

# Indexieren von Emotionen bei Videos

Kathrin Knautz, Evelyn Dröge, Susanne Finkelmeyer, Daniel Guschanski, Kerstin Juchem, Cynthia Krzmyk, Daniel Miskovic, Jan Schiefer, Evrim Sen, Julia Verbina, Nils Werner und Wolfgang G. Stock, Düsseldorf

*Gegenstand der empirischen Forschungsarbeit sind dargestellte wie empfundene Gefühle bei Videos. Sind Nutzer in der Lage, solche Gefühle derart konsistent zu erschließen, dass man deren Angaben für ein emotionales Videoretrieval gebrauchen kann? Wir arbeiten mit einem kontrollierten Vokabular für neun Emotionen (Liebe, Freude, Spaß, Überraschung, Sehnsucht, Trauer, Ärger, Ekel und Angst), einem Schieberegler zur Einstellung der jeweiligen Intensität des Gefühls und mit dem Ansatz der broad Folksonomy, lassen also unterschiedliche Nutzer die Videos taggen. Versuchspersonen bekamen insgesamt 20 Videos (bearbeitete Filme aus YouTube) vorgelegt, deren Emotionen sie indexieren sollten. Wir erhielten Angaben von 776 Probanden und entsprechend 279.360 Schiebereglerinstellungen. Die Konsistenz der Nutzervoten ist sehr hoch; die Tags führen zu stabilen Verteilungen der Emotionen für die einzelnen Videos. Die endgültige Form der Verteilungen wird schon bei relativ wenigen Nutzern (unter 100) erreicht. Es ist möglich, im Sinne der Power Tags die jeweils für ein Dokument zentralen Gefühle (soweit überhaupt vorhanden) zu separieren und für das emotionale Information Retrieval (EmIR) aufzubereiten.*

## Indexing of Emotions in Videos

*Object of our empirical research study are depicted and provoked emotions in videos. Are users able to index such emotions consistently? Are the users' votes usable for emotional video retrieval? We worked with a controlled vocabulary for nine basic emotions (love, happiness, enjoyment, surprise, desire, sadness, anger, disgust, and fear), a scroll bar for adjusting the emotions' intensities, and the approach of broad folksonomies. Different users tagged the same videos. The test persons had the task to index the emotions of 20 videos (reprocessed movies from YouTube). We got data from 776 participants and could analyze all in all 279,360 scroll bar values. The consistency of the users' votes is very high; there are stable tag distributions for the emotions of the particular videos. The final shape of the distributions will be reached by the tagging activities of only few users (less than 100). Applying the approach of power tags it is possible to separate the pivotal emotions of every document – if there is any feeling at all. Those document-specific emotions establish the basis of an emotional information retrieval (EmIR) system.*

## Einleitung

„Ich habe Angst. Ich habe Angst, Dave (...)“ entgegnet HAL in Stanley Kubricks 1968er Meisterwerk *2001: Odyssee im Weltraum* dem Astronauten Dave Bowman. HAL ist ein mit einem künstlichen Wesen versehener Supercomputer der Serie 9000, der das Raumschiff „Discovery“ autark lenkt. HAL ist zu einer Gefahr für die Besatzung geworden, und Dave Bowman hat sich Zugang zu dem zentralen Computerraum verschafft und beginnt damit, den Computer schrittweise abzuschalten. Computer(systeme), die die Emotionen ihres menschlichen

Gegenübers nicht nur wahrnehmen, sondern diese ferner verarbeiten sowie interpretieren – Rosalind W. Picard bündelt dies unter dem Begriff „Affective Computing“: „(...) computing that relates to, arises from, or influences emotions“ (Picard, 1995, 1) – und darüber hinaus auch auf einer emotionalen Ebene interagieren, sind längst keine bloße Illusion mehr: „Computers are beginning to acquire the ability to express and recognize affect, and many soon be given the ability to ‚have emotions‘“ (Picard, 1995, 1).

Wir beziehen den Ansatz des Affective Computing auf den Bereich Human-Computer-Interaction (HCI) (Boehner et al., 2007) bzw. auf die Forschungen zum Information Retrieval (Stock, 2007a). Bei unserer Forschungsarbeit geht es nicht um emotional reagierende Computer, sondern darum, ob wir in Retrievalsystemen nach Gefühlen recherchieren können, die in Dokumenten entweder ausgedrückt werden oder die Gefühle beim Betrachter hervorrufen. Wie können wir solche Gefühlsreaktionen überhaupt suchbar speichern und abfragbar machen? Welche Indexierungsverfahren sind einsetzbar? Sind Gefühleinschätzungen derart konsistent, dass sie auch eindeutig Dokumenten zugeordnet werden können? Dies sind Fragen des jungen informationswissenschaftlichen Forschungsgebietes des Emotionalen Information Retrieval (EmIR) (Schmidt und Stock, 2009). Emotionen schwingen in diversen Dokumenttypen mit. „Gefühlsbeladene“ Dokumente finden wir bei

- Texten (z.B. Romanen oder Gedichten),
- Musik,
- Bildern,
- Videos,
- Multimedia-Dokumenten (z.B. Weblogs).

