

Projektbericht

Technologien des Wissensmanagements

Prof. Dr.-Ing. Axel Hahn & Dipl.-Inform. Jürgen Geuter

Sommersemester 2013

Ausgearbeitet von

Daniel Fay, Bernd Hermes, Shimal Ibrahim

Matrikel-Nr.

(1952327), (1986957), (1091507)

Abgabe

Oldenburg, den 30.06.13

Inhaltsverzeichnis

1. Einleitung.....	3
2. Das Konzept.....	3
User und Artikel.....	3
Änderungen	4
Versionen.....	5
Kategorisierung.....	6
Grafische Benutzeroberfläche	6
3. Analyse des Artikels	7
Welchen Arten von Änderungen gibt es?.....	7
Welche Muster aus Änderungen?.....	8
4. Analyse des Users (Benutzers).....	9
Welche Artikel ändert ein User?	9
Wie gehören diese Artikel des Users zusammen?	9
Welche Arten von Änderungen führt der Benutzer durch?.....	10
Kann man Typen von Nutzern erkennen?.....	11
5. Analyse der Daten.....	11
Welche Nutzer sind sich ähnlich?.....	11
Wie kann man Kooperationsfähigkeit zweier Nutzer messen?	12
Wie kann man die Reputation von Nutzern bewerten?	13
6. Sonstige Datenanalyse	13
Expertensuche: Wer kennt sich mit „Berlin“ und „Kaffee“ aus?	13
Zeitgeist: Wie ändert sich die Nutzung eines Begriffs in der Datenbank über die Zeit? ..	14

1. Einleitung

Die vorliegende Seminararbeit im Rahmen der Veranstaltung Technologien des Wissensmanagements führt die einzelnen Entscheidungen und Ideen auf, die während der Erarbeitung des Modells und darüber hinaus besprochen wurden. Die Aufgabe bestand in der Entwicklung eines Softwaremodells, welches u.a. Analysen ermöglichen sollte und diese Analysen entsprechend visualisiert darzustellen. Die relevanten Konzepte mussten zudem extrahiert und modelliert werden. Die Datenbank, die zur Verfügung gestellt wird, ist die Wissensdatenbank Wikipedia.

Die Informationen sollten in einer MySQL-DB abgelegt, der Zugriff und das Abspeichern der benötigten Wikipedia-Informationen sollte durch eine API-Implementation hergestellt werden. Die Analysen wären anschließend bereits durchführbar, jedoch wäre dies noch nicht visuell. Die Entscheidung zur Erstellung einer GUI, grafische Benutzeroberfläche war somit gefallen.

Zur Überprüfung der Analysen wird ein Java-Programm entwickelt, welches zur Beantwortung der Aufgabenstellung eingesetzt wurde. Jedes Mitglied beschäftigte sich mit einer Thematik. Neben der API-Implementation, ist eine MySQL-Umsetzung und eine Vergleichsbetrachtung umgesetzt worden. Mit Beendigung der ersten beiden Teile, wurde der Vergleichsteil vermehrt besprochen. Um die gestellten Aufgaben durchzuführen, sind anschließend vordefinierte Artikel, Kategorien und Benutzer vorausgewählt worden.

Im Kapitel 2 Unsere Konzepte werden die relevanten Daten für die Analysen sowie die GUI erläutert. Mit diesem Werkzeug und den Informationen werden im Anschluss die Kapitel 3 bis 6 beantwortet.

2. Das Konzept

Zur Bearbeitung der Projektaufgabe wird ein Konzept konzipiert, das folgende Ideen und Inhalte umsetzen wird.

User und Artikel

Für die Beantwortung der weiteren Fragen sind nachfolgend diese Benutzer ausgewählt worden:

- APPER, AHZ , Aka
- Belsazar
- Kaisersoft (Kaisersaft ist nicht vorhanden)

- Weissbier

Die Artikel, welche im weiteren Verlauf betrachtet werden sind die folgenden:

- Homöopathie
- Christian_Lindner
- U2 (Band) (U2_Band wird nicht gefunden)
- Prism_Überwachungsprogramm
- Datenschutz und Deutschland
- Hacker und Erfrischungsgetränk

Änderungen

Ein Benutzer hat bei Wikipedia mehrere Möglichkeiten bei der Artikelverwaltung. Er kann einen bestehenden Artikel ändern, er kann einen weiteren neuen Artikel hinzufügen und er kann einen Artikel zur Löschung freigeben. Im letztgenannten Fall wird die Löschung durch einen Administrator durchgeführt. [Quelle: Wikipedia]

Die Artikeländerungen können von unterschiedlichen Typen sein. Welche Typen im Rahmen dieses Projektes generiert wurden, werden im Verlauf dieser Arbeit erläutert.

