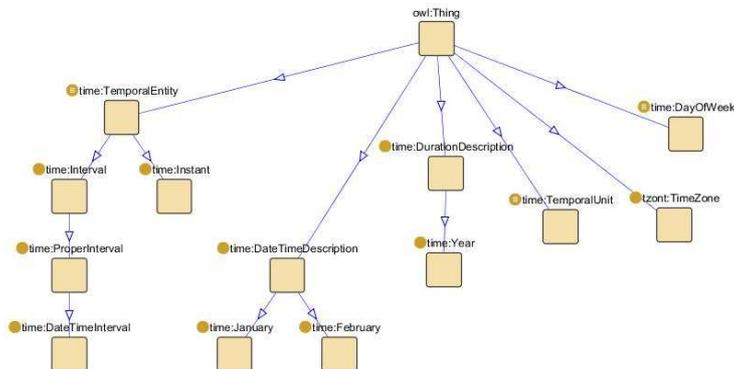


Abbildung 1.2: OWL-Time: Hierarchische Architektur

derum hat die Unterklasse *DateTimeInterval*. Nur *DateTimeInterval* sowie *Instant* können *DateTimeDescriptions* haben, das Herzstück der Ontologie, denn sie geben Zeitpunkte in einem standardisierten Format an. Eine *DateTimeDescription* muss eine Einheit (*TemporalUnit*) haben. Dies ist die kleinste benutzte Einheit, also hat die Uhrzeit 10:15 die Einheit min. Optional können *DateTimeDescriptions* außerdem folgende Werte zugewiesen werden: *year*, *month*, *week*, *day*, *dayOfWeek*, *dayOfYear*, *hour*, *minute*, *second*, and *timeZone*. Dadurch können Daten und Uhrzeiten für verschiedene Verarbeitungsweisen und für verschiedene Zeitzonen vergleichbar angegeben werden. Zeitzonen sind in der importierten W3-Ontologie⁹ definiert. Zur einfachen und individuellen Benutzung werden Unterklassen von *DateTimeDescription* und *DurationDescription* definiert. Soll bspw. der Monat Januar als Klasse definiert werden, wird eine Unterklasse der *DateTimeDescription* benutzt, die als Einheit *month* hat, wobei *month* auf -01 und alle anderen Einheiten auf 0 gesetzt wird. Das gleiche Paradigma gilt für *DurationDescriptions*. Möchte man die Zeitdauer in Jahren ausdrücken, wird hier eine Unterklasse definiert und die Kardinalität von *years* auf 1 gesetzt.[17]

Durch die topologischen Relationen können zeitliche Bezüge zwischen *TemporalEntity* hergestellt werden. Ist zum Beispiel A vor (*before*) B, dann bedeutet dies, dass das Ende von A vor dem Ende von B eintritt und durch *inside* wird ausgedrückt, dass ein Zeitpunkt während eines Intervalls geschieht. Weiterhin gibt es eine ganze Reihe von Relationen um Intervalle in Relation zueinander zu setzen wie *intervalAfter*, *intervalDuring* oder *intervalOverlaps*. Diese Relationen können alle nur auf echte Intervalle angewendet werden, haben also als domain und range *ProperInterval*.

Relevanz für Trendontologien

Zeit ist eine zentrale Einheit in vielen Bereichen und ganz besonders in der Trenderkennung, denn bei Trends handelt es sich um Veränderungen in einem gewissen Zeitraum. Zum einen kann es verschiedene Bewegungen desselben Themas zu verschiedenen Zeitpunkten geben, die erkennbar voneinander abgegrenzt

⁹<http://www.w3.org/2006/timezone#> (5/2011)

werden müssen. Zum anderen aber ist für die Erkennung von Trends in der Regel sehr wichtig, in welchem Zeitraum diese auftreten. Es ist sicherlich ein wichtiger Unterschied ob ein positiver Trend noch eine Woche oder mehrere Jahre anhält. Des Weiteren werden in Trendontologien Annahmen über Beziehungen und Namen getroffen, die nicht unbedingt dauerhaft sind. Hier ist es vorteilhaft, einen erwarteten Lifecycle dieser Konzepte definieren zu können. OWL-Time bietet einfache und individuell benutzbare Möglichkeiten, um diese Anforderungen zu erfüllen. Zeitpunkte, Intervalle und ihre Informationen können definiert werden und entweder durch Relationen, die einem zeitlich relevanten Konzept eine *TemporalEntity* geben, mit Konzepten in Verbindung gebracht werden oder die Konzepte selbst werden als Unterklassen von *TemporalEntity* Klassen definiert und können damit jede Information erhalten, die der entsprechenden *TemporalEntity* zugeschrieben werden kann.

Kapitel 2

Entwicklung der Metaontologie

Ausgehend von den Konzeptsammlungen für Trenderkennung, die im Rahmen des TREMA-Projektes, durchgeführt an der FU-Berlin in Kooperation mit betrieblichen Partnern¹, für die Domänen Finanzmarkt und Marktforschung entwickelt wurden, wird eine möglichst allgemeingültige Metaontologie zur Trenderkennung entwickelt². Die inhaltlichen Anforderungen durch Kompetenzfragen im Ontology Requirements Specification Document (OSRD) spezifiziert, das die in [14] vorgeschlagene Methodik benutzt, siehe A.6. Hierbei traten vor allem Verwandtschaften von verschiedenen Konzepten, deren Auswirkungen aufeinander sowie die Bewertung dieser Auswirkungen hinsichtlich spezifischer Gegebenheiten und Problemen als notwendig hervor. Dabei ist die Bewertung der Auswirkung nur schwer umzusetzen, da Bewertungen nicht nur von Domäne zu Domäne, sondern auch innerhalb einer Domäne verschieden sein können und häufig sogar von aktuellen Gegebenheiten und Ereignissen abhängen. Zudem sind sie abhängig von menschlichen Bewertungskriterien, die sich je nach Sachlage und Sichtweise schnell ändern können. Ein Beispiel dafür ist die Atomkraftdiskussionen in Deutschland und den diesbezüglichen Meinungsumschwung in Zusammenhang mit dem großen Erdbeben in Japan im März 2011³.

¹Finanzierung: Investitionsbank Berlin; Kooperation: neofonie GmbH, JRC GmbH, Metri-nomics GmbH

²Zur Modellierung der Ontologie wurde Protegé 3.4.1 genutzt.

³<http://www.bbc.co.uk/news/world-asia-pacific-12748215> (12/2012)

2.1 TREMA Ontologien

“Extending the keyword-based trend ontology we observed the emergence of so-called term fields in market research, which correspond to the semantic fields from the Semantic Field Theory.[10] Relying on the semantic field idea, the extension of concept definition by adding term fields to the concept seemed reasonable. However, defining which term belongs to the concept field and whether a given term is trend indicating or not is difficult.” [13]

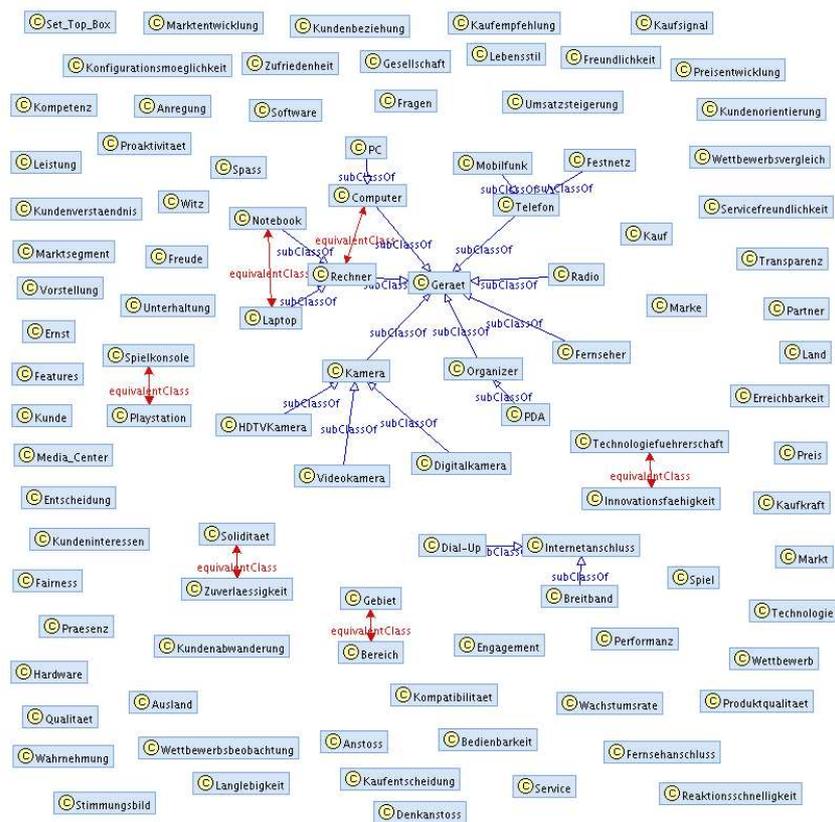


Abbildung 2.1: Untere Ebene der Marktforschungsontologie

Das TREMA-Projekt erforschte die Möglichkeiten der Trenderkennung aus gemischten Daten (Zahlen und Text). Dabei lag das Augenmerk vor allem auf Semantic Web Technologien sowie Data und Text Mining. Eines der Ziele war die Formalisierung von signifikanten Konzepten, mit deren Hilfe statistische Lernmethoden und Textclustering-Methoden für die automatische Trenderkennung verbessert werden können.

Die Ontologien, die im Rahmen des TREMA-Projektes entwickelt wurden, sind Konzeptsammlungen und bauen auf Studien aus den Jahren 2006/2007 auf. Hierbei wurden Texte hinsichtlich Themenstrukturen und Themenentstehung untersucht.

Es wurden für Marktforschungs- und Finanztrendontologien relevante Konzepte identifiziert, im Sinne der semantischen Feldtheorie zu Themenfeldern mit Hilfe von *part_of*- und *belongs_to*-Relationen zusammengefasst und trendspezifisch annotiert. Die Ontologien beinhalten Metainformationen von relevanten Konzepten, typische Keywords, sowie Informationen über den trendanzeigenden Charakter von Begriffen, Relationen und Regeln.[4]

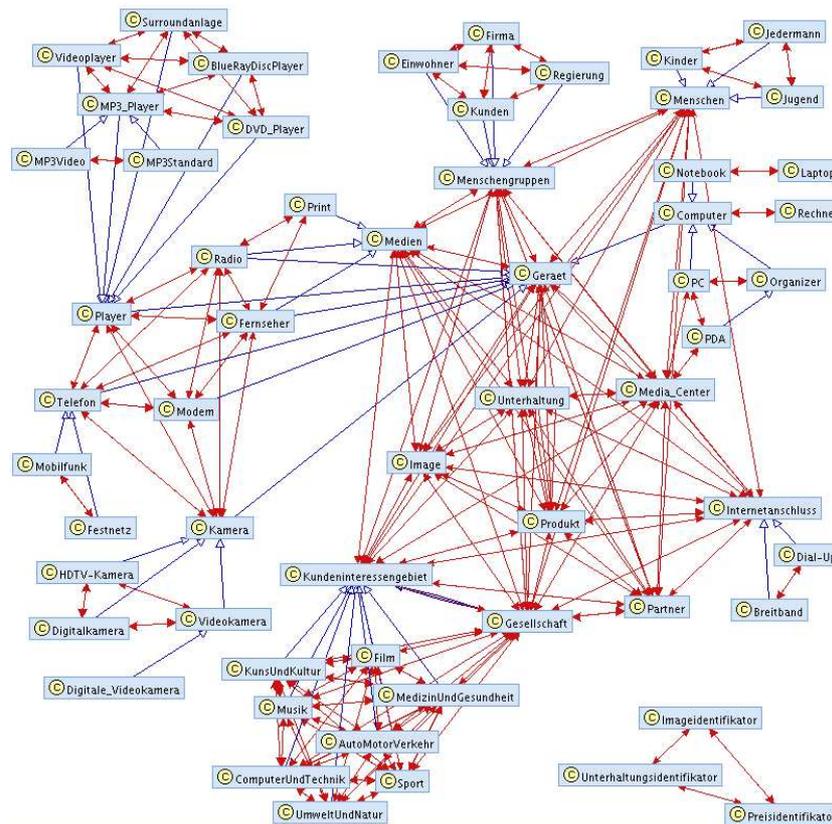


Abbildung 2.2: Mittlere Ebene der Marktforschungsontologie

Im Fall der Marktforschung basiert die Ontologie⁴ auf einem drei-stufigen Ansatz: in erster Ebene wurden relevante Konzepte mit Keywords definiert (siehe 2.1), diese bilden in der zweiten Ebene Themenfelder (siehe 2.2) und gliedern sich in dritter Ebene in ein zeitlich unabhängiges Schema (siehe 2.3), in dem Klassifikationen und Stimmungsdefinitionen festgelegt sind. Die erste Ebene basiert auf einem einfachen Schema und kann einfach angewendet werden, um bspw. maschinelle Lernmethoden zu erweitern.