Wir berichten in diesem Artikel über das Indexieren von Basisemotionen bei Videos, nachdem für Musik (Lee und Neal, 2007) sowie für Bilder (Schmidt und Stock, 2009) bereits Ergebnisse vorliegen. Wir versuchen, folgende Forschungsfragen (Ff) einer ersten Beantwortung nahezubringen:

- (Ff 1): Ist zufriedenstellendes Retrieval für Filme im Web derzeit praktisch einsetzbar?
- (Ff 2): Was sind Basisemotionen?
- (Ff 3): Wie können wir Basisemotionen indexieren, um sie im emotionalen Retrieval praktisch einzusetzen?
- (Ff 4): Wie konsistent sind nutzerorientierte Verfahren beim Taggen von Emotionen in Videos?
- (Ff 5): Wie viele unterschiedliche Nutzer braucht man, bis sich eine stabile Verteilung der Gefühls-Tags ergibt?

Gibt es überhaupt stabile Verteilungen bei Gefühls-Tags?

- (Ff 6): Können wir aus den Tagverteilungen Power Tags separieren, die den dargestellten bzw. empfundenen Gefühlen entsprechen?
- (Ff 7): Korrelieren gewisse Emotionen miteinander?

Als Methode setzen wir eine Befragung von Versuchspersonen ein. Insgesamt 776 Probanden haben im Frühsommer 2009 jeweils 20 Filme (bearbeitete YouTube-Dokumente) mittels Schieberegler hinsichtlich neun verschiedener Basisemotionen getaggt.

## Videoretrieval (Forschungsfrage 1)

Wie lassen sich Videos überhaupt für Retrievalsysteme zugänglich machen? Die Retrievalforschung unterscheidet grob zwei Ansätze:

- Concept-based Information Retrieval (über Begriffe bzw. Wissensordnungen),
- Content-based Information Retrieval (nur über Inhalte des Dokuments).

Ein zeit- und personalunabhängiger Ansatz ist die automatische, contentbasierte Erschließung von Videos (Aigrain, Zhang, und Petkovic, 1996; Colombo, Del Bimbo und Pala, 1999; Del Bimbo, 1999; Enser, 2008a; Enser, 2008b; Gupta und Jain, 1997). Um contentbasiertes Retrieval (Stock, 2007a, Kap. 31) zu ermöglichen, muss zunächst eine Analyse erfolgen, um das Video für weitere Schritte vorzubereiten. Ein Film besteht aus vielen verschiedenen Szenen, die sich in Kameraeinstellungen, sog. Shots, einteilen lassen. Die Shots wiederum bestehen aus einzelnen Frames (Bildern), welche die kleinsten Videoeinheiten darstellen (Smeaton, 2007).

Dies sind visuelle Komponenten, welche unter den Low-Level Features zusammengefasst werden. Allgemein bezeichnen Low-Level Features Metadaten (Smeaton et al., 2008; Auld und Petrelli, 2008) und elementare Videoattribute, die sich aus visuellen, auditiven und textuellen Komponenten zusammensetzen (Smeaton, 2007). Folgende Low-Level Features kommen für eine automatische Indexierung infrage:

- Dimension (Einzel-)Bild
  - Farbe
  - Textur
  - Gestalt
- Dimension Ton
  - gesprochener Text
  - Musik (Tonhöhe, Rhythmus, Harmonie, Klangfarbe)
  - Geräusche
- Dimension Video
  - Segmentierung: Szenen und Shots

- Bewegung: Kamerabewegung (Schwenken oder Zoomen) und aufgenommene Bewegung.

Die Informationen, die sich aus diesen Low-Level Features erschließen lassen, sind jedoch begrenzt und deshalb nur bedingt für eine Analyse des bedeutungsträchtigen und emotionalen Contents geeignet. Doch gibt es bereits Ansätze, welche versuchen, gefühlsbeschreibenden Content aus den Low-Level Features automatisch zu extrahieren (Chan und Jones, 2005; Chen et al., 2007; Salway und Graham, 2003; Soleymani et al., 2008; Xu et al., 2005). Colombo, Del Bimbo und Pala (1999, 41) geben am Beispiel der Farbdimension zu bedenken: „The mapping of low-level color primitives into emotions is quite complex.“

Wir wollen an einem Beispiel verdeutlichen, wie contentbasiertes emotionales Videoretrieval ablaufen könnte. Hanjalic und Xu (2004) versuchen, von Low-Level Features auf den emotionalen Gehalt von Videos zu schließen. Sie verwenden ein zweidimensionales Modell, um Emotionen zu erfassen. Eine erste Dimension ist die Intensität, die andere die Valenz. Die Intensität gibt den Grad der Emotionen an, und deren Skala bewegt sich von energiegeladen, aufgeregt und wachsam bis hin zu ruhig, schläfrig und friedlich. Die Dimension der Valenz oder Wertigkeit zeigt auf, ob die dargestellten Emotionen erfreulich bzw. positiv oder unerfreulich bzw. negativ sind. Auf der Basis dieser beiden Dimensionen soll nun der emotionale Content von den Low-Level Features abgeleitet werden. Zum Erschließen der Wertigkeit der Emotionen werden drei Komponenten berücksichtigt. Diese sind zum einen die Komponente der Bewegung, welche anhand von aufeinanderfolgenden Frames gemessen wird, zum anderen die Komponente des Rhythmus, welche sich aus der Frequenz von einzelnen Schnitten erschließen lässt, sowie die Komponente der Energie des Soundspektrums. Die Dimension der Valenz wiederum wird nur anhand von einer Komponente gemessen. Hier wird lediglich die durchschnittliche Tonhöhe berücksichtigt, wobei eine hohe durchschnittliche Tonhöhe Freude und eine niedrige durchschnittliche Tonhöhe Traurigkeit ausdrücken soll. Das vorgestellte Modell misst diese beiden Dimensionen letztendlich anhand von automatisch extrahierten Zahlenwerten, weshalb eine differenzierte emotionale Indexierung der Videos somit nicht gegeben sein kann. Hierdurch kann lediglich eine Basis geschaffen werden, um besonders positive oder negative emotionale Sequenzen zu bestimmen. Die Autoren geben zu, dass die bis dato erforschten Relationen zwischen den Low-Level Features und der emotionalen Ebene vage sind, weshalb sich eine Abbildung des emotionalen