Das Konzept zur Feststellung von Änderung ist wie folgt implementiert:

Generell besitzt ein Artikel n-Revision. Beim Erreichen von n, werden die Revisionen n sowie n-1 betrachtet und miteinander verglichen. In der nachfolgenden Abbildung 1 ist ein Beispiel einer Revisionsstruktur aufgezeigt.

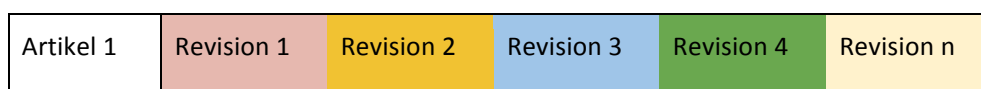


ABBILDUNG 1: ARTIKEL-REVISIONEN

Die Revisionen haben für die Analyse jeweils nachstehende Eigenschaften:

- maximale Anzahl an Zeilen
- maximale Anzahl an Wörter
- eine Liste von einzigartigen Zeilen (Zeilen, die nicht in der anderen Vergleichsrevision enthalten sind)
- eine Liste von einzigartigen Wörtern (Wörter, die nicht in der anderen Vergleichsrevision enthalten sind)

Werden jetzt zwei Revisionen miteinander verglichen, kann dabei die Anzahl an einzigartigen Zeilen und Wörtern festgestellt werden. Das funktioniert wie folgt:

Schritt 1: Zunächst werden die Zeilen miteinander verglichen sowie auf Ähnlichkeiten geprüft. Falls sie identisch sind, werden diese aus den beiden zu vergleichenden Revisionen entfernt (vgl. Abb. 2).

Artikel n	Revision n-1	Zeilen	Wort 1	Wort 2	Wort 3	Wort 4	Wort 5
			Wort 1	Wort 2	Wort 3	Wort 4	Wort 5
			Wort 1	Wort 2	Wort 3	Wort 4	Wort 5
			Wort 1	Wort 2	Wort 3	Wort 4	Wort 5
	Revision n	Zeilen	Wort 1	Wort 2	Wort 3	Wort 4	Wort 5
			Wort 1	Wort 2	Wort 3	Wort 4	Wort 5
			Wort 1	Wort 2	Wort 3	Wort 4	Wort 5
			Wort 1	Wort 2	Wort 3	Wort 4	Wort 5

ABBILDUNG 2: UNTERSCHIEDLICHE REVISIONSZEILEN EINES ARTIKELS

Schritt 2: Die nicht übereinstimmenden Zeilen beider Revisionen werden genommen und gemeinsam Wort für Wort verglichen. Werden identische Wörter gefunden, so werden diese aus den Zeilen der Revisionen entfernt (vgl. Abb. 3).

Artikel n	Revision n-1	Wort 1	Wort 2	Wort 3	Wort 4	Wort 5	Wort 6
		Wort 1	Wort 2	Wort 3	Wort 4	Wort 5	Wort 6
	Revision n	Wort 1	Wort 2	Wort 3	Wort 4	Wort 5	Wort 6
		Wort 1	Wort 2	Wort 3	Wort 4	Wort 5	Wort 6

ABBILDUNG 3: UNGLEICHE WÖRTER DER EINZELNEN REVISIONEN

Schritt 3: Wie viele "einzigartige" Zeilen und Wörter der Artikel n besitzt, wird an dieser Stelle gezählt. Im vorliegenden Fall wurden bei einer Revision n-1 keine (einzigartigen) Zeilen und keine (einzigartigen) Wörter gefunden. Hingegen sind bei der nächsten Revision n, zwei Zeilen und sieben Wörter gefunden worden.

Versionen

Sowohl die letzte wie auch die vorletzte Revision werden betrachtet. Im Blickpunkt steht zunächst immer die "ältere", also die vorletzte Revision eines Artikels. Später vergleichen wir diese mit der neueren Revision, bei der dann die Art der Änderung festgestellt wird. Näheres wird im Kapitel 3 (Analyse des Artikels) erörtert.