Die zweite Ebene fügt Metakonzepte hinzu: *identifier*, *diversificator*, *sensitive*, *satisfier* and *disatisfier*.⁵ Diese werden benutzt um Zufriedenheit oder Unzufriedenheit anzugeben. Die dritte Ebene letztlich besteht aus den drei zentralen

⁴Im Folgenden wird die Ontologie dieses Anwendungsfalls als Marktforschungsontologie bezeichnet

⁵Diese Ausdrücke werden im „marketing Satisfaction Research“ benutzt.[13]

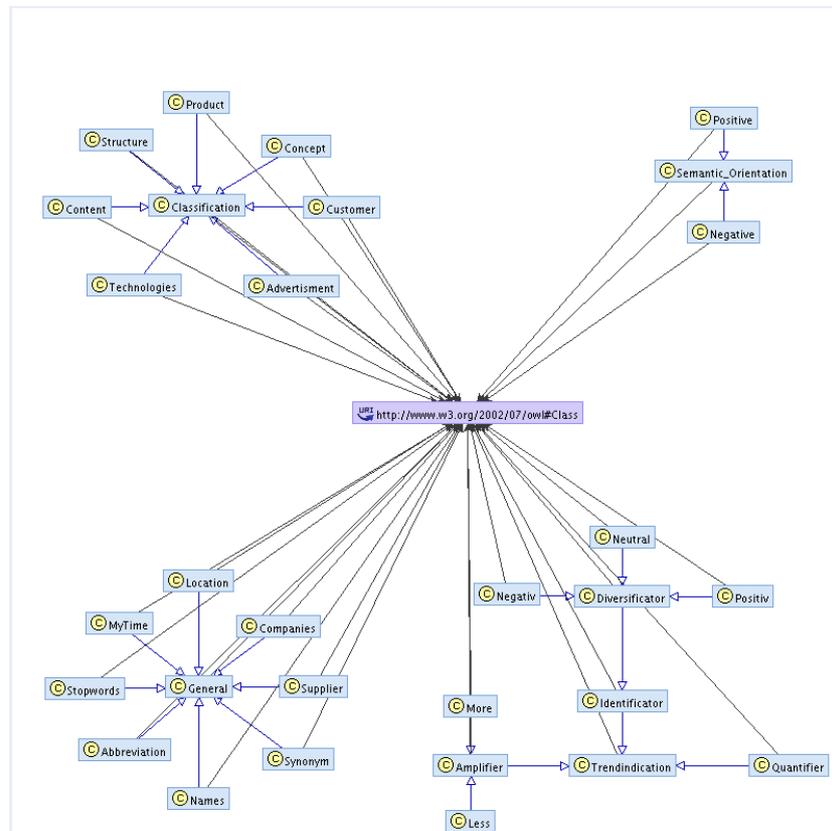


Abbildung 2.3: Meta-Ebene der Marktforschungsontologie

Klassen: *general*, *quantification* und *classification*. Dabei beinhaltet *general* die wichtigsten Konzepte, *quantification* die Idee von *identifier*, *diversificator* und fügt *amplifier*⁶ als neues Metakonzept hinzu. *classification* letztlich definiert den Kontext für *quantifier*. Die in dieser Arbeit thematisierte Metaontologie entspricht konzeptionell der dritten Ontologieebene.[13]

2.2 Metaontologie

2.2.1 Konzeptionelle Anforderungen

Um eine angemessene Metaontologie zu erarbeiten, müssen zuerst einmal die Anforderungen der Ontologie geklärt werden. Bei der Erarbeitung von Kompetenzfragen mit Hilfe des OSRD zeigte sich, dass es sich bei den Anforderungen hauptsächlich um die Einbettung von für Trends relevante Konzepte in das Gesamtbild bzw. Teile davon handelt. Angenommen man interessiert sich für den Ölpreis, dann sollte die Ontologie beispielsweise zum einen Informationen dazu liefern, welche Konzepte allgemein Einfluss auf den Ölpreis nehmen bzw. von

⁶ Amplifier sind Adverben, also: sehr, viel, wenig.[13]

ihm beeinflusst werden und zum anderen mit welchen Konzepten der Ölpreis zurzeit in Verbindung steht. Denn nur durch die Differenz aus allgemeiner Verwandtschaft und zeitspezifischen Verbindungen lassen sich schlüssige Informationen zur Entwicklung eines Themas und damit eines Trends bekommen. Des Weiteren werden Aussagen zur Art der Verwandtschaft von Konzepten und ihrer Einflussnahme aufeinander benötigt. Trends entstehen aus dem dynamischen Zusammenwirken einer Vielzahl miteinander in Verbindung stehender Teile. Als Beispiel fungiert hier einmal mehr das schon angesprochene Erdbeben in Japan diesmal in Zusammenhang mit Aktien von Firmen im Sektor der erneuerbaren Energien. Deren Kurse stiegen nach dem Erdbeben stark an, obwohl sie an sich nichts mit Erdbeben oder Japan zu tun haben. Zu Erklären ist dieser Anstieg durch die Verkettung verschiedener Umstände: Das Erdbeben als Auslöser, die damit einhergehenden Probleme am Atomkraftwerk Fukushima, die Unsicherheit gegenüber Atomkraft im Allgemeinen und letztendlich Investitionen in den Sektor erneuerbare Energien in der Hoffnung einer breiten Trendwende in diese Richtung. Derartige Entwicklungen können nur erkannt und verstanden werden, wenn die einzelnen Teile der Wirkungskette mit einbezogen werden, weshalb derartige Abhängigkeiten zu modellieren sind.

Hierfür werden Aussagen zu Relevanz, Wirkungsgrad, Risiko, Vertrauen sowie Indikatoren (stark, schwach, hoch, runter, positiv, negativ) benötigt, die kombiniert auf Verwandtschaften angewendet werden können. Die Metaontologie stellt die nötige Funktionalität einer derartigen Analyse zur Verfügung, sie wird jedoch in der Programmierung nicht aufgegriffen. Weitere wichtige Punkte sind die Kompatibilität mit anderen Ontologien sowie die individuelle Erweiterbarkeit der Metaontologie. Auf Grund der starken Abhängigkeit der Konzepte und Zusammenhänge von Domäne und Anwendungsbereich ist eine modulare Entwicklung und Nutzung von Trendontologien von Nöten. SKOS als einfaches, anerkanntes Wissensorganisationssystem bietet sich als Grundlage der Ontologie an und sollte die einzige Anforderung sein, um eine Ontologie zu integrieren.

2.2.2 Allgemeine Beschreibung

Die Architektur der Metaontologie (siehe 2.4) baut auf SKOS sowie OWLtime auf und benutzt eine von dem Verfasser erstellte Relation Ontologie. Alle Klassen der Metaontologie erben von *skos:Concept*, sodass die typischen SKOS-Relationen *related*, *member* und *broader/narrower* auf alle Klassen und Instanzen anwendbar sind. Die Auswahl der Klassen basiert auf der Auswertung der in A.6 spezifizierten Kompetenzfragen. Dort sind ausserdem intendierte Nutzung und Nutzer zu festgelegt, sowie eine Auflistung der wichtigsten Konzepte der Kompetenzfrage mit ihrer respektiven Umsetzungen in der Ontologie.

Die Ontologie gruppiert sich in vier Hauptklassen: *Player*, *Indication*, *Relational* und *ValuePartition*. *Player* sind die Dinge, die an und für sich relevant sind und ohne Verbindung zu anderen Konzepten Aussagekraft haben. Die Klasse ähnelt der Klasse *general* der prototypischen Trendontologie, siehe 2.1. Der Unterschied besteht im Fokus auf die Metaebene. Während *general* aus direkt in Texten vorkommende Konzepte wie *company*, sowie semantischen Komponenten besteht, fokussiert *Player* der Metaontologie die Benutzung und Einordnung solcher Konzepte durch *Grouping*, *Topic*, *Source* und *Trendstage*, während die Konzepte wie *supplier* auf Grund ihrer Domänenzugehörigkeit nicht in der Metaontologie enthalten sind und semantische Komponenten ausgelagert

Bindungsstärken, zur Relation. Diese Ontologie besteht nur aus zwei Klassen, *RelationalThing* und *RelatingThing*, sowie einer Relation *scalableRelated*. Ein *RelationalThing* ist hierbei etwas, was in Verbindung gesetzt werden kann, ein *RelatingThing* etwas, das eine Verbindung beschreibt. Durch die *Relational.owl* wird zum einen sichergestellt, dass eine *RelatingThing* nicht für sich stehen kann, da es zwingend ein *RelationalThing* braucht, und zum anderen wird die Möglichkeit von Verkettungen garantiert. Für eine Visualisierung der Möglichkeiten siehe 2.5. Ein Beispiel aus der Trendontologie wäre ein Ausdruck wie „Der Ölpreis hat starke Auswirkungen auf den Benzinpreis“. Bezogen auf *Relational.owl* wären Ölpreis und Benzinpreis *RelationalThings* und starke Auswirkung ein *RelatingThing*. Der Satz kann also dargestellt werden als: *Ölpreis scalableRelated starke Auswirkung scalableRelated Benzinpreis*. Hierbei ist *Auswirkung* ein *RelatingThing*, also eine zum Konzept erhobene Relation, die Dank ihres Konzeptdaseins durch den *Amplifier*⁷ stark modifiziert werden kann. Dies eröffnet eine ganze Bandbreite von Möglichkeiten hinsichtlich der Modifizierung wie auch der Assoziation von Relationen.

2.2.3 Klassen und Relationen

Klassen

Im Folgenden werden die zentralen Klassen der Metaontologie im Detail beschrieben und ihre Funktion erläutert.

Player *Player* ist die Zentrale Klasse der Ontologie und dem grammatischen Substantiv nachempfunden. Wird die Ontologie populiert, spannt sich ein Netz aus Dingen und Verbindungen. *Player* sind die Einstiegspunkte in das Netz und werden zur Trenderkennung betrachtet. Alle anderen sind dem *Player* in der Trenderkennung untergeordnet. Sie können zwar für sich existieren, doch signifikante Aussagekraft erhalten sie erst in Verbindung mit einem *Player*. *Player* sind all jene Dinge, die eine Wirkung haben; im Unterschied zu *Relational* und *Indication* die eine Wirkung beschreiben. Das heißt in der Praxis, dass anstelle von allen möglicherweise trendrelevanten Faktoren zur Trenderkennung nur die relevanten *Player* und ihre Verbindungen betrachtet werden. Themenbereiche werden in *Topic* eingebunden.