Gehalts eines Videos auf Basis der Low-Level Features als schwierig gestaltet. Contentbasiertes Videoretrieval steckt noch immer in der Entwicklung, obwohl schon einige Fortschritte in Hinblick auf die automatische Extraktion der Low-Level Features zu bemerken sind (Chan und Jones, 2005; Xu, Chia, und Jin, 2005). Jedoch müssen nach Smeaton (2007) weitere Aufgabenschwerpunkte gesetzt werden. Das Ermitteln einzelner Features muss verbessert werden, denn noch sind die Ergebnisse, besonders im automatischen Herausfiltern von semantischen Merkmalen, viel zu ungenau. Bisherige Ansätze im Videoretrieval basieren zudem hauptsächlich auf Keyframe- und Bildretrieval (Auld und Petrelli, 2008; Kim und Kim, 2009), jedoch entspricht dieser statische Ansatz nicht dem gewünschten Videoretrieval, da hier der temporale Aspekt, die Bewegung der Objekte und der Kamera fehlt. Auf Grund dessen eignen sich Keyframes nur bedingt als Kurzfassung von Videos, denn die Auswahlverfahren für diese setzen meist einfach in der Mitte eines Shots an und nicht dort, wo sich der Inhalt des Videos am besten repräsentieren lässt (Kim und Kim, 2009). Des Weiteren bemängelt Smeaton, dass momentan nur einzelne Video-Shots ausschlaggebend sind, was den User bei seiner Videosuche nur bedingt zufriedenstellen kann. Denn anhand eines Shots können kaum semantische Inhalte, wie die Stimmung und Botschaft eines Videos, übermittelt werden. Deshalb sollte stattdessen eine Szene aus bedeutungsträchtigen Shots zusammengestellt werden, die den „wahren“ Content widerspiegeln, um auf die Bedürfnisse der Nutzer einzugehen. Smeaton (2007) hebt einen personalisierten Videozusammenschnitt (Smeaton et al., 2008), vergleichbar mit Videotrailern, hervor. Dieser könnte dem User helfen, seine Suchergebnisse besser zu bewerten. Doch momentan beschränkt sich Videoretrieval nur auf einfaches Browsen und einer Keyframe-Darstellung der Ergebnisse. Das contentbasierte Videoretrieval im Allgemeinen und das Retrieval nach Emotionen im Besonderen bringen also noch einige Herausforderungen mit sich. Wichtig wäre zu bestimmen, welcher konkrete Content (sagen wir: ein lächelndes Gesicht; Bailenson et al., 2008) bestimmte Emotionen beim Betrachter (hier: Freude) auslöst. Forschungen zu solchen „emotionalen Referenzpunkten“ (im Anschluss an die „cognitive reference points“ von Eleanor Rosch (1975)) stehen noch aus. Wenden wir uns nunmehr dem conceptbasierten Videoretrieval zu! Snoek und Worring (2009) haben bezüglich des Retrievals von Videos über 300 Literaturquellen ausgewertet und festgestellt, dass in diesen zumeist conceptbasierte Lösungen als vielversprechende Alter-