Kategorisierung

Die Kategorien, die zu den oben genannten Artikeln zugeordnet werden sind:

- Chaos Computer Club? Datenschutz
- Deutschland Erfrischungsgetränk Hackerkultur Homöopathie
- Hydrotherapie?
- Rockband The Dubliners U2 Irische Band
- Unternehmensberater New Economy

Grafische Benutzeroberfläche

Die Visualisierung der Analysen wird durch eine entsprechende grafische Benutzeroberfläche präsentiert, die mehrere Funktionalitäten aufweist.

Neben der Wiki-API Schnittstelle ist eine MySQL-Datenbankverbindung, sowie ein Panel für die Analysen und eine Darstellung im Form eines Graphen zu den einzelnen Artikeln ersichtlich. In der nachfolgenden Abbildung wird die GUI dargestellt.

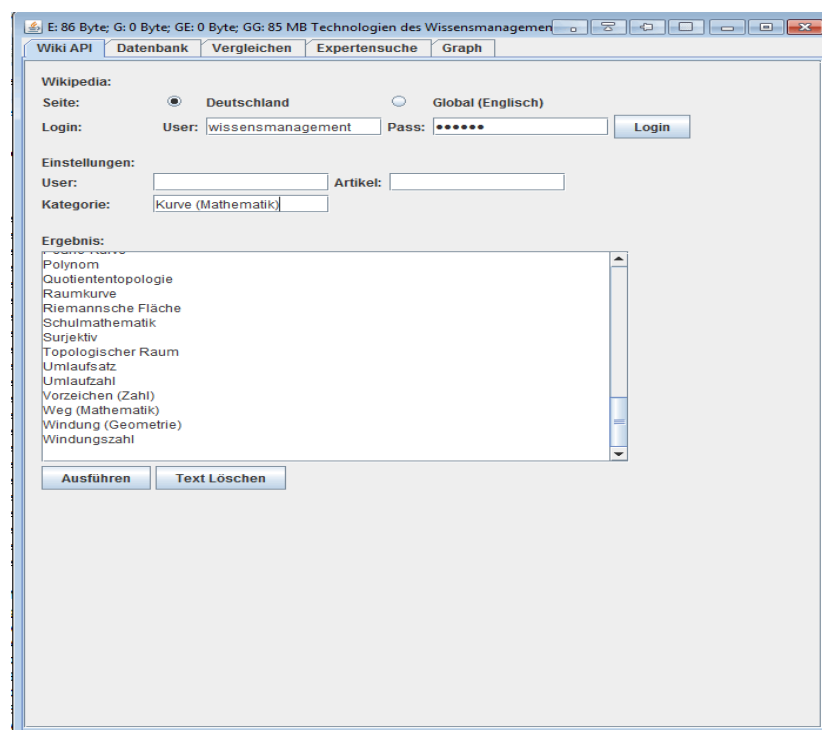


ABBILDUNG 4: GUI DER ENTWICKELTEN SOFTWARE

3. Analyse des Artikels

Welchen Arten von Änderungen gibt es?

In Wikipedia werden die zwei Typen kleine Änderung und große Änderung von Artikeln unterschieden. Diese Änderungstypen werden jedoch vom User (Autor), der die Änderung durchgeführt hat, festgelegt. Was bei den beiden Typen geändert wurde kann jedoch nur noch grafisch festgestellt werden.

Um die gestellten Fragen zu beantworten, klassifizierten wir diese beiden Arten von Änderung als ungeeignet. Daraufhin definierten wir eigene Typen von Änderungen. Insgesamt werden nach unserer Ansicht sechs Arten von Änderungen unterschieden. Im Folgenden werden diese mit jeweils einer kurzen Erläuterung aufgelistet:

- **Korrektur**

Dieser Änderungstyp wird festgestellt, wenn sowohl die Anzahl der unterschiedlichen Zeilen als auch die Anzahl der einzigartigen Wörter der neueren Revision kleiner sind als die der älteren Revision.