Topic *Topic* ist eine *skos:Collection*, das heißt ihm können beliebig viele *Concepts* als *skos:member* zugewiesen werden. Einem Thema werden Klassen, die eine Einordnung in dieses Thema anregen, als *skos:member* hinzugefügt. Idealerweise haben Themen eine oder mehrere Quellen (*Source*), in denen Informationen wie Datum, AutorIn, URI, Art der Quelle etc. vermerkt werden, um besondere – beispielsweise zeitliche – Abfragen zu erlauben und Rückverfolgungen zu ermöglichen.

Trendstage *Trendstage* erlaubt die Kategorisierung von trendrelevanten Faktoren in Trendstadien nach den in [15] definierten Kategorien.⁸ Durch diese Ka-

⁷Im Sinne der Metaontologie (siehe 2.4)

⁸Trendstadien: Trend creator, Trendsetter, Trendfollower, Mainstreamer, Conservative, Anti-innovator. Die Kategorien early und late mainstream wurden in der Ontologie nicht berücksichtigt da die Granularität für eine Trendontologie zu fein erachtet wurde.

tegorisierung werden Abfragen nach trenderzeugenden oder trendbeendenden Konzepten möglich oder auch gefilterte Abfragen ohne Mainstream Konzepte, sodass sich trendverstärkende, trenderhaltende und trendschwächende Konzepte extrahieren lassen. Dadurch ist eine genauere Einschätzung der Trendentwicklung möglich. Ein Beispiel aus dem Bereich des Börsenhandels sind die Analysten als Trendsetter, da sie es sind, die einen Trend unterstützen oder ihm entgegenstehen.

Indication *Indication* wurde mit kleinen Veränderungen aus der Marktforschungsontologie (siehe Kapitel 5.1) übernommen und spielt die Rolle des Adjektivs. Die Elemente dieser Klasse zeigen Richtung (*Indicator*) und Kraft (*Amplifier*) an. Durch *Role* kann, falls gewünscht, festgelegt werden, ob die Kraft oder Richtung positiv oder negativ bewertet wird. Es kann sowohl Relationen der *Relational* Klasse modifizieren als auch direkt mit einem *Player* in Verbindung gesetzt werden und diesen beschreiben. Da auch *Indication* von *skos:Concept* erbt, kann hier *skos:related* angewandt werden. Durch diese Architektur können in domänenspezifischen Ontologien neue Unterklassen oder Instanzen je nach Bedarf durch bereits definierte Bausteine wie *Strong* oder *Up* zusammengesetzt werden, was die Anzahl benötigter neuer Klassen reduziert, die Einheitlichkeit erhöht und automatisch die semantischen Beziehungen zwischen Indikatoren erhält.

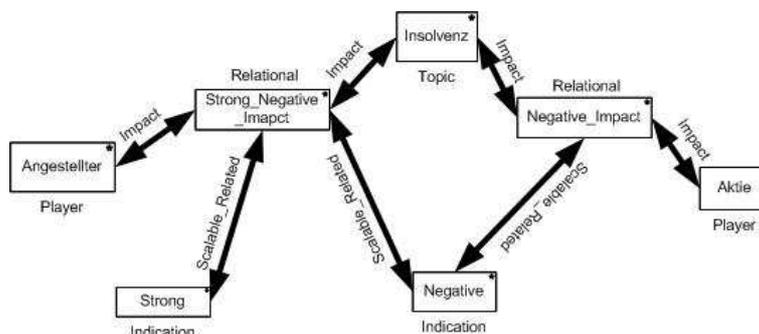


Abbildung 2.5: Beispielhafter Aufbau einer semantischen Verkettung mit skalierbaren Relationen dank *Relational.owl*

Relation *Relation* beschreibt Arten von Wirkungen auf gewisse Dinge. Hier kann *Impact*, *Relevance*, *Risk* und *Trust* von Etwas für Etwas definiert werden. Die Klassen in *Relation* können dabei als skalierbarer Mittler behandelt werden. Durch die Einbindung von *Relation.owl* und die Benutzung von *rel:scalableRelation* wird sichergestellt, dass Dreiecksverbindungen von direkten Verwandtschaften abgegrenzt werden und zudem es wird zudem angezeigt, dass diese Verbindung in ihrem Informationsgehalt skalierbar ist wie in 2.5 veranschaulicht wird. *ValuePartition* beinhaltet benutzerdefinierte Datentypen. In der Metaontologie wird hier nur *Keyword* benutzt.

Das Problem der Synonyme

In unseren natürlichen Sprachen sind viele Begriffe äquivalent. Sei es, dass sie aus verschiedenen Sprachen kommen und dasselbe bedeuten, sei es, dass ein Begriff eine Abkürzung ist oder auch, dass etwas mehrere Bezeichnungen hat. Diese Synonyme müssen einander zugeordnet werden, wenn es darum geht, Texte zu lesen und zu analysieren. Vielfach werden Synonyme als einzelne Klassen bzw. Instanzen realisiert mit einer Relation, die ihre Gleichheit anzeigt.

Dies vergrößert jedoch die Trendontologie unnötig. Denn eine Klasse sollte die Idee bzw. das Konzept von etwas darstellen, was bei Synonymen per Definition ein und das Selbe ist. Was sich unterscheidet sind die Namen der Konzepte und als solche sollten sie auch behandelt werden. Namen werden in der Klasse *Keyword* behandelt. Diese ordnet jedem Konzept die sie identifizierenden Namen zu. Ein *Keyword* kann durch *alternativeKeyword* beliebig viele Wörter zugeordnet bekommen. Diese Wörter entsprechen den synonymen Benennungen eines Konzeptes. Dadurch sind sowohl keine synonymen Klassen vorhanden, was redundant wäre, als auch alle Synonyme in dem entsprechenden *Keyword* in einer Klasse zusammengefasst, sodass jedes Konzept genau eine *Keyword*-Instanz benötigt; wobei es nicht zwingend auf nur eine beschränkt ist.

Relationen

Die von der Metaontologie verwendeten Relationen beschränken sich weitestgehend auf die von SKOS, OWL-Time sowie Relation.owl zur Verfügung gestellten Relationen, denn diese sollen hauptsächlich benutzt werden. Die zusätzlich entwickelten Relationen dienen den wenigen zwingend notwendigen Anforderungen an Klassen, nämlich der Zuordnung von Keywords durch *keyword* und den synonymen Begriffen durch *alternativeKeyword* sowie der Zuordnung eines *Counters* zur zählbaren Erfassung von Bindungen mittels *countableRelated* und *counter*. Alle anderen Relationen – *amplifies*, *indicates* und *role* – sind für den Fall, dass man *Relationals* oder *Role* modifizieren möchte, anstatt eine verwandtschaftliche Verbindung darzustellen. Letztlich gibt es noch die Relation *specializes* mit domain und range *skos:ConceptScheme*. Diese bietet die Möglichkeit Ontologien in eine Hierarchie zu bringen und dadurch festzuhalten, welche zu einer anderen passt.

2.2.4 Funktionalität

Die Metaontologie soll ermöglichen eine Menge von Texten zu parsen, die Ergebnisse in der Ontologie festzuhalten und aussagekräftig auszuwerten. Die Ontologie sollte daher alle Zusammenhänge, die durchs Parsen zum Vorschein treten, abspeichern können, sodass sich ein Netz aus Konzepten bildet, wobei die Platzierung im Netz Aufschluss über den derzeitigen Zustand eines Konzeptes gibt. Es bildet sich also semantische Felder aus, die Aufschluss über die thematische Umgebung eines Konzeptes geben. Die Änderung dieser Umgebung innerhalb eines Zeitraumes bildet eine thematische Änderung in diesem Zeitraum ab und zeigt einen Trend an. Dieser Ansatz basiert auf der in 5.1 erwähnten Feldtheorie. Hierzu werden mehrere gemeinsam benutzbare Funktionen zu Verfügung gestellt. Zum einen werden unter *Topic* verschiedene Themenbereiche entweder manuell oder maschinell erzeugt und ihnen entsprechende Konzeptmengen in

Form von *skos:Collection* zugeordnet. Dadurch lässt sich ein Netz aus Themen spannen und individuell nutzen. Weiterhin werden zahlenmäßige Bindungsstärken durch *countableRelated* und semantische Bindungsstärken durch *Indication* angewendet auf Relation ausgedrückt. Durch *skos:related* lassen sich einfache Verwandtschaften modellieren und falls eine semantische Erkennung geschrieben wird, die verschiedene Arten von Bindungen unterscheiden kann, können diese durch Relational beschrieben werden.

Ein zentraler Faktor in der Trenderkennung ist die Zeit und darf nicht vernachlässigt werden. Jeder *Player* ist eine *TemporalEntity*, wodurch *Lifecycles* und andere zeitliche Perspektive mittels *DurationDescriptions* festgelegt werden können. Dies kann dazu verwendet werden die Zeiträume von Trends oder Faktoren, die einen Einfluss ausüben, anzuzeigen, aber auch um die Validitätszeiträume von in Bedeutung oder Wirkung veränderlichen Konzepten zu definieren. Sei es bspw. um dieses Konzept dann zu löschen oder anzuzeigen, dass es geprüft werden muss. Genügt dies nicht, kann in Unterklassen auf *DateTimeInterval* mit einer *DateTimeDescription* spezialisiert werden.

Wie die Ontologie letztlich benutzt wird und was die Beziehungen aussagen, sollte auf der domänenspezifischen Ebene definiert werden. Mittels *scalableRelation* bzw. *rel:relationalThing* werden Funktionalitäten zur statistischen Beziehungssetzung, die in dieser Arbeit vorrangig benutzt wird, mittels *Counter*, sowie für semantische Beziehungen mittels bspw. *Impact* oder *Relevance* zur Verfügung gestellt. Letztere können durch semantisches Tagging und die Erkennung von Prädikaten durch Parsen automatisch extrahiert werden. Zusätzlich existieren die Möglichkeiten von SKOS mittels *skos:member* Themenbereiche zu modellieren und mittels *skos:related* allgemeinere, weniger spezifische oder manuell definierte Relationen darzustellen. All dies kann und soll bei der Anwendung der Ontologie auf bestimmte Domänen durch Unterklassen für die domänenspezifische Bedeutung spezifiziert werden. Dadurch wird ein anpassungsfähiges und mächtiges Werkzeug zur automatischen Generierung von sehr detaillierten semantischen Netzen zu Verfügung gestellt und nachfolgend die Abfrage nach thematischer Einbettung, Kontext und Wirkung zwischen Konzepten, Themen und Themenbereichen ermöglicht.

Kapitel 3

Anwendung der Metaontologie

Die Anwendung auf die Domäne Börsenhandel verdeutlicht die Funktionalität der Metaontologie und bietet die Möglichkeit zur Evaluierung. Die Börsenontologie wurde mit einem Datensatz von Aktienanalysen aus den Jahren 2007/2008 von Yahoo gefüllt und anschließend getestet. Diese Analysen in deutscher Sprache betreffen nationale wie auch internationale Aktien, vorrangig von in Deutschland und den USA ansässigen Unternehmen.