nativen gesehen werden. Im conceptbasierten Videoretrieval arbeiten wir mit Begriffen, die zwar prinzipiell auch dem Content (vor allem dem gesprochenen Text) automatisch entnommen werden könnten, aber derzeit vorwiegend intellektuell zugeordnet werden (Schweins, 1997). Ein erster Weg führt über Wissensordnungen (beispielsweise Thesauri) und professioneller Inhaltserschließung. Diesen Weg wählen beispielsweise die deutschen und österreichischen Fernsehanstalten bei der Inhaltserschließung von Fernsehproduktionen. Es kommen – bei intellektueller Indexierung – sowohl Klassifikationen als auch ein Thesaurus zum Einsatz (ARD/ORF/ZDF, 2008). Gefühle werden jedoch nicht beschrieben. Beim Vorliegen einer Begriffsordnung für Emotionen wäre es zumindest prinzipiell möglich, den Videos kontrolliertes Vokabular zu Gefühlen intellektuell zuzuordnen (Kim und Park, 2003). Diese Methode ist jedoch stark abhängig von der Person des jeweiligen Indexers. Ein zentrales Problem allen Concept-based Videoretrievals dürfte die Indexierungskonsistenz sein, die bei Bildern – und wohl auch bei Filmen – sehr schwach ausgeprägt ist (Markey, 1984). „The output of the indexing process seemed to be quite inconsistent“ (Markkula und Sormunen, 2000, 273). Wir lernen von Goodrum (2000, 64), „manual indexing suffers from low term agreement across indexers (...), and between indexers and user queries.“ „There is evidence that current systems for image access often fail the user“, erinnert Jörgensen (1998, 162). Es gibt bisher keinerlei Erfahrungen mit der intellektuellen Indexierung von Emotionen. „The influence of the photograph's emotional tone on categorization has not been discussed much in previous studies“, berichten Laine-Hernandez and Westman (2006). Letztlich steht solch einem Unterfangen im World Wide Web ein praktisches Problem entgegen. Es ist schlicht unmöglich (weil unbezahlbar), die Millionen von verfügbaren Videos durch professionelle Indexer auswerten zu lassen, denn „manual assignment of textual attributes is both time consuming and costly“ (Goodrum, 2000, 64). Etwas platt ausgedrückt: So geht es nicht (Enser, Sandom, Hare, und Lewis, 2007). Der geringere Zeit- und Personalaufwand würde für usergenerierte Tags – also Social Tagging oder kooperatives Indexieren (Jörgensen, 2007) – sprechen, wie sie beispielsweise auf Video-Plattformen wie YouTube zu finden sind (Geisler und Burns, 2007; Melenhorst et al., 2008; van Velsen und Melenhorst, 2009). Folksonomies (Peters, 2009; Smith, 2008; Weller, Peters und Stock, 2010) sind entweder „narrow“ (nur der Urheber des Dokuments darf Tags vergeben; Beispiel: YouTube), „extended narrow“ (außer dem Urheber dürfen auch seine Freunde Tags

verteilen, jeden Tag aber nur einmal; Beispiel: Flickr) oder „broad“ (jeder Nutzer des Informationsdienstes darf Tags vergeben, die demnach auch mehrfach einem Dokument zugeordnet werden können; Beispiel: Del.icio.us). Bei narrow und extended narrow Folksonomies stellt sich das Problem der Indexierungskonsistenz – hier sogar besonders pointiert, da keine Information Professionals, sondern Laien indexieren (Peters und Stock, 2007; Peters und Stock, 2008). Bei broad Folksonomies und einer ausreichend großen Zahl an taggenden Nutzern sollten sich bei den hauptsächlich vergebenen Tags keinerlei Konsistenzprobleme ergeben, allerdings entstammen die Tags keinem normierten Vokabular, sondern der Sprache der Nutzer. Da YouTube eine narrow Folksonomy nutzt, können wir die dort durchgeführte Indexierung nicht für unsere Zwecke gebrauchen. So endet die Forschungsfrage 1 mit einem negativen Ergebnis: Weder content- noch conceptbasiertes Videoretrieval führt derzeit zu praktisch brauchbaren Ergebnissen für Filme im WWW.

### Basisemotionen (Forschungsfrage 2)

Zur Klärung dessen, was „Emotionen“ sind, greifen wir auf Ergebnisse und Diskussionen aus der Psychologie zurück. Die Emotionsforschung kann auf eine lange Tradition zurückschauen, und dennoch war es bisher nicht möglich, eine einheitliche exakte Definition für das Konzept der Emotion vorzulegen. Eine Definition bereitet insofern Schwierigkeiten, da zwar jeder weiß, was eine Emotion ist, die Formulierung einer konkreten Begriffsbestimmung jedoch schwierig ist (Fehr und Russell, 1984). Auch Schmidt-Atzert stellt fest, dass bisher „kein Konsens festzustellen [ist], was man unter einer Emotion zu verstehen hat“ (Schmidt-Atzert, 1996, 18). Nach Izard (1994, 20) muss eine vollständige Definition von Emotion drei Aspekte beinhalten:

- das Erleben oder das bewusste Empfinden des Gefühls,
- die Prozesse, die sich im Gehirn und im Nervensystem abspielen und
- das beobachtbare Ausdrucksgebaren, besonders im Gesicht.

Bereits 1981 untersuchten Kleinginna & Kleinginna 100 Aussagen und Definitionen aus einschlägigen Spezialwerken, Wörterbüchern und Einführungstexten und versuchten, eine Arbeitsdefinition zu erstellen: Emotion ist „ein komplexes Muster von Veränderungen, das physiologische Erregung, Gefühle, kognitive Prozesse und Verhaltensweisen einschließt, die in Reaktion auf eine Situation auftreten, welche ein Individuum als persönlich bedeutsam wahrgenommen