- **Formatierung**

Die Formatierung steht im Rahmen dieses Projektes für minimale Änderung der neueren Revision gegenüber ihre nächst älteren. Der Änderungstyp Formatierung wird festgestellt, wenn die Anzahl der unterschiedlichen Zeilen als auch die Anzahl der einzigartigen Wörter der neueren Revision gleich der älteren Revision ist, aber dennoch beide ungleich null sind.

- **Verbesserung**

Ändert ein Nutzer eine Revision und ist sie gegenüber ihrer nächst älteren um einen bestimmten Prozentsatz größer, so wird der Typ der Änderung als Verbesserung klassifiziert. Diese Verbesserung kann auf zwei Wegen festgestellt werden:

- Falls die Anzahl der unterschiedlichen Zeilen, wie auch die Anzahl der einzigartigen Wörter der älteren Revision (die als Grundlage für den Vergleich dienen) nicht Null sind. Dieser Änderungstyp wird daraufhin generiert, wenn die Menge der Änderung in der neueren Revision maximal 15 % beträgt.
- Falls die Anzahl der unterschiedlichen Zeilen, wie auch die Anzahl der einzigartigen Wörter der älteren Revision gleich Null sind, dann wird der Vergleich anhand der maximalen Zeilen und Wörter durchgeführt. Die maximalen Zeilen und Wörter sind

die Anzahl der im Revision vorhanden Zeilen und Wörter, die vor dem Ausführen der Operationen festgestellt wurden. Auch hier wird geschaut, wie groß die Menge der Änderung beträgt. Ist die Differenz der maximalen Zeilen und Wörter beider betrachteten Revisionen maximal 15 %, so handelt es sich abermals um eine Verbesserung.

- **Wissensproduktion**

Eine Wissensproduktion ist vorhanden, wenn die Anzahl der unterschiedlichen Zeilen wie auch die Anzahl der einzigartigen Wörter der neueren Revision im Vergleich zu der älteren Revision zwischen 15 und 100 % liegen. Auch hier bestehen zwei Möglichkeiten diesen Typ der Änderung festzustellen. Die beiden Möglichkeiten sind bereits in der Verbesserung beschrieben worden. Bei der Wissensproduktion wird verglichen mit der Verbesserung eine zusätzliche Prüfung vorgenommen. Es wird untersucht, ob die Differenz der Anzahl der maximalen Zeilen und Wörter der älteren und neueren Revisionen zwischen 15 und 100 % liegt.

- **Überarbeitung**

Eine Überarbeitung ist gegeben, wenn die Anzahl unterschiedlichen Zeilen als auch die Anzahl der einzigartigen Wörter der neueren Revision gegenüber ihre älteren über 100 % liegen. In gleicher Weise wie zuvor bei den beiden Typen Verbesserung und Wissensproduktion, können zwei Möglichkeiten zur Generierung des Typen Überarbeitung genutzt werden.

- **Keine Änderung**

Keine Änderung wird festgestellt, falls folgende Größen der betrachteten Revisionen gleich groß sind:

- Anzahl der einzigartigen Wörter
- Anzahl der unterschiedlichen Zeilen
- Anzahl der maximalen Zeilen
- Anzahl der maximalen Wörter

Welche Muster aus Änderungen?

Die Muster aus Änderungen ergeben sich aus den Typen der Änderungen. Es kann bei einem Artikel über die Zeit ein chronologischer Ablauf der Änderungsarten festgestellt werden. Ein Artikel wird erstellt und enthält unmittelbar danach eine erste Revision. Diese Revision ist die Basisrevision und dient zum Vergleich mit der nächsten Revision.

Die Revision ist am Anfang keinen Änderungstypen zugeordnet. Die darauf folgende Revision kann, im Vergleich zu ihrer älteren Revision, einen Änderungstypen zugeteilt sein. Dieser kann z.B. Verbesserung lauten. Wiederrum kann eine weitere untersuchende Revision den Änderungstyp Wissensproduktion, Überarbeitung oder Korrektur aufweisen. Mit der Zeit werden die Änderungen immer schwächer und können aus diesem Grunde den Typ Formatierung und Verbesserung zugeordnet werden.

4. Analyse des Users (Benutzers)

Diese Fragen können durch entsprechende Analysen, wie zuvor bei den Artikeln erläutert, untersucht und beantwortet werden. Im Folgenden fand die Beantwortung der Fragen durch Benutzung von SQL-Befehle statt. Als Datengrundlage dienten jeweils die letzten 500 Datensätze der Artikeln oder Kategorien.