3.1 Anwendung für Aktienanalysen

Die in diesem Kapitel beschriebene Ontologie ist daher vorrangig auf Aktien aus diesem Bereich und Analysen auf Deutsch ausgerichtet. 3.1 zeigt eine Übersicht der Unterklassen von Grouping in der für den Börsenhandel spezialisierten Ontologie. Auf Grund der zeitlichen Begrenzung der Arbeit musste ein einfaches System für die thematische Extraktion der Daten gewählt werden. Um eine einfache Verarbeitung zu gewährleisten und gleichzeitig Themen zu extrahieren, die später als Themenfelder dargestellt werden, werden Dokumente manuell definierten Themenbereich zugeordnet. Der passende Themenbereich wird instanziiert und erkannte Konzepte durch skos:member der Instanz zugeordnet. Dadurch lässt sich ein außerordentlich hoher Grad der Automatisierung erreichen, da nur Konzepte und Schlüsselwörter sowie einige wenige Themenbereiche definiert werden müssen.

Dies funktioniert gut, da die Dokumente des Testkorpus zum einen verhältnismäßig kurz und beschränkt in ihrer thematischen Breite sind und da zum anderen in ihnen bereits Keywords extrahiert und entsprechend annotiert wurden. Bei längeren Texten bietet sich eine Volltext-Extraktion mit einer „Thema in Thema“ Zuordnung an, bei der Absätze oder Sätze zu Themen gruppiert werden und die Menge der (Ab)Sätze einem Oberthema (Dokument) zugeordnet werden. Durch die Einbettung der einzelnen Konzepte in die Themenstruktur wird die Einschätzung der Analytinnen und Analysten, die die Analysen des benutzten Datensatzes produzierten, bezüglich dieses Konzeptes sichtbar.

3.1.1 Klassen

Um den Finanzmarkt betrachten zu können, wird die Ontologie durch börsenrelevante Konzepte erweitert. Dabei handelt es sich vor allem um Firmen, deren Anteile an der Börse gehandelt werden, sowie um handelbare Gegenstände. Weiterhin sind Kennzahlen wie Umsatz, Zinssatz, Kurs u. A. relevant, wie auch die Art des Konzerns (z.B. Zulieferer), die Art der Anlage – Aktie (Share), Anleihe, handelbare Gegenstände (z.B. Resource) und Werte (z.B. MoralValue). Kennzahlen sind nichts weiter als spezifische Datentypen, die einem Konzern zugeordnet werden und gehören als solche zur Klasse ValuePartition. Alle anderen Klassen können an und für sich Relevanz haben und gehören als solche zur Klasse Player. Letztlich fehlen noch börsenspezifische Themenbereiche. Die Konzepte Gold, Aktie, Insolvenz, Umsatz und Unternehmen (Company) wurden aus der Finanzontologie (siehe Kapitel 5.1) übernommen, jedoch nicht durch finanztechnische sondern semantische Logik definiert.

3.1.2 Themenbereiche

Finanznachrichten wurden in sieben Teilgruppen geteilt. Die Zuordnung von Dokumenten zu den Themenklassen erfolgt einerseits durch entsprechende Keywords die einen betreffenden Themenkomplex anzeigen, wie z.B. bei Insolvenz oder Kursziel, und andererseits durch die Zuordnung von Klassen zu Themenbereichen, wenn sie beim Vorkommen in einem Dokument auf diesen Themenbereich hinweisen. Durch *skos:related* werden weiterhin Verbindungen zwischen Themen und Klassen angezeigt, ohne dass die Klassen auf eine Einordnung auf dieses Thema hinweisen. Beispielsweise ist *BusinessVolume* Konzern und Kennzahl durch *skos:member* zugeordnet worden, weil zu einem Bericht über Unternehmenszahlen das Unternehmen und spezifische Finanzzahlen gehört. Außerdem ist es auch durch *skos:related* verwandt mit *low_risk* weil Umsatzmeldungen bzw. allgemein Zahlenwerte von Unternehmen meist eindeutig sind, da sie berechnet werden können, und daher wenig Risikopotenzial bergen. Nachstehend sind die sieben entwickelten Teilgruppen aufgeführt.

- **BusinessVolume:** Informationen zur technischen Marktstellung von Konzernen, also Umsatznachrichten, Quartalszahlen, Schulden usw.
- **Themenbildende Klassen:** Umsatz, Company



Abbildung 3.1: Klassen der Boersen-Trendontologie

- **ConcernNews:** Informationen zu Konzernen, die mehr inhaltlicher Natur sind. Also bspw. Marktstellung, Kundenentwicklung, Aufsichtsratsänderung.
 - **Themenbildende Klassen:** ConcernType, Company
- **FinancialEvent:** Für den Finanzmarkt relevant Ereignisse; Insolvenz, Naturkatastrophen, Krieg
 - **Unterklassen:** Insolvenz (Keyword: insolvenz), Uebernahme (Keyword: uebernahme)
 - **Themenbildende Klassen:** Location
- **MarketNews:** Allgemeine Marktberichte; Konjunkturschwäche, Marktstimmung, Krisen, Blasen.
 - **Themenbildende Klassen:** Group, Political, Sentiment
- **Recommendation:** Anlageempfehlung
 - **Themenbildende Klassen:** Analyst, Share, FinancialInstrument
- **ShareNews:** Allgemeine Nachrichten zu Aktien
 - **Themenbildende Klassen:** Company, Share
- **SharePriceChange:** Spezifische Nachrichten zu Aktien mit Kurszielangaben.
 - **Unterklasse:** Kursziel
 - **Themenbildende Klassen:** Share, Kurs

In diese Themen können Dokumente anhand von themenidentifizierenden Keywords oder passenden Kombinationen von Konzepten innerhalb eines Dokumentes sortiert gespeichert werden und anschließend nach gewünschten Themenbereichen ausgegeben werden.

Kapitel 4

Experimente und Evaluierung

Eine Evaluation der Trendontologie kann einerseits durch den Vergleich der letztendlichen Möglichkeiten mit anfangs erarbeiteten Anforderung und andererseits durch Vergleich mit Fakten geschehen.

4.1 Evaluierung hinsichtlich Kompetenzfragen

Es wurden Kompetenzfragen sowohl allgemeiner als auch finanzspezifischer Natur definiert, die die Ontologie durch Benutzung des Programms beantworten können soll. Beispielhaft werden nun einige dieser Fragen anhand von Ontologieabfragen mit Hilfe des Programms behandelt. Es ist zu beachten, dass zahlenbasierte Ergebnisse in der Praxis erst im Vergleich mit anderen Werten Signifikanz gewinnen. Eine vollständige Auflistung der Kompetenzfragen mit entsprechenden Abfragen der Ontologie findet sich in Anhang A.6.

Wie relevant ist ein Konzept allgemein (Frage 9)?

Angenommen es stellt sich die Frage, wie sehr Deutschland thematisiert wird. Die Abfrage der Ontologie nach diesem Konzept ergibt folgende Ausgabe¹:

```
../TrendDetection\_v2\#Germany (9137) has Topic structure:  
../TrendDetection\_v2\#Financial : 1142  
../TrendDetection\_v2\#kaufen : 1003  
../TrendDetection\_v2\#MaschinenbauIndustrie : 650  
../TrendDetection\_v2\#Share : 606  
../TrendDetection\_v2\#Kurs : 562  
../TrendDetection\_v2\#Up : 520  
../TrendDetection\_v2\#Industrie : 510  
../TrendDetection\_v2\#Anlage : 468  
../TrendDetection\_v2\#Zulieferer : 422  
../TrendDetection\_v2\#AutomobilIndustrie : 414
```

Der für diese Frage relevante Wert ist 9137, die absolute Häufigkeit des Vorkommens von Deutschland. Wie bereits erwähnt, hat dieser Wert an und für sich

¹Alle Ausgaben dieses Kapitels wurden auf ihre wesentlichen Bestandteile gekürzt.

keine Aussagekraft. In Relation mit anderen Konzepten lässt sich jedoch mehr erfahren:

```
../TrendDetection_v2\#USA (4808)
../TrendDetection_v2\#Deutsche\Telekom (442)
../TrendDetection_v2\#Allianz (433)
../TrendDetection_v2\#Schweiz (382)
../TrendDetection_v2\#Starbucks (104)
```

Es wird nun deutlich, dass Deutschland ein äußerst relevantes Konzept ist. Unter Berücksichtigung des Testdatensatzes, der vorrangig Analysen deutscher Unternehmen beinhaltet, ist dies kaum verwunderlich. Weiß man jedoch nicht, dass dies der Fall ist, kann man es an der Ausgabe ablesen.

Wie relevant ist ein Konzept in einem speziellen Fall (Frage 10)?

Als Beispiel für diese Fragestellung wird nach der thematischen Relevanz von Branchen je nach Land gefragt. Es wird eine prozentuale Ausgabe benutzt bei der die spezifische Bindungszahl prozentual zur absoluten Bindungszahl des Konzeptes angegeben werden, da diese den Anteil des Vorkommens pro Konzept angeben. Diese Werte sind im Gegensatz zu absoluten Werten, deren Bedeutung von der Gesamtzahl der Verbindungen abhängt, direkt vergleichbar. Es werden Deutschland und die USA verglichen.

```
../TrendDetection_v2\#Germany (9137) has Topic structure:
../TrendDetection_v2\#Financial : 0.124
../TrendDetection_v2\#kaufen : 0.11
../TrendDetection_v2\#MaschinenbauIndustrie : 0.071
../TrendDetection_v2\#Share : 0.066
../TrendDetection_v2\#Kurs : 0.062
../TrendDetection_v2\#Up : 0.057
../TrendDetection_v2\#Industrie : 0.056
../TrendDetection_v2\#Anlage : 0.051
../TrendDetection_v2\#Zulieferer : 0.046
../TrendDetection_v2\#AutomobilIndustrie : 0.045

../TrendDetection_v2\#USA (4808) has Topic structure:
../TrendDetection_v2\#Kurs : 0.116
../TrendDetection_v2\#_Neutral : 0.115
../TrendDetection_v2\#Pharma : 0.11
../TrendDetection_v2\#kaufen : 0.105
../TrendDetection_v2\#InformationTechnologies : 0.09
../TrendDetection_v2\#Dienstleister : 0.069
../TrendDetection_v2\#halten : 0.056
../TrendDetection_v2\#Strong_Down : 0.05
../TrendDetection_v2\#Up : 0.048
```

Es lässt sich erkennen, dass Informationstechnologien und Dienstleister im Hinblick auf die USA sehr viel häufiger thematisiert werden, wogegen in Zusammenhang mit Deutschland die Industrie und speziell Maschinenbau und Automobilherstellung relevante Themen sind.

Wie ist die Bewertungslage einer Aktie (Frage 31)?

```
../TrendDetection\_v2\#Novartis (182) has Topic structure:
../TrendDetection\_v2\#Schweiz : 54
../TrendDetection\_v2\#Pharma : 54
../TrendDetection\_v2\#kaufen : 28
../TrendDetection\_v2\#Share : 20
../TrendDetection\_v2\#Up : 14
../TrendDetection\_v2\#Übernahme : 2
../TrendDetection\_v2\#halten : 2
../TrendDetection\_v2\#Strong\_Up : 2
../TrendDetection\_v2\#FinancialMarket : 2
../TrendDetection\_v2\#Chemie : 2
../TrendDetection\_v2\#Strong\_Down : 2
../TrendDetection\_v2\#\_Neutral : 2

../TrendDetection\_v2\#Porsche (154) has Topic structure:
../TrendDetection\_v2\#AutomobilIndustrie : 50
../TrendDetection\_v2\#Germany : 50
../TrendDetection\_v2\#Up : 12
../TrendDetection\_v2\#\_Neutral : 10
../TrendDetection\_v2\#Kurs : 8
../TrendDetection\_v2\#FinancialMarket : 8
../TrendDetection\_v2\#Share : 6
../TrendDetection\_v2\#halten : 4
../TrendDetection\_v2\#nachhaltigkeit : 2
../TrendDetection\_v2\#verkaufen : 2
../TrendDetection\_v2\#Company : 2
```

Anhand der Anzahl an Indikatoren (kaufen/Strong_Up/Up, halten/_Neutral, verkaufen/Down/ Strong_Down) kann eine Rangliste der Bewertungslage verschiedener Unternehmen bzw. Aktien aufgestellt werden aus der sich die relative Bewertungslage einer Aktie erschließt. Dies wird in 4.2 demonstriert. Die Aufstellung ergibt ein Bewertungsverhältnis von 37/3/2 für Novartis und 6/9/2 für Porsche und ergibt eine Bewertung von 50,5 (Novartis) zu 2,95 (Porsche) auf einer Skala von 0 (negativ) bis 100 (positiv).²

Wie ist die Bewertungslage einer Branche (Frage 32)?