hat“ (Kleinginna und Kleinginna, 1981). Auch Meyer, Reisenzein und Schützwohl (2001) betreiben ihre Forschung auf Basis einer Arbeitsdefinition, da die Definition selbst eine zentrale Fragestellung innerhalb der Emotionsforschung abbildet und damit vielmehr „nicht die Voraussetzung der wissenschaftlichen Emotionsforschung ist, sondern deren Resultat“ (Meyer, Reisenzein und Schützwohl, 2001, 50). Dennoch lassen sich Emotionen nach Meyer, Reisenzein und Schützwohl (2001, 24) durch verschiedene Merkmale etwas genauer charakterisieren. So sind Emotionen aktuelle psychische Zustände von Menschen und damit abgrenzbar von anderen Konzepten wie Dispositionen und Stimmungen. Emotionen sind zudem auf ein bestimmtes, nicht unbedingt real existierendes Objekt bzw. auf Objekte gerichtet, welche diese Emotionen auslösen. Als weiteres Merkmal nennen die Autoren, dass sich Emotionen in der so genannten Reaktionsstria aus subjektiven, behavioralen und physiologischen Aspekt zeigen. Auch haben Emotionen eine bestimmte Qualität, Intensität und Dauer. Die Qualität fungiert als Gruppierungsmerkmal, so dass man bei dem Vorkommen einer Emotion in einer Situation wie beispielsweise in „Person A freut sich“ von einer Zuordnung des Qualitätstyps *Freude* spricht. Die konkreten Realisierungen der Qualitätstypen zeichnen sich, neben einer zeitlichen Begrenzung, zudem durch unterschiedliche Intensitäten aus. Die Stärke einer Emotion lässt sich nach Meyer, Reisenzein und Schützwohl auf einer Skala von niedriger über mittlerer bis hin zu einer starken Intensität festlegen (2001, 29-30).

Wie Emotionen entstehen, versuchen verschiedene Theorien der Emotionsforschung mit unterschiedlichen Ausrichtungen zu klären. Behavioristische Ansätze betonen den Verhaltensaspekt und die auslösenden Bedingungen. Emotionen werden entweder als angeboren angesehen (Furcht, Wut und Liebe) oder entstehen als konditionierte Reaktionsmuster auf bestimmte Reize (Watson, 1930). Weiterhin gibt es evolutionspsychologische Ansätze, welche die evolutionär bedingten adaptiven Funktionen von Emotionen, wie beispielsweise ihre Überlebensfunktion, betonen. Sie gehen auf Darwin (1872) zurück, dessen Hauptanliegen es war, die phylogenetische Entwicklung von Emotionen nachzuweisen, und wurden von diversen Emotionsforschern (McDougall, 1926; Plutchik, 1980; Izard 1971, 1977, 1991; Ekman et al. 1982; Tomkins 1962, 1963) weitergeführt. Dass Emotionen durch die Wechselwirkung zwischen physiologischen Veränderungen bestimmt werden, wird in kognitiv-physiologische Theorien postuliert. Durch die (in-)direkte Wahrnehmung dieser Veränderung entwickelt

und adaptiert der Mensch seine Emotionen (James, 1884; Panksepp, 1982; Frijda, 1986). Attributionstheorien bzw. attributionale Theorien beschreiben, wie der Mensch mittels Ursachenzuschreibung versucht, seine Umwelt zu verstehen und zu kontrollieren. Emotionen sind nach diesen Ansätzen damit Reaktionen auf Handlungsergebnisse (Arnold, 1960; Weiner und Graham, 1984; Lazarus, 1960).

Im Zuge unserer Forschung stellt sich die Frage, welche Emotionen für das Projekt ausgewählt und damit in Hinblick auf das emotionale Videoretrieval herangezogen werden sollen. In der Psychologie findet sich hinsichtlich einer Kategorisierung von Emotionen der Ansatz, Emotionen auf eine kleine, aber relativ fixe Anzahl zu reduzieren. Diese werden als fundamentale Emotionen bzw. als Basisemotionen bezeichnet (Ortony und Turner, 1990). Über ihre Anzahl herrscht jedoch auf Grund der verschiedenen Theorieausrichtungen der Emotionsforscher Uneinigkeit, so dass die unterschiedlichen Basisemotionen in Tabelle 1 aufgezeigt werden.

Sowohl das emotionale Musikretrieval (Lee und Neal, 2007) als auch das emotionale Bildretrieval (Schmidt und Stock, 2009) arbeiten mit fünf Basisemotionen: Traurigkeit, Glück, Ärger, Angst und Ekel. Angesichts der psychologischen Fachliteratur haben wir die Menge der fundamentalen Emotionen

angereichert. Als grundlegende Emotionen werden folgende für das Projekt ausgewählt: Trauer, Ärger, Angst, Ekel, Überraschung, Sehnsucht, Freude und Liebe. Unabhängig von diesen in der Emotionspsychologie postulierten Basisemotionen wird zusätzlich in Form von Humor bzw. Witz ein weiterer Aspekt mit aufgenommen, um auch dieser medialen Komponente gerecht zu werden. Unter *Trauer* kann man „die mehr oder weniger stereotypisierten Reaktionen auf Verluste [sehen]. Der Verlust betrifft in der Regel eine geliebte Person, aber auch der Wegfall von materiellen Gütern oder der Verlust von sozialen Werten wie Ehre kann zum Anlass für eine Trauerreaktion werden“ (Hoffmann, 1983, 183). Nach Izard dient Trauer (bzw. Kummer) „äußerst nützlichen Aufgaben, indem er einem selbst und anderen mitteilt, dass nicht alles in Ordnung ist, indem er den Menschen motiviert, das Notwendige zu tun, um den Kummer zu verringern. Kummer macht einen aufgeschlossen für die eigenen Probleme und die Probleme der Welt“ (Izard, 1994, 109-110). *Ärger*-Emotionen werden in allen Kulturen gleich ausgedrückt und besitzen eine gleich Erlebnisqualität (Ekman & Friesen, 1971). „Ärger dient zur Regulierung interpersoneller Beziehungen. Er bringt die Androhung zum Ausdruck, die den zielgerichteten Handlungsablauf einer Person störende Barriere beseitigen oder zerstören zu wollen“ (Bornewas-