Welche Artikel ändert ein User?

Durch den SQL-Befehl `"Select Artikel from revision where User = '*' group by Artikel;"`, wobei das Sternchen für den jeweiligen User steht, wird eine Übersicht über die Artikel des jeweiligen Users gruppiert ausgegeben. Zur Demonstrationszwecken werden drei Benutzer in die Betrachtung gezogen.

Der User "AHZ" verfasste die folgenden Artikel: Homöopathie, Deutschland.

"Aka" wiederum ist an diesen Artikeln beteiligt gewesen: Christian_Lindner, Datenschutz, Deutschland und Homöopathie, U2 (Band).

Der dritte User bei dieser Betrachtung ist "Belsazar", der den Artikel Homöopathie verfasste bzw. dort Änderungen durchführte.

Wie gehören diese Artikel des Users zusammen?

Die folgenden SQL-Befehle sind genutzt worden: `Select Kategorie from kategorie where Artikel ="artikel1";`.

Die Variablen "artikel1" und "artikel2" stehen für die beiden zu betrachteten Artikel des Users. Für den jeweiligen User werden zunächst die Kategorien ausgegeben.

Welche Zusammengehörigkeiten Artikel des jeweiligen Users aufweisen, wird durch die Betrachtung der Kategorien aufgezeigt und erläutert.

Der User AHZ hat Artikel verfasst, die zu den folgenden Kategorien zugeordnet werden: Deutschland, Chaos Computer Club?, Homöopathie.

Die Kategorien Deutschland und Homöopathie besitzen eine Zusammengehörigkeit, weil die Homöopathie eine Behandlungsmethode, nach Vorstellung eines deutschen Arztes ist [Quelle: Wikipedia].

Der User Aka interessiert sich für Artikel, die den nachfolgenden Kategorien angehören: Chaos Computer Club?, Deutschland, Homöopathie und Rockband, Irische Band. Er verfasste oder änderte u.a. die Artikel Datenschutz, Deutschland und Homöopathie, die der Kategorie Deutschland am ehesten zugeordnet werden kann. Sowohl beim Chaos Computer Club, wie auch bei der Homöopathie wird auf Deutschland eingegangen. Artikel U2 (Band) gehört den Kategorien Rockband und Irische Band an und besitzt zu den erstgenannten Artikeln keine Verbindung.

Bei dem User Belsazar kann keine Aussage zur der Zusammengehörigkeit von den Artikeln gegeben werden, weil nur ein Artikel, die Homöopathie von ihm verfasst wurde.

Welche Arten von Änderungen führt der Benutzer durch?

Diese Frage kann durch eine Analyse und durch Nutzung der Änderungstypen, wie bei der Analyse des Artikels beschrieben, beantwortet werden. Es wird hier nicht weiter darauf eingegangen.

Der User hat sowohl kleine als auch große Veränderungen im Artikel durch-geführt.

Neben diesen Änderung hat stellt unsere Software , wie im Kapitel “Analyse des Artikels des Artikels” beschrieben, noch weiter Änderungsarten fest. Dabei kann man feststellen, dass folgende User folgende Änderungen durchgeführt haben:

Artikel	User	Änderung
Christian_Lindner	Kleiner Stampfi	Korrektur
Datenschutz	Mikered	Keine Änderung, da er/sie die erste Revision erstellt hat
Datenschutz	134.91.6.204	Formatierung, und Korrektur (User hat zwei Revisionen bearbeitet

ABBILDUNG 5: ARTEN VON ÄNDERUNGEN, DIE EIN USER AUSFÜHRT

Kann man Typen von Nutzern erkennen?

Wenn ein beliebiger User mehrere Artikel verfasste und diese regelmäßig bei Veränderungen der Thematik oder bei sonstigen Veränderungen durch verschiedene User aktualisiert, so kann dieser Benutzer als ein Stammmitglied von Wikipedia angesehen werden. Wenn dies über eine längere Zeit anhält, dann kann der User zudem Administratorstatus erhalten.

Aus einem anderen Winkel betrachtet, ist ersichtlich, dass bestimmte Nutzer oft eine bestimmte Art von Änderung durchführen.