Durch die Abfrage nach Branchen lassen sich ganze Wirtschaftsbereiche analysieren.

```
../TrendDetection\_v2\#Bank (1987) has Topic structure:
../TrendDetection\_v2\#Dienstleister : 442
../TrendDetection\_v2\#Germany : 396
../TrendDetection\_v2\#Financial : 197
../TrendDetection\_v2\#Immobilie : 182
../TrendDetection\_v2\#Broker : 178
../TrendDetection\_v2\#halten : 160
```

²Es handelt sich um Indikatorverhältnisse nach dem Schema positiv/neutral/negativ. Die Bewertung erfolgt nach Quant 3. Das Prinzip der Berechnung wird in 4.2 erläutert.

```

../TrendDetection\_v2\#kaufen : 120
../TrendDetection\_v2\#USA : 112

../TrendDetection\_v2\#Industrie (1844) has Topic structure:
../TrendDetection\_v2\#Germany : 510
../TrendDetection\_v2\#Zulieferer : 316
../TrendDetection\_v2\#MaschinenbauIndustrie : 260
../TrendDetection\_v2\#Anlage : 174
../TrendDetection\_v2\#AutomobilIndustrie : 116
../TrendDetection\_v2\#halten : 106
../TrendDetection\_v2\#kaufen : 100
../TrendDetection\_v2\#Dienstleister : 100
../TrendDetection\_v2\#Kurs : 86
../TrendDetection\_v2\#Up : 76

../TrendDetection\_v2\#Dienstleister (3820) has Topic structure:
../TrendDetection\_v2\#kaufen : 580
../TrendDetection\_v2\#Financial : 578
../TrendDetection\_v2\#InformationTechnologies : 524
../TrendDetection\_v2\#Bank : 442
../TrendDetection\_v2\#Kurs : 408
../TrendDetection\_v2\#halten : 362
../TrendDetection\_v2\#USA : 334
../TrendDetection\_v2\#Broker : 224
../TrendDetection\_v2\#Immobilie : 204
../TrendDetection\_v2\#Up : 164

```

Einmal mehr zeigen die Indikatorverhältnisse von Industrie 138/106/0, Dienstleistern 662/362/0 und Banken 120/160/0 die Bewertungslage der betreffenden Branchen an. Es ist zu erkennen, dass Dienstleister zur Zeit der Analysen in einer besseren Lage waren als die Industrie und Banken weit abgeschlagen sind.³ Wird mit einbezogen, dass Banken selbst Dienstleister sind wird die Diskrepanz zwischen Dienstleistung und Bankensektor besonders auffällig. Eine solche Erkennung der Problembereiche kann wertvolle Hinweise bieten, in welche Bereiche es lohnenswert ist zu investieren und besonders in welche nicht, geschieht automatisiert und kann als Zahl zu automatischen Berechnungen weiterverwendet werden.

Exkurs: Erweiterung der thematischen Umgebung

Dank der Einbettung in das semantische Feld kann die betrachtete thematische Umgebung eines Konzeptes erweitert werden.

Normale Abfrage:

```

../TrendDetection\_v2\#Schweiz (382) has Topic structure:
../TrendDetection\_v2\#Pharma : 0.168
../TrendDetection\_v2\#Share : 0.151
../TrendDetection\_v2\#Novartis : 0.141

```

³Kaufen-Halten-Verhältnis: Dienstleister 1,83; Industrie 1,3; Banken 0,75.

```
../TrendDetection\_v2\#kaufen : 0.136
../TrendDetection\_v2\#Bank : 0.089
../TrendDetection\_v2\#Chemie : 0.047
../TrendDetection\_v2\#\_Neutral : 0.047
../TrendDetection\_v2\#Up : 0.037
../TrendDetection\_v2\#halten : 0.031
../TrendDetection\_v2\#Financial : 0.031
```

Erweiterte Abfrage:

```
../TrendDetection\_v2\#Schweiz (32637) hast Topic structure:
../TrendDetection\_v2\#Germany : 0.15
../TrendDetection\_v2\#USA : 0.098
../TrendDetection\_v2\#Dienstleister : 0.089
../TrendDetection\_v2\#Financial : 0.08
../TrendDetection\_v2\#kaufen : 0.08
../TrendDetection\_v2\#halten : 0.058
../TrendDetection\_v2\#Kurs : 0.05
../TrendDetection\_v2\#Bank : 0.04
../TrendDetection\_v2\#Up : 0.035
../TrendDetection\_v2\#Share : 0.034
```

Während die Schweiz in erster Ebene thematisch hauptsächlich vom Pharmakonzern Novartis definiert ist die weiterreichende thematische Ausrichtung der Schweiz über vier Ebenen weit mehr von internationalen Beziehungen geprägt. Durch den starken Abfall von kaufen-Indikatoren bei erweiterter Themenfeldbetrachtung und die Betrachtung der Indikatoren von Novartis (Frage 31) lässt sich außerdem feststellen, dass die überdurchschnittlich hohe Zahl dieser Indikatoren hauptsächlich auf Novartis zurückzuführen ist und nicht unbedingt dem Marktklima der Schweiz entspricht.

4.2 Vergleich von Prognose und Kursentwicklung

Der eingeleseene Dokumentencorpus umfasst über 5000 Aktienanalysen des Zeitraumes Frühjahr 2007 bis Frühjahr 2008. Es können daher beispielhafte Prognosen erstellt werden und diese anschließend mit den tatsächlichen Kursentwicklungen bis 7/2011 verglichen werden. Die Ontologie ist auf einem Informationsstand, wie er schon im Frühjahr 2008 möglich gewesen wäre. Auf Basis dieses Standes wird eine Aktienausswahl gewählt und die Kursentwicklung dieser Aktien beobachtet. In Anhang C findet sich eine Übersicht von 21 Aktien mit der Ontologie entnommenen Trendindikatoren, der prozentualen Kursentwicklung, sowie aus den Empfehlungen errechneten Kennzahlen. Die Aktien wurden nach der Häufigkeit ihres Vorkommens in der Ontologie ausgewählt. Es sind also diejenigen, die im Dokumentencorpus am häufigsten thematisiert werden.

4.2.1 Auswahl der Indikatoren

Die betrachteten Indikatoren gliedern sich in drei Gruppen: positiv, negativ und neutral. Dies entspricht der Zuordnung der Role-Klasse der Metaontologie. Die jeweils zu derselben Gruppe gehörenden Indikatoren werden gewichtet addiert, um das Gewicht der jeweiligen Gruppe je Aktie zu bestimmen. Die Gewichtung nimmt für klare Indikatoren, d.h. Erwähnungen von bspw. kaufen und overweight, eins an und für schwache oder mehrdeutige Indikatoren 0,5.

- Positive Indikatoren
 - kaufen (Keywords: kaufen, buy), Gewichtung: 1
 - Strong_Up (Keywords: overweight, overperform), Gewichtung: 1
 - Up (Keywords: hoch, über, up), Gewichtung: 0,5
- Neutrale Indikatoren
 - halten (Keywords: halten, hold), Gewichtung: 1
 - neutral (Keywords: neutral, durchschnitt), Gewichtung: 0,5
- Negative Indikatoren
 - verkaufen (Keywords: verkaufen, sell), Gewichtung: 1
 - Strong_Down (Keywords: underperform, underweight), Gewichtung: 1
 - Down (Keywords: unter, runter, down) , Gewichtung: 0,5

4.2.2 Berechnung der Kennzahlen

Die Kennzahlen sind so ausgelegt, dass sie groß sind, wenn die Empfehlungslage gut ist. Es wurden zwei Kennzahlen erstellt.

Quant1 Die erste Kennzahl benutzt das Verhältnis aus positiv/neutralen Indikatoren zu negativen und berechnet sich aus:

$$\frac{\text{positiv} + \frac{\text{neutral}}{4}}{2 * \text{negativ}}$$

Der Faktor von zwei hinsichtlich negativer Indikatoren erschließt sich aus der Verlustaversionskennzahl von 2,5; sie wurde auf zwei reduziert, da kein unmittelbarer Verlust auftritt.[11] Der Divisor von vier der neutralen Indikatoren zeigt an, wie gut neutrale Indikatoren im Vergleich zu positiven gewertet werden, und wird sich je nach Markthaltung ändern.⁴

Quant2 Quant1 berücksichtigt allerdings nicht direkt Verhältnisse zwischen positiven und neutralen Indikatoren, weshalb sie bei einer großen Zahl an neutralen und wenigen positiven Indikatoren unverhältnismäßig gute Ergebnisse liefert. Dies wird in der zweiten Kennzahl kompensiert:

$$\frac{\text{positiv} + 2 * \frac{\text{positiv}}{\text{neutral}} * \frac{\text{positiv}}{\text{negativ}} + \frac{\text{neutral}}{8}}{4 * \text{negativ}}$$

Diese Kennzahl basiert hauptsächlich auf den Verhältnissen zwischen positiven und anderen Indikatoren. Der hohe Faktor von vier bei den Verhältnissen wie auch der erhöhte Divisor von neutral legt das Augenmerk auf diese Verhältnisse. Der Negativ-Faktor wurde erhöht, um das häufigere Auftreten von positiv im Zähler zu kompensieren.

Quant3 Da Quant 2 bei einer großen Menge von neutralen Indikatoren unverhältnismäßig klein wird, werden Quant 1 und Quant 2 kombiniert, wobei Quant 2 vierfach so stark gewichtet wurde wie Quant 1, da letztere bei durchschnittlichen Werten unverhältnismäßig größer ist als Quant 2 und ohne Reduzierung das Ergebnis verfälscht. Dies ist die kombinierte Kennzahl Quant 3:

$$\left(\frac{\text{Quant1}}{4 * \max(\text{Quant1})} + \frac{\text{Quant2}}{\max(\text{Quant2})} \right) * 100$$

Wie schon in Abschnitt 4.1 angesprochen, gewinnen die der Ontologie entnommenen Daten, sofern es sich wie in dieser Arbeit um eine quantitative Analyse handelt, erst in Relation zueinander an Bedeutung. Daher lag auch eine Normierung auf den aktuellen Maximalwert der Aktienbewertung nahe, der sich bei Änderung des Ontologiezustandes mit verändert.

4.2.3 Vergleich der Kursentwicklung

Es wird ein Korb der laut Kennzahlen besten Aktienwerte (höchste Werte hinsichtlich Quant 3) ausgewählt und damit prognostiziert, dass sich die ausgewählten Werte in den Folgejahren der Analysen⁵ überdurchschnittlich entwickeln.

⁴Dieser Faktor basiert auf persönlicher Einschätzung anhand der Stimmungslage des Analysenzeitraumes. Er war geprägt vom Ausklingen einer langen euphorischen Stimmungslage und dem Platzen der Investmentblase, was bedeutende negative Einschläge verursachte. Je euphorischer die Markthaltung ist, desto geringer sind neutrale Indikatoren zu bewerten.