ser und Mummendey, 1983, 156). *Angst* und *Furcht* werden in der Emotionspsychologie teilweise unterschieden (Izard, 1994), teilweise aber auch synonym verwendet. Nach McDougall hat *Furcht* die biologische Funktion des Vermeidens von Verletzungen oder Tod (1908/1960). Nach Schwarzer ist *Angst* „die Besorgtheit und Aufgeregtheit angesichts von Situationen, die subjektiv als bedrohlich und ungewiss eingeschätzt werden“ (Schwarzer, 1983, 147). McDougall definiert *Ekel* als ein Abstoßungsinstinkt, welcher den Handlungsimpuls des Zurückweizens bzw. Zurückweichens hervorruft, um Schädigung, Krankheit oder Tod zu vermeiden (McDougall, 1908/1960). Als Ursache kann beispielsweise „[m]aterieller oder psychologischer Verfall (,alles Verdorbene‘) angesehen werden“ (Izard, 1994, 111). *Ekel* ist eine der fünf Zivilisations-schranken und „ist eine Provokation, weil wir uns gegen die Nähe wehren und das Ekelhafte gleichzeitig [...] unser Interesse weckt“ (Vendrell Ferran, 2008, 233 f.). *Überraschung* als Emotion ist ein vorübergehender Zustand und „resultiert aus einem starken Anstieg in neuraler Stimulierung, typischerweise herbeigeführt durch ein plötzliches, unerwartetes Ereignis. *Überraschung* dient dem sehr nützlichen Zweck, das Nervensystem von gerade bestehender Emotion und gerade bestehende Kognition zu befreien, so dass das Individuum angemessen auf die Reizsituation und die plötzlicher Veränderung reagieren kann, die es erlebt hat“ (Izard, 1994, 109). *Sehnsüchte* sind „intensive längerfristige und gegenwärtig schwer – oder gar nicht – erfüllbare Wünsche nach Personen, Dingen, Ereignissen oder Erfahrungen aus ihrer Vergangenheit, Gegenwart oder Zukunft“ (Wilhelm, 2008). *Freude* ist „gekennzeichnet durch ein Gefühl von Selbstvertrauen und Bedeutsamkeit, ein Gefühl geliebt zu werden und liebenswert zu sein. [...] Freude ist nicht nur gekennzeichnet durch Akzeptierung des Selbst, sondern auch von der Akzeptierung von und gute Beziehungen mit anderen Menschen und der Welt im Allgemeinen, insbesondere mit dem Gegenstand der Freude. Freude an anderen führt zu Vertrauen in andere“ (Izard, 1994, 272). Die romantische *Liebe* „versetzt den gesamten Körper und Geist in Hochstimmung und macht den Menschen offen für Erregung und Ekstase und anfällig für affektiv-kognitive Orientierungen“ (Izard, 1994, 117 f.). Damit stellt *Liebe* eine besondere Beziehung zwischen den Menschen dar. Das Konzept von *Spaß*, *Humor* und *Witz* ist eng verwandt mit dem der Freude und ist fester Bestandteil der sozialen Interaktion. *Humor* im Sinne des emotionalen Retrieval bei medialen Dokumentationseinheiten ist eine Mischung aus dem standardisierten Humor (wie z.B. in Witzen oder Komödien) und spontanem Humor.

Tabelle 1: Basisemotionen und Vertreter (verändert und erweitert nach Ortony und Turner, 1990, 316).

Fundamental Emotion	Basis for Inclusion	Reference
rage and terror, anxiety, joy	Hardwired	Gray (1982)
expectancy, fear, rage, panic	Hardwired	Panksepp (1982)
pain, pleasure	Density of neural firing	Mowrer (1960)
anger, interest, contempt, disgust, distress, fear, joy, shame, surprise	Unlearned emotional states	Tomkins (1984)
acceptance, anger, anticipation, disgust, joy, fear, sadness, surprise	Relation to adaptive biological processes	Plutchik (1980)
anger, aversion, courage, dejection, desire, despair, fear, hate, hope, love, sadness	Relation to action tendencies	Arnold (1960)
anger, disgust, anxiety, happiness, sadness	Do not require propositional content	Oatley & Johnson-Laird (1987)
anger, disgust, fear, joy, sadness, surprise	Universal facial expressions	Ekman, Friesen & Ellsworth (1982)
anger, disgust, elation, fear, subjection, tender-emotion, wonder	Relation to instincts	McDougall (1908/1960; 1926)
anger, contempt, disgust, distress, fear, guilt, interest, joy, shame, surprise	Hardwired	Izard (1971)
desire, happiness, interest, surprise, wonder, sorrow	Forms of action readiness	Frijda (1986)
fear, grief, love, rage	Bodily involvement	James (1884)
fear, love, rage	Hardwired	Watson (1930)
happiness, sadness	Attribution independent	Weiner and Graham (1984)



Abbildung 1: Schieberegler-Tagging von Emotionen.