In der folgenden Tabelle werden einige Beispiele aufgeführt:

Artikel	User	Änderung
Christian_Lindner	Miltrak	Korrektur (3x)
Christian_Lindner	Bluemangenius	Keine Änderung (1x) und Korrektur (Kein Typ erkennbar)
Deutschland	Horst-schlaemma	Formatierung (2x), Korrektur (1x) und Wissensproduktion (1x) (Kein Typ erkennbar)

ABBILDUNG 6: TYPEN VON NUTZER

5. Analyse der Daten

Welche Nutzer sind sich ähnlich?

Um die einzelnen Nutzer miteinander zu vergleichen, bietet die Wikipedia-API mehrere Einsatzmöglichkeiten. Zum einen lassen sich grundlegende Informationen über einem Benutzer erfahren. Zum Beispiel der Benutzername oder das Registrierungsdatum, zum anderen kann ein Benutzer auch eigene Informationen über sich bei Wikipedia hinterlegen. Zu diesen Informationen gehört das Geschlecht oder bspw. der echte Name. Wikipedia erlaubt zusätzlich das Erstellen einer eigenen Seite auf der sich der User mit seinen Tätigkeiten präsentieren kann.

Des Weiteren lassen sich von einem Benutzer die einzelnen Beiträge durch Wikipedia einzeln anzeigen. Zusätzlich lässt sich abfragen, welche Rechte ein Benutzer hat und welche Gruppe er angehört. Als Rechte werden dabei die erlaubten Tätigkeiten eines Benutzers bezeichnet, also ob er neue Seiten anlegen darf und dergleichen. Mit der

Zuordnung zu einer Gruppe werden einem User ebenfalls automatisch Rechte zugewiesen.

Anhand dieser Daten, gibt es verschiedene Ansatzpunkte mit denen User verglichen werden können. Wenn man für die Ähnlichkeit die Gruppen und Rechte in Betracht zieht, so erhält man schon einen ersten Eindruck darüber, welche Stellung die jeweiligen Benutzer bei Wikipedia besitzen. Des Weiteren lässt sich zum Beispiel anhand des Registrierungsdatums feststellen, ob zwei Benutzer in etwa gleich lange Wikipedia nutzen. Die Anzeige der irritierten Beiträge gibt weiterhin an, wie erfahren der User ist. Es kann sich beim jeweiligen Benutzern um einen erfahrenen Editor oder Neuling handeln. Zudem lassen sich über die Bestimmung der Artikelveränderungen Aussagen treffen. Dadurch können ähnliche Interessen zwischen zwei oder mehr Benutzern entdeckt werden.

Wie kann man Kooperationsfähigkeit zweier Nutzer messen?

Um die Kooperationsfähigkeit zweier Nutzer zu bewerten, muss zunächst geprüft werden, an welchen Artikeln die beiden Benutzer gemeinsam gearbeitet haben. Anschließend sind einzelnen Revisionen der jeweiligen Artikel zu betrachten. Dabei ist vorab zu beachten, dass die bearbeiteten Revisionen der User möglichst dicht aufeinanderfolgen, weil das eher für eine Kooperation spricht. Als nächstes ist bei der jeweiligen Revision zu untersuchen, ob Daten hinzugefügt oder entfernt wurden. Ist die Prüfung abgeschlossen, kann entweder eine Gemeinsamkeit bei der Bearbeitung eines Artikels zwischen zwei Usern festgestellt werden oder andererseits das Gegenteil hiervon. Zum Beispiel zeigt die Artikelgröße in Bytes an, um wie viel größer oder kleiner der Artikel von Revision zu Revision geworden ist. So zeigt zum Beispiel, wenn ein Artikel an große Mengen an Bytes mit einer Revision hinzugewonnen hat und mit der darauf folgenden Revision eine ähnliche Menge Bytes sofort wieder abgenommen hat, dass dort zwei Benutzer sich über den Inhalt nicht einig sind. Tritt nun dieses Verhalten häufiger bei den gleichen Personen auf, dann wird von einem „Editwar“ gesprochen, wobei alle beteiligten Parteien kontraproduktiv zueinander handeln.

Alternativ können auch zwei Benutzer kooperativ einen Artikel bearbeiten, solange sie zum Beispiel unterschiedliche Segmente hinzufügen oder auch bei den jeweils anderen kleinere Korrekturen wie Rechtschreibfehler durchführen. Auch dieses lässt sich anhand der Betrachtung der Revisionen feststellen, wenn bspw. auf eine große Änderung bei der ein Abschnitt hinzugefügt wurde eine kleine Änderung folgt, in der lediglich Stil, Grammatik und Rechtschreibung überarbeitet wurden.