⁵Der Dokumentenkörper besteht aus Analysen der Jahre 2007/2008. Der betrachtete Zeitraum ist 7/2007 - 7/2011.

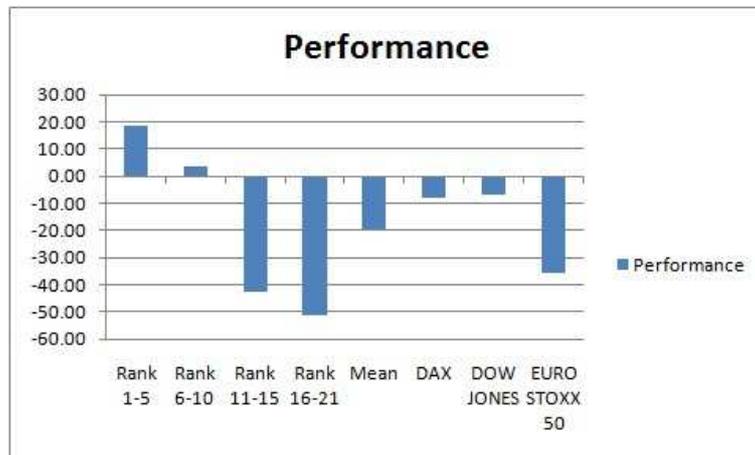


Abbildung 4.1: Wertentwicklung der Aktien nach Quant3 in 5er-Gruppen sortiert und Vergleich mit üblichen Indizes im Zeitraum 7/2007 - 7/2011.

Dieser wird mit der durchschnittlichen Kursentwicklung aller 21 Aktien⁶ sowie der in den letzten vier Jahren (7/2007 - 7/2011) real besten (höchster Kursgewinn) Aktien verglichen. Dabei werden einmal die besten fünf und einmal die besten acht Aktien verglichen. Dadurch wird einerseits genügend Spielraum gewährleistet, um von dem Durchschnitt deutlich abweichende Ergebnisse zu erhalten und andererseits ist die Menge genügend groß, um Zufallstreffer zu vermeiden. Eine Aufstellung der Aktien nach realer vier-Jahres-Performance im Vergleich zur Bewertung durch die Kennzahlen findet sich in Anhang C. Durch die Auswahl wurde in 70% der Fälle eine Genauigkeit von 80% oder mehr erreicht und bei weiteren 10% lag die Abweichung unter einem Drittel.

Abbildung 4.1 zeigt die Wertentwicklung der Aktienausswahl des Testcopus gruppiert hinsichtlich ihrer Bewertung und Abbildung 4.2 zeigt sie im Vergleich zu der retrospektiv besten Auswahl. Der Kursgewinn der Auswahl von acht Aktien lag in dem betrachteten Zeitraum 45% über dem des Durchschnitts und 9% unter der Performance der real acht besten Aktien. Durch eine kleinere Auswahl erhöht sich das Gewinnpotenzial, allerdings auch das Risiko einer unterdurchschnittlich abscheidenden Auswahl. Die Aktien der Auswahl der nach Quant 3 besten fünf Werte stehen in der Rangliste der vier-Jahres-Performance auf Rang 2, 6, 4, 1 und 10 in dieser Reihenfolge. Der Kursgewinn dieser Auswahl lag in vier Jahren 61% über der durchschnittlichen Entwicklung und 12% unter dem Kursgewinn der real besten fünf Aktien.

Es wäre also durch die ontologiegestützten Auswahl innerhalb von vier Jahren eine 45% bzw. 61% höhere Performance erzielt worden, als durch den gleichmäßigen Kauf aller Werte, die den Durchschnitt abbilden. Gleichzeitig liegt die erreichte Performance 9% bis 12% unter der bei den betrachteten Werten maximal möglichen. Die Werte wurden gerundet. Die exakten Werte und ihre Herkunft sind Anhang C zu entnehmen. Auffällig wird bei einem Vergleich der Aktienperformance und ontologiegestützter Bewertung, dass Bankwerte im

⁶Es wurden die im Corpus in signifikanter Häufigkeit vorkommenden Aktien gewählt, d.h. mindestens fünfzehn Erwähnung in Verbindung mit Instanzen von *Indication*

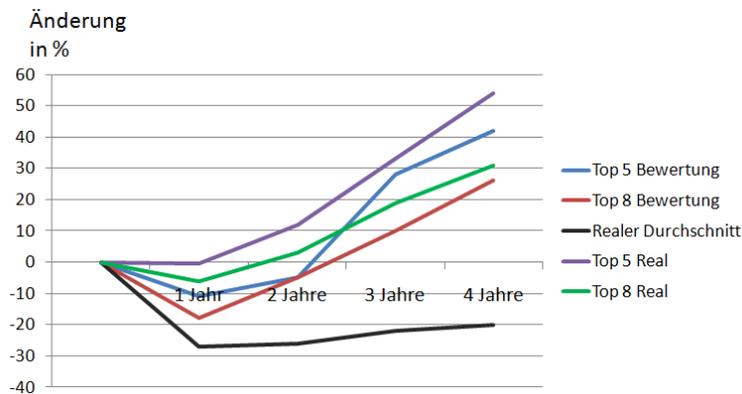


Abbildung 4.2: Wertentwicklung der Aktien im Vergleich hinsichtlich durchschnittlicher Entwicklung, Entwicklung der besten nach Ranking und Entwicklung der real besten Aktien.

Vergleich zur Bewertungslage relativ schlecht abschnitten. Wie in Abschnitt 4.1 bei Frage 32 gezeigt, ließ sich das schlechte Abschneiden des Bankensektors durch eine Ontologieabfrage nach Branchen erkennen. Diese Erkenntnis ist in das Ranking der Aktien nicht mit eingeflossen und hätte das Ergebnis noch weiter verbessern können.

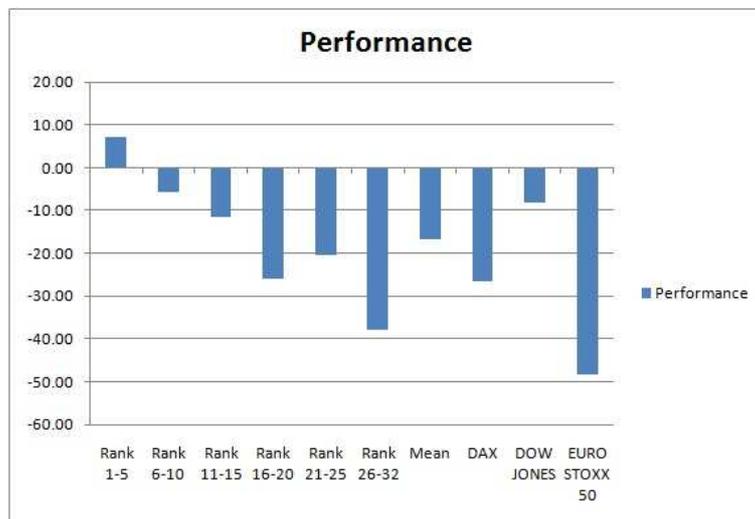


Abbildung 4.3: Anwendung der Auswahlmethodik auf einen 2. Dokumentenkörper. Darstellung der Wertentwicklung im Zeitraum 7/2007 - 7/2011.

Die Anwendung der Auswahlmethodik mittels Quant3 auf einen weiteren Corpus führte zu ähnlich guten Ergebnissen wie in Abbildung 4.3 erkennbar ist. Der dort benutzte Testcorpus war mit 14000 Dokumenten deutlich größer und umfasste eine größere Bandbreite an Dokumenten. Während der erste Datensatz ausschließlich aus professionellen Analysen hinsichtlich gewisser Unter-

nehmen sowie Unternehmens- oder wenigstens Branchenspezifischer Nachrichten besteht, beinhaltet der zweite Datensatz auch Foren- und Blogeinträge und allgemeine Marktberichte.

Kapitel 5

Fazit und Ausblick

Die Vorteile von automatischen Trenderkennungsverfahren wurden bereits eingangs in 1 und 1.2 beschrieben. Von höheren Gewinnspannen durch bessere Vermarktungsstrategien über Verhaltensdaten für die Sozialforschung bis hin zum politischen Entscheidungshelfer sind die Möglichkeiten weitreichend. Offen blieb bisher die Betrachtung der Möglichkeiten wie auch der Probleme von Trendontologien.

5.1 Möglichkeiten

Einer der grundlegenden Vorteile von Trendontologien sind ihre vielfältigen Anwendungsmöglichkeiten und vielseitigen Implementierungen. Durch die Trendontologie können nicht nur Trends aus Dokumenten extrahiert, sondern im Fall des Börsenhandels auch fundamentale Kenndaten direkt in Verbindung mit ihren Konzepten gespeichert werden. Hierdurch lassen sich zum Beispiel kombinierte Abfragen nach Themenfeldausrichtung, Präsenz im öffentlichen Diskurs, sowie Kennzahlen wie z.B. das KGV¹ realisieren. Dadurch lassen sich prinzipiell die Datenbanken für Kennzahlen in die Trendontologie integrieren. Die Zuordnung von Wörtern zu Konzepten und das hierarchische und assoziative Netz der Ontologie Definition ermöglicht weiterhin:

- Filtern nach verschiedenen Konzepten: Länder, Firmen, Gruppen, Parteien.
- Einordnung in Themenbereiche. Sortierung und Abfrage in bestimmten Bereichen ist möglich. Ausgabe kann anzeigen, welche Themenbereiche vor allem im Diskurs auftreten. Es muss nicht bekannt sein, was untersucht wird, siehe 4.1
- Bei Nutzung in verschiedenen Domänen gleichzeitig können diese differenziert betrachtet werden, ohne dass ihre Stellung zueinander vernachlässigt wird. Manuelle Definition von Verwandtschaften, Wirkungen etc. ist möglich, siehe 2.2
- Implementierung von Ähnlichkeiten verschiedener Konzepte in verschiedenen Domänen. Siehe Beispiel zu Indication in 2.2

¹Kurs-Gewinn-Verhältnis

- Großes ungenutztes Potenzial durch die Möglichkeit von verbesserten semantischen Analysen. Je mehr Semantik erkannt wird, desto mehr kann die Ontologie speichern und bei der Ausgabe verwertet werden.

Im Gegensatz zu statistischen Analysen sind Trendontologien noch weitestgehend unerforscht und stehen am Anfang ihrer Entwicklung. Allgemein ist – wie bereits eingangs erwähnt – die maschinelle Trenderkennung ein junges Thema, dessen Potenzial noch nicht ausgeschöpft wird.

5.2 Probleme

Die Nachteile von Trendontologie liegen in den hohen Anforderungen an Ontologie und Textanalyse, die für aussagekräftige Aussagen nötig sind. Während ihr Nutzen ebenso wie ihre Vielseitigkeit potentiell äußerst groß sind, muss viel Arbeit investiert werden, um die Möglichkeiten von Trendontologien auszuschöpfen. Für einen bestimmten Anwendungsbereich eine einfache Trendontologie zu erstellen ist kein großes Problem. Um allerdings die Möglichkeiten von domänenübergreifenden Abfragen mit guten automatischen Textanalysen zu verbinden, wird nicht nur eine Trendontologie benötigt, sondern eine Metaontologie mit einer Vielzahl von domänenspezifischen und ggf. anwendungsspezifischen Ontologien. Diese Ontologien sollen gemeinsam nutzbar und in ihren Ergebnissen vergleichbar sein um interdisziplinäre Aussagen zu bearbeiten und interdisziplinäre Ergebnisse zu erhalten. Um Dokumente in die Ontologie zu integrieren müssen mindestens Keywords identifiziert und implementiert bzw. automatische Extraktionen und Zuordnungen von Keywords zu Konzepten erstellt werden. Beides wird in vielen Fällen domänenspezifisch zu handhaben sein. Selbiges gilt für weiterführende semantische Analysemethoden wie die Erkennung von Aktionen und Auswirkungen auf Konzepte und Stimmungslage. Weiterhin werden äußerst große Datenmengen benötigt, sodass der Rechenaufwand nicht unerheblich ist. Da das Einbetten von Konzepten in Themen und die aufgespannten Themenfelder dadurch Ergebnisse liefern, dass sich wichtige Muster herauskristallisieren, braucht man sehr viele Daten um zufällige Zusammenhänge und falsche Interpretationen zu verhindern.