„Humor durchbricht unsere Routine auf positive Weise, er unterrichtet uns über soziale Interaktion, er lobt und bestraft für soziales Verhalten in fitnessrelevanten Themen [im Sinne Darwins, Anm. d. Autoren] und er schafft Bindungen und Abgrenzungen“ (Schwender, 2001, 220).

### Gefühls-Tagging mithilfe von Schiebereglern (Forschungsfrage 3)

Die Beantwortung von Forschungsfrage 1 führt – in positiver Formulierung – zur Erkenntnis, dass die sinnvollste Methode, Videos im WWW zu indexieren, im Einsatz einer broad Folksonomy liegt, dass also viele unterschiedliche Personen dieselben Dokumente indexieren. Forschungsfrage 2 leitet uns dazu, neun Basisemotionen zu akzeptieren. Des Weiteren wissen wir, dass Emotionen über Intensitäten verfügen. Auf diese drei Ergebnisse aufbauend, arbeiten wir beim Indexieren von Gefühlsmomenten

- mit einem kontrollierten Vokabular für unsere neun Basisemotionen,
- mit Schiebereglern zur Einstellung der Intensität pro Emotion (Skala: 0 [keine Ausprägung] bis 10 [starke Ausprägung]),
- mit dem Ansatz der broad Folksonomy, lassen also unterschiedliche Nutzer die Dokumente durch Schieberegler-einstellungen taggen.

Wir folgen damit dem Ansatz von Lee und Neal (2007) für emotionales Musikretrieval und von Schmidt und Stock (2009) für emotionales Bilderretrieval. Einer Idee in Schmidt und Stock (2009, 865) folgend, unterscheiden wir bei den Gefühlen in

dargestellte Emotionen und in beim Betrachter ausgelöste Emotionen. Diese müssen schließlich nicht übereinstimmend sein. Stellen wir uns beispielsweise vor: Ein breit lächelnder Hooligan (Lächeln ist Ausdruck von Freude) schlägt auf andere Fußballfans ein, dann wird dies nicht bei allen Betrachtern ebenso Freude auslösen, sondern eher Ärger oder Ekel. Die Nutzer haben ebenso stets die Möglichkeit festzustellen, dass im Video gar keine Emotion ausgedrückt wird bzw. dass sie selbst beim Betrachten nichts empfinden. Zusätzlich baten wir die Nutzer (in einer Art Kommentarfeld) anzugeben, welche Gründe für die Emotionen bzw. deren Intensitäten für sie in Betracht kommen. Abbildung 1 zeigt das von uns eingesetzte Schieberegler-Tagging von Basisemotionen.

### Methode: Online-Befragung zu Bearbeitungen von YouTube-Videos

Die methodische Bearbeitung des Forschungsprojektes durchläuft folgende fünf Schritte:

- (1) Auswahl und Bearbeitung von Videos als Versuchsdokumente
- (2) Pretest des Versuchs
- (3) Auswahl und Anschreiben der Probanden
- (4) Speichern der Antworten der Versuchspersonen
- (5) Auswerten der Rohdaten

(1) Insgesamt wurden aus YouTube 20 Videos ausgewählt: je zwei, von denen die Versuchsleiter meinten, dass sie einer der neun Basisemotionen entsprächen und zwei gefühlsneutrale Filme. Pro Fun-

damentalgefühl hat ein Film eine musikalische Untermalung und der andere nicht. Hiermit wollten wir uns den Weg offen halten zu analysieren, ob die Musik Einfluss auf die zugeschriebenen bzw. beobachteten Gefühle hat. Diese Daten harren noch der Auswertung. (Ein weiterer Projektschritt, über den wir hier allerdings auch nichts berichten, hat 20 weitere Filme – diesmal jedoch in schwarz/weiß – als Forschungsgegenstand. Hieraus erhoffen wir uns Ergebnisse über das Gefühlstaging bei Farb- bzw. Schwarz-Weiß-Filmen).

Mit einer derart großen Anzahl an Videos, die die Versuchspersonen ansehen und taggen sollen, spielt die Videolänge bei der Selektion und Bearbeitung eine zentrale Rolle. Es galt eine Videodauer zu finden, die einerseits ausreicht, um die gezeigten Emotionen zu transportieren und andererseits so kurz wie möglich ausfällt, um die zumutbare Aufmerksamkeitsspanne der Probanden nicht zu überschreiten. Bei der Einschätzung der zumutbaren Aufmerksamkeitsdauer stützten wir uns auf die takeGas-Studie von DoubleClick & Tomorrow Focus AG (Rhode und Lustig, 2007) und die Studie von Tubemogul (Burch, 2008). Beide Studien untersuchten die durchschnittliche Betrachtungsdauer von Online-Videos. Die Ergebnisse der Studien weichen recht deutlich voneinander ab. Während die takeGas-Studie eine Ausstiegsrate von 45 Prozent ab einer Videolänge von 20 Sekunden verzeichnete, zeigte sich in der Tubemogul-Studie ein vergleichbarer Wert erst ab einer Länge von 60 Sekunden. Da sich beide Studien sowohl in Kontext als auch bezüglich der gezeigten Videoinhalte unterscheiden, können wir die Werte lediglich als Orientierungspunkt verwenden. Die Entscheidung fiel auf eine maximale Länge von 25 Sekunden pro Video. Beim Videoschnitt kam das Videobearbeitungsprogramm Jahshaka zum Einsatz, mit dessen Hilfe die Videos auf die gewünschte Länge zugeschnitten und in ein geeignetes Videoformat konvertiert werden konnten.