Wie kann man die Reputation von Nutzern bewerten?

Um zu bestimmen welchen Status ein User innerhalb der Wikipedia-Community besitzt, können alle benutzerspezifischen Informationen herangezogen werden. Die dabei wohl wichtigsten Merkmale sind die Rechte die ein Nutzer hat, bzw. die Gruppen denen er angehört. Dazu ist zu sagen, dass die Rechte und Gruppen beantragt werden müssen, wobei dann von Benutzern die in der Gruppe Bureaucrats sind entschieden wird, ob die beantragten Rechte genehmigt werden. Damit soll gewährleistet sein, dass Rechte nicht an die falschen Benutzer vergeben werden. Weiter gilt zu beachten, dass auch ein Gast einen Artikel bearbeiten darf, was in der Vergangenheit schon zu Kontroversen geführt hat. Insgesamt lässt sich sagen, dass ein Benutzer der mehr Rechte hat auch eine höhere Reputation genießt.

Als genaues Gegenteil dazu können Nutzer auch geblockt werden. Diesen Nutzer wird die Möglichkeit genommen Beiträge zu editieren. Geblockte Nutzer fallen auf, in dem sie Artikel unsachgemäß bearbeiten oder zweifelhafte Kommentare verfassen. Wird ein solches Verhalten bei einem Wikipedia-Admin gemeldet, kann dieser einen Benutzer über einen Zeitraum hin sperren. Der Benutzer hat demnach auch eine eher niedrigere Reputation.

6. Sonstige Datenanalyse

Expertensuche: Wer kennt sich mit „Berlin“ und „Kaffee“ aus?

Mithilfe der Wiki-API ist es möglich zu ermitteln, welche Nutzer wie viel zu einem Wikipedia Artikel beigetragen haben. Um dies zu ermitteln können verschiedene Punkte mit einbezogen werden. Zum einen ist da die Anzahl der Beiträge die ein Nutzer bei einem Artikel hinzugefügt hat. Ferner lässt sich bei jedem Beitrag anzeigen, wie viel Byte der Nutzer hinzugefügt oder entfernt hat. Das lässt sich zum Beispiel als eine genauere Gewichtung hinsichtlich des geleisteten Beitrages nehmen.

Diese Kriterien werden nun auf die Artikel angewendet, bei denen festzustellen ist, ob es gemeinsame Experten gibt. Die Vorgehensweise ist dabei wie folgt. Zuerst wird der erste Artikel überprüft, um anschließend eine Liste an Nutzern zu erstellen. Diese Liste muss nun mit der Liste eines weiteren Artikels verglichen werden. Die Überschneidungen geben dabei an, welcher der Benutzer an beiden Artikeln gearbeitet hat. Nun ist es z.B. mit einer Überprüfung der jeweiligen Beiträge der einzelnen Benutzer möglich, zu untersuchen welcher Benutzer den größten Anteil an den jeweiligen Artikeln geleistet

hat. Diese Verfahrensweise lässt sich auf einzelne Benutzer anwenden. Dadurch werden die Benutzer herausgefiltert, die auf zwei Themengebieten Expertenwissen vorweisen.

Zeitgeist: Wie ändert sich die Nutzung eines Begriffs in der Datenbank über die Zeit?

Um die Veränderung eines Begriffes über die Zeit zu ergründen, sind alle in Wikipedia verzeichneten Homonyme eines Wortes zu prüfen. Dafür werden die einzelnen in der Begriffserklärung aufgeführten Artikel herausgesucht und nach dem Datum ihres ersten Erstellens sortiert. Im weiteren Schritt sind die einzelnen Revisionen nach Ihren Datum zu sortieren und daraufhin jeweils nach Tage und Monate zu aggregieren. Die so gewonnenen Daten zeigen nun, wann welcher Artikel wie oft editiert wurde. Nun wird davon ausgegangen, dass je populärer eine Bedeutung ist, er häufiger editiert wird. Bildet man diese Daten nun über einen Zeitstrahl ab, werden die einzelnen Hochphasen der jeweiligen Bedeutung sichtbar.