5.3 Fazit

Die Semantic Field Theory bietet die Möglichkeit mit einfachen Methoden komplexe Strukturen und Zusammenhänge darzustellen. Die Metaontologie produziert mit Hilfe dieser Theorie ein Netz der thematischen Einbettung und stellt dafür viele verschiedene Möglichkeiten der Einbindung in Themenfelder bereit, ohne diese zu erzwingen. Damit unterstützt die Metaontologie komplexe und detaillierte Verbindungen zur Auswertung detaillierter Beziehungen, ebenso eine einfache thematische Einbettung mit simplen Methoden. Durch Eingliederung der einschlägig bekannten Ontologien SKOS und OWL-Time in die Metaontologie steht diese auf einer breiten Basis und genügt den Anforderungen von Einfachheit, Kompatibilität und Erweiterbarkeit. Die Evaluation anhand von Kompetenzfragen ist erfolgreich verlaufen (siehe 4.1). Die Metaontologie stellt die Möglichkeit der Beantwortung aller Fragen zur Verfügung (siehe Anhang

B). Der nächste Schritt der Evaluierung ist ein Vergleich mit statistischen Daten über dem Textkorpus. Die Möglichkeiten von Trendontologien sind groß, die Anforderungen an Textanalyse und Ontologie, und der Zeit- und Kostenaufwand allerdings auch. Trendontologien bieten bisher in der Trenderkennung unbekannte Möglichkeiten, Merkmale verschiedener Bereiche zu verknüpfen und ihre gemeinsame Wirkungsweise automatisch zu analysieren. Um das volle Potenzial von Trendontologien nutzen zu können, wird, wie in 5.2 erwähnt, eine ganze Reihe von spezialisierten Ontologien benötigt, die mit der Metaontologie und damit miteinander kompatibel sind. Hierdurch können simple Abfragen mit einem hohen Variantenreichtum und vielseitigen Spezialisierungsmöglichkeiten realisiert werden.

Hochspezialisierte Trendontologien für eingeschränkte Anwendungsbereiche, wie die in 3 besprochene Ontologie, können dagegen auf Grund ihres hohen Automatisierungsgrades mit relativ geringem Aufwand erstellt und genutzt werden und ermöglichen in diesem Bereich die Erstellung von präzisen Prognosen (siehe 4.2). Arbeitsbedarf besteht unter anderem in der maschinell gestützten Entwicklung von domänenspezifischen Trendontologien, um die Entwicklung wie auch den Angliederungsprozess von spezialisierten Ontologien zu beschleunigen und zu vereinfachen. Eine eingehendere Untersuchung der praktischen Möglichkeiten von Trendontologien und ihres Mehrwertes wie auch ihrer Nachteile gegenüber anderen Trenderkennungsverfahren ist zur Abschätzung des Kosten-Nutzen-Verhältnisses wünschenswert. Weiterhin würde eine Zusammenarbeit mit oder Eingliederungen von Textanalyseverfahren wie z.B. Open Calais² den potenziellen Informationsgehalt signifikant erhöhen. Automatisierte Bewertungsmuster zur Einordnung von Konzepten in Bewertungskategorien bezüglich bestimmter Themen oder Domänen bergen das Potenzial der direkten Abfrage von positiven oder negativen Einflüssen hinsichtlich eines Themas. Letztlich bleibt die Frage nach Rechenaufwand und Speichereffizienz.

Es besteht noch erheblicher Forschungsbedarf im Bereich der automatischen Trenderkennung, was sich unter anderem an den in letzter Zeit häufiger erscheinenden Publikationen erkennen lässt. Besonders die ontologiegestützte Trenderkennung ist ein weitestgehend unerforschtes Feld, dessen immenses Potenzial bisher noch nicht genutzt werden kann. Wie eingangs dargestellt sind die Einsatzbereiche von Trendontologie reichen von Themen der Wirtschaft über Politik bis hin zu gesellschaftlichen Fragen und beantworten Fragen zu Meinungen, Risiken, Potenzialen, Entwicklungen und Defiziten und Vielem mehr. Gleichzeitig können Trendontologien äußerst individuell spezialisiert und eingesetzt werden, ohne ihre ursprüngliche Allgemeingültigkeit zu verlieren. Diese Möglichkeiten sollten nicht ungenutzt bleiben.

Der Report ergibt sich aus der Bachelorarbeit zum Thema "Ontologien für wissensbasierte Trenderkennung" an der Freien Universität Berlin von Lars Wiffler. Die Arbeit wurde von Olga Streibel betreut.

²<http://www.opencalais.com> (7/2011)

Literaturverzeichnis

- [1] Nikos Askitas and Klaus F. Zimmermann. Google Econometrics and Unemployment Forecasting. *Applied Economics Quarterly*, 55 (2):107–120, 2009.
- [2] Nikos Askitas and Klaus F. Zimmermann. Prognosen aus dem Internet: Weitere Erholung am Arbeitsmarkt erwartet. *Wochenbericht des DIW Berlin*, 25:402–405, 2009.
- [3] A. Constan and K. F. Zimmermann. Im Angesicht der Krise: US-Präsidentenwahlen in transnationaler Sicht. *Wochenbericht des DIW Berlin*, 44:688–701, 2008.
- [4] G. Coskun, R. Heese, M. Luczak-Rösch, R. Oldakowski, R. Schäfermeier, and O. Streibel. Towards Corporate Semantic Web: Requirements and Use Cases. Technical Report B-08-09, Freie Universität Berlin, 2008.
- [5] Oxford Dictionary. Definition for trend, 6/2011. <http://oxforddictionaries.com/definition/trend>.
- [6] Jeremy Ginsberg, L. Brammer, L. Brilliant, M. H. Mohebbi, R. S. Patel, and M. S. Smolinski. Detecting Influenza Epidemics using Search Engine Query Data. *Nature*, 457:1012–1014, 2009.
- [7] N. Guarino. *Formal Ontology and Information Systems*. IOS Press, 1998.
- [8] U. Hellinger. Trend Detection in formal models of dynamic domains. Technical report, Universität Karlsruhe, 2007.
- [9] K. A. Kholodilin and B. Silverstovs. Geben Konjunkturprognosen eine gute Orientierung? *Wochenbericht des DIW Berlin*, 13:207–214, 2009.
- [10] A. Lehrer. *Semantic fields and lexical structure*. North-Holland linguistic series. North-Holland, 1974.
- [11] S. Levermann. *Der entspannte Weg zum Reichtum*. Carl Hanser Verlag, 2010.
- [12] N. Manikandan, M. Sakthi, P. reddy, and P. Venkata. TDPA: Trend Detection and Predictive Analytics. *International Journal on Computer Science and Engineering*, 3:1033–1039, 2011.
- [13] Olga Streibel and Malgorzata Mochol. Trend Ontology for Knowledge-based Trend Mining on textual Information. In *7th International Conference on Information Technology*, 2010.

- [14] Mari Carmen Suárez-Figueroa, Asunción Gómez Pérez, and Mariano Fernández López. *NeOn Methodology for Building Ontology Networks: Specification, Scheduling and Reuse*. PhD thesis, Universiti Putra Malaysia, 2010.
- [15] Hendrik Vejlgaard. *Anatomy of a Trend*. McGraw-Hill, 2007.
- [16] World Wide Web Consortium. Skos primer, 6/2011. <http://www.w3.org/TR/skos-primer/>.
- [17] World Wide Web Consortium. Time ontology reference, 6/2011. <http://www.w3.org/TR/owl-time>.
- [18] Qing Zhang, Chaoyi Pang, Qing Xie, S. Mcbride, D. Hansen, and Yanchun Zhang. ADA: An online trend pattern detection system. In *2nd International Conference on Software Engineering and Data Mining*, 2010.
- [19] K. F. Zimmermann. Schadensbegrenzung oder Kapriolen wie im Finanzsektor. *Wirtschaftsdienst*, 12, 2008.
- [20] K. F. Zimmermann. Prognosenkrise: Warum weniger manchmal mehr ist. *Wirtschaftsdienst*, 2, 2009.

Anhang A

Ontology Specification Requirements Document

A.1 Zweck

Das Ziel der Trendontologie ist eine standardisierte Basis zur (textbasierten) Trenderkennung. Die Ontologie soll allgemein die automatische Trenderkennung auf Texten (im Web) unterstützen. Ein Trend in Texten ist ein aufkommendes Thema, dessen Relevanz und Einfluss sich in einer bestimmten Zeitspanne verändert.

A.2 Anwendungsbereich

Es soll ein gemeinsames Wissensmodell der textbasierten Trenderkennung geschaffen werden, dass unabhängig von Domänen und Zeitparametern valide ist. Domänenspezifisches Wissen soll einfach in das allgemeine Modell eingebunden werden können.

A.3 Grad der Formalisierung

Die Formalisierung der Ontologie sollte so gering wie möglich sein, um einen möglichst großen Freiraum für Entwickler von zu integrierenden Ontologien zu lassen, wobei genug Formalität gewahrt sein muss, um als standardisierende Basis fungieren zu können. SKOS-Kompatibilität ist das bevorzugte Level der Formalisierung. Weiterhin wird für die Realisierung themenfelderorientierter Ontologien Klassen als Instanzen angesehen werden, was nur in OWL Full möglich ist.

A.4 Zielgruppen

- Direkte Nutzung (alleinstehend):
 - Ontologiedesigner, die die Ontologie als Basis für eigene domänenspezifische Trendontologien benutzen.

- Ontologiedesigner, die verschiedene domänenspezifische Trendontologien gemeinsam benutzen möchte und dafür eine standardisierte Basis benötigen.
- Indirekte Nutzung (in Verbindung mit domänenspezifischen Ontologien):
 - Kommerzielle Nutzung im Bereich der Trenderkennung, in dem durch maschinelle Textanalyseverfahren eine Aussage über zukünftige Bewegungen eines Themas getroffen und Trendontologien zur Verbesserung des Ergebnisses eingesetzt werden.
 - Trendforscher, die auf Basis von erkannten Trends die Ergebnisse evaluieren und Techniken der Trenderkennung verbessern. (Eingeklammert, da diese Nutzungsform nicht zur zentralen Aufgabe gehört. Es geht hier um die Idee den Trend selbst in der Ontologie zu definieren und alte Trends zu speichern und zu evaluieren.

A.5 Vorgesehene Nutzung

- Richtlinie zur Entwicklung von Trendontologien, sodass diese später austauschbar oder auch gemeinsam benutzt werden können.
- Basis zur Integration verschiedener Trendontologien.
- In Kombination mit domänenspezifischen Ontologien:
 - Wissensbasis zur Verbesserung und Vereinfachung der textbasierten automatischen Trenderkennung.
 - Wissensbasis zur statistischen und logischen Evaluierung des Systems.

A.6 Kompetenzfragen

Trend

- 1 Welche Konzepte sind relevant für einen Trend?
- 2 In welche Richtung geht der Trend (wichtiger/unwichtiger)?
- 3 Wie stark ist der Trend?
- 4 Was sind die Auswirkungen des Trends?
- 5 Auf welche Konzepte wirkt der Trend?
- 6 Wann ist der Trend?
- 7 Wie lang ist der Trend?

Konzept

- 8 Welche Konzeptidentifizier (keywords) werden für bestimmte Konzepte in bestimmten Domänen verwendet?
- 9 Wie relevant ist ein Konzept allgemein?
- 10 Wie relevant ist ein Konzept in einem speziellen Fall?
- 11 Beeinflusst ein Konzept den Trend positiv oder negativ?
- 12 Welche Konzepte wirken auf einander?
- 13 Welche Konzepte sind verwandt?
- 14 Welche Konzepte können unter einem gemeinsamen Oberbegriff (Gruppe) zusammengefasst werden?
- 15 Welche Konzeptidentifizier (keywords) sind Synonym?
- 16 ... sind Abkürzungen?

Stimmung

- 17 Wie wirkt sich ein Ereignis auf die Stimmung interessierter Personen aus?
- 18 Wie wirkt sich eine Stimmung auf einen Trend aus?

Zeit

- 19 Wie ist die zeitliche Perspektive (Lifecycle)?

Sicherheit/Vertrauen

- 20 Wie gesichert ist eine Aussage? <-> Wieviel Risiko birgt sie?
- 21 Wieviel Vertrauen wird in die Quelle gesetzt?

Interesse/Relevanz

- 22 Wie groß ist das öffentliche Interesse am Thema (wie viel wird es diskutiert)?

Finanzsektor (in A.7 nicht enthalten)

- 23 Beeinflusst ein Ereignis einen Aktienindex positiv (steigt) oder negativ?
- 24 Wie relevant ist die Bewegung des Indexes für die Auswirkung des Ereignisses auf die Weltwirtschaft?
- 25 Welche Menschengruppen werden vom Ereignis beeinflusst?
- 26 Welche Branchen verlieren durch das Ereignis, welche gewinnen?
- 27 Welche Konzerne gehören zu diesen Gruppen?
- 28 In welchem Index sind sie vertreten?

- 29 Beeinflusst das Konzernergebnis den Aktienkurs?
- 30 Gibt es eine Finanzblase oder sind die Strategien vorsichtig?
- 31 Wie ist die Bewertungslage eine Aktie?
- 32 Wie ist die Bewertungslage einer Branche?

A.7 Glossar der Begriffe

Oberbegriff	Begriff	Umsetzung in der Metaontologie	Vorkommen
Konzept		Player	14
	Auswirkung	impact	6
	Kozeptidentifier	keyword	3
	Relevanz	relevnace	3
	(Konzept)gruppe	skos:Collection	1
	Synonym/Abkürzung	alternateKeyword	2
	Verwandschaft	related/countableRelated	1
Trend			7
	Trendrichtung	Indication	2
	Stärke	Amplifier	1
Zeit			2
	Dauer	time:DurationDescription	1
	Zeitpunkt	time:DateTimeDescription	1
Stimmung		Sentiment	3
	Person	foaf:Person/Agent	1
	Ereignis	Incident	1
Risiko		Risk	2
	Quelle	Source	1
	Aussagekraft	Source impact Player	1
	Vertrauen	Trauen	2
Interesse		Public relevance	2
	Öffentlichkeit	Public	2

Anhang B

Kompetenzfragen und korrespondierende Abfrage

Die folgende Tabelle stellt die Kompetenzfragen dar und durch welche Abfragen die Antworten für diese Fragen abgelesen werden können. Die letzte Spalte stellt die Antwort da. Die Aufstellung ist vergleichbar mit eine SPARQL-Abfrage¹; die zweite Spalte korrespondiert mit der Abfrage (Where-Clause), die letzte Spalte beinhaltet das, was ausgegeben und betrachtet wird (Select-Clause).

	Kompetenzfrage	Abfrage nach..	Indikator für Antwort
1	Welche Konzepte sind relevant für einen Trend?	Trendzeitraum + Kernkonzepte	In Beziehung stehende Konzepte, die in dem Zeitraum häufig vorkommen
2	In welche Richtung bewegt sich der Trend (wichtiger/unwichtiger)?	Trendzeitraum + Kernkonzepte	Änderung von Membercount (Häufigkeit des Vorkommens)
3	Wie stark ist der Trend	Trendzeitraum + Kernkonzepte	Membercount (Häufigkeit des Vorkommens)
4	Was sind die Auswirkungen des Trends	Trendzeitraum + Kernkonzepte	Einbettung trendrelevanter Konzepte und nachfolgende Änderung
5	Aus welche Konzepte wirkt der Trend	Trendzeitraum + Kernkonzepte	Einbettung trendrelevanter Konzepte
6	Wann ist der Trend	Kernkonzepte	Zeitraum eines signifikanten Einbettungsgrades relevanter Konzepte hinsichtlich Zeitpunkt
7	Wie lang ist der Trend	Kernkonzepte	siehe 6 hinsichtlich Zeitdauer

¹<http://www.w3.org/TR/rdf-sparql-query> (5/2012)

8	Welche Konzeptidentifizierer (keywords) werden für bestimmte Konzepte in bestimmten Domänen verwendet.	Konzept mit keyword-Relation und Keyword in entsprechender Domänenontologie	Keyword
9	Wie relevant ist ein Trend allgemein	Konzept + countable Related	Count
10	Wie relevant ist ein Konzept in einem speziellen Fall	Konzept + für den Fall relevante Konzepte + countableRelated	Count
11	Beeinflusst ein Konzept den Trend positiv oder negativ	Konzept + Trendkernkonzepte	Role
12	Welche Konzepte wirken aufeinander	Konzept + Relational	verbundene Konzepte
13	Welche Konzepte sind verwandt	Konzept + related/countableRelated	verbundene Konzepte
14	Welche Konzepte können unter einem Oberbegriff zusammengefasst werden		Häufig vorkommende Schnittmengen von skos:Collection
15	Welche keywords sind Synonym	keyword	alternativeKeyword
16	..sind Abkürzungen	s.o.	s.o.
17	Wie wirkt sich ein Ereignis auf die Stimmung interessierter Personen aus?	Incident + relevantPersons	Häufige Sentiments
18	Was bewirkt die Stimmung in Bezug auf den Trend	Sentiment + Kernkonzepte + Trendzeitraum	Relational
19	Wie ist die zeitliche Perspektive	Player	Lifecycle
20	Wie gesichert ist die Aussage	Source	Risk
21	Wieviel Vertrauen wird in die Quelle gesetzt?	Source	Trust
22	Wie groß ist das öffentliche Interesse am Thema (wie viel wird es diskutiert)?	Topic	Anzahl an Verbindungen (countable Related)
23	Beeinflusst ein Ereignis einen Aktienindex positiv (steigt) oder negativ?	Incident	Index + Role

24	Wie relevant ist die Bewegung des Indexes für die Auswirkung des Ereignisses auf die Weltwirtschaft?	Incident + Weltwirtschaft	Stärkste Änderungen an Konzeptverhältnissen im Vergleich zu durchschnittlichen Werten der letzten Jahre
25	Welche Menschengruppen werden vom Ereignis beeinflusst?	Incident	Häufig auftretende Gruppen
26	Welche Branchen verlieren durch das Ereignis, welche gewinnen?	Incident + Branch/Country	Company + Role
27	Welche Konzerne gehören zu diesen Gruppen?	Branch/Country	Company
28	In welchem Index sind sie vertreten?	Company	Index
29	Beeinflusst das Konzernergebnis den Aktienkurs?	BusinessVolume + Company	Kursentwicklung bei Veröffentlichung von BusinessVolume
30	Gibt es eine Finanzblase oder sind die Strategien vorsichtig	Indikatoren	Gesamtverhältnis der Indikatoren (Gute Bewertung -> Euphorie, schlechte -> Pessimismus)
31	Wie ist die Bewertungslage einer Aktie	Company	Verhältnis der Indikatoren
32	Wie ist die Bewertungslage einer Branche	Branche	Verhältnis der Indikatoren

Anhang C

Aktienkennzahlen und Berechnung

Die nachfolgende Tabelle stellt kompakt die Daten dar, auf deren Basis die Ranking-Berechnung der Aktien in Abschnitt 4.2 vorgenommen wurden. Die Spalten "Positiv", "Neutral" und "Negativ" geben die Häufigkeit der der Ontologie entnommenen Indikatoren an. Die Zuordnung erfolgte nach dem in 4.2 vorgestellten Schema auf Basis des 1. Datensatzes. Die Indikator-Werte wurden mit Eins initialisiert um Divisionen durch Null zu vermeiden. Die Kurswerte sind www.comdirect.de entnommen und wurden gerundet. Die hinteren Spalten geben das Ergebnis der Quant-Berechnung nach 4.2 an, wobei die normierten Spalten in Prozent des höchsten vorgekommenen Wertes angegeben werden. Die Tabelle ist nach der realen Wertentwicklung über vier Jahre sortiert.

	Positiv	Neutral	Negativ	Kurs 7/2007	Kurs 7/2008	Kurs 7/2011	Performance 1Jahr	Performance 4 Jahre	Quant 3	Quant 1 norm	Quant 1	Quant 2 nomi	Quant 2
Apple	30	6	4	92	109	230	18.48	150.00	8.22	35.39	15.75	7.59	11.30
Orade	17	2	1	14.9	13.4	22.2	-10.07	48.99	59.83	78.65	35.00	100.00	148.81
Starbucks	15	8	2	19.5	10.7	26.3	-45.13	34.87	7.81	38.20	17.00	6.07	9.03
Lanxess	21	5	1	41	28	55	-31.71	34.15	43.95	100.00	44.50	62.90	93.61
Fraport	1	18	4	53.5	43.1	54.2	-19.44	1.31	0.84	6.18	2.75	0.14	0.21
Novartis	37	3	2	41	37	41	-9.76	0.00	50.50	84.83	37.75	79.80	118.76
Meditec	10	12	4	16	10.3	15.4	-35.63	-3.75	2.24	14.61	6.50	0.83	1.24
Google	14	6	2	390	377	335	-3.33	-14.10	7.72	34.83	15.50	6.73	10.01
Ebay	33	11	6	24	17	20	-29.17	-16.67	4.75	26.78	11.92	2.81	4.18
Deutsche Telekom	36	30	4	13.8	10.5	10.4	-23.91	-24.64	7.85	48.88	21.75	3.48	5.18
Thyssen	6	8	4	44	39	33	-11.36	-25.00	1.37	8.99	4.00	0.48	0.72
Daimler	23	3	4	67	40	49	-40.30	-26.87	7.53	26.69	11.88	8.39	12.48
Adva	9	15	2	7.3	1.5	4	-79.45	-45.21	4.49	28.65	12.75	1.82	2.71
Yahoo	13	15	4	19.9	14.2	10.6	-28.64	-46.73	2.90	18.82	8.38	1.10	1.63
Allianz	34	79	6	173	111	92	-35.84	-46.82	5.78	40.26	17.92	1.50	2.23
Porsche	6	9	2	113	78	53	-30.97	-53.10	2.95	18.54	8.25	1.27	1.89
E.ON	16	21	6	41	42	19	2.44	-53.66	2.36	15.92	7.08	0.75	1.11
CentroSolar	36	32	12	11	9.1	4	-17.27	-63.64	2.43	16.48	7.33	0.75	1.11
Solarworld	9	17	4	37.5	30.3	9.2	-19.20	-75.47	2.19	14.89	6.63	0.67	0.99
Commerzbank	24	28	14	28.4	15.7	3	-44.72	-89.44	1.44	9.95	4.43	0.40	0.60
Citigroup	39	8	10	383	108	27.5	-71.80	-92.82	3.28	18.43	8.20	1.95	2.90