(2) In einem dem Versuch vorgeschalteten Pretest konnten die Probanden, 14 Studierende der Informationswissenschaft, auf Unstimmigkeiten sowohl formaler als auch inhaltlicher Natur hinweisen. Nach wenigen Korrekturen ist der Versuchsablauf optimiert worden.

(3) Die Ermittlung der Teilnehmer an der Studie erfolgte mit Hilfe des Email-Verteilers des Rechenzentrums der Heinrich-Heine-Universität Düsseldorf. Die dort aufgeführten Studenten erhielten eine Email, in der auf das Forschungsprojekt hingewiesen und um Teilnahme an der Umfrage gebeten wurde. Aus urheberrechtlichen Gründen wurde die Umfrage durch ein in der Rundmail mitgeschicktes Passwort vor Zugriffen durch universitätsfremde Nutzer geschützt.

(4) Um die Speicherung der Umfragedaten und damit die Auswertung zu gewährleisten, stellte die Heinrich-Heine-Universität Speicherplatz auf einem Webserver mit PHP-Unterstützung und eine MySQL Datenbank zur Verfügung. Daraufhin wurde eine PHP-Anwendung entwickelt, welche die HTML-Seiten der Umfrage dynamisch erzeugte und die Umfragedaten der Nutzer in der Datenbank speicherte. Innerhalb der PHP-Applikation waren die grafischen und funktionalen Eigenschaften der Umfrage hinterlegt. Die einzelnen Umfrageseiten basierten auf einer generischen Vorlage, so wurde jeweils nur die Information des abzuspielenden Videos verändert. Die Anzahl der Videos konnte somit beliebig verändert werden, ohne in den eigentlichen Quellcode der Anwendung eingreifen zu müssen. Jedem Studenten wurde zu Anfang der Umfrage eine eindeutige SessionID zugewiesen, welche über ein Cookie im Browser des Teilnehmers gespeichert wird. Diese SessionID ermöglichte es der PHP-Anwendung, den Nutzer auf jeder Seite der Umfrage zu identifizieren. Eine Löschung des Cookies erfolgte nach der Schließung des Browsers.

Die Hauptbestandteile der Datenbank bildeten drei Tabellen. In einer ersten Tabelle, *emo\_user* (vgl. Abbildung 2), wurden die Angaben der Teilnehmer bezüglich ihrer Fakultät und Geschlechts verzeichnet. Auch eine Speicherung der jeweiligen SessionID und des Anfangszeitpunktes erfolgte in diesem Teil der Datenbank. Die eingegebenen Video-Bewertungen mittels Schieberegler wurden in den Tabellen *emo\_data* und *emo\_data\_complete* abgelegt. Erfolgte eine vollständige Bewertung aller 20 Videos, zog dies eine Speicherung in der Tabelle *emo\_data\_complete* nach sich. In der Tabelle *emo\_data* hingegen wurden auch die bereits abgegebenen Bewertungen bei abgebrochenen Durchläufen gespeichert. Dadurch wären wir zusätzlich in der Lage gewesen, diese Daten bei der Auswertung zu berücksichtigen. Um den Zugriff auf die vollständigen Datensätze (vgl. Abbildung 3) und damit die Auswertung zu gewährleisten, erfolgte eine Übertragung der Daten auf eine unabhängige, Passwort-gesicherte Seite mit Exportfunktion.

776 Probanden haben alle 20 Videos getagt. Hiervon sind 36 Prozent männlich und 64 Prozent weiblich. Wir hatten also insgesamt 279.360 Schieberegler-einstellungen (776 Personen \* 20 Filme \* 18 Emotionen) als Basis für weitere Berechnungen zur Verfügung. Wir verzichteten auf die Berechnung von Signifikanzwerten, da bei solch großen Zahlen (nahezu) alle Resultate zu statistisch signifikanten Ergebnissen führen. Für unsere Zwecke reichen Werte der deskriptiven Statistik vollständig aus, da wir ja ausschließlich



Abbildung 2: Attribute der Datenbank *emo\_user*.



Abbildung 3: Datensatzbeispiel.

Tabelle 2: Konsistenz der emotionalen Voten bei den dargestellten Emotionen.

dargestellte Basisemotion	Standardabweichung	Frauen	Männer
Liebe	0,88	0,94	0,82
Ekel	0,92	0,92	0,93
Trauer	1,11	1,09	1,12
Ärger	1,21	1,11	1,30
Spaß	1,46	1,48	1,44
Glück	1,54	1,56	1,52
Sehnsucht	1,55	1,69	1,42
Angst	1,61	1,69	1,54
Überraschung	1,94	2,15	1,74

Anmerkung: N = 776 Versuchspersonen; 279.360 Schieberegler-einstellungen

Tabelle 3: Konsistenz der emotionalen Voten bei den empfundenen Emotionen.

empfundene Basisemotion	Standardabweichung	Frauen	Männer
Liebe	0,66	0,72	0,63
Trauer	0,91	0,90	0,93
Ärger	1,15	1,16	1,14
Sehnsucht	1,15	1,23	1,07
Ekel	1,24	1,28	1,19
Angst	1,36	1,37	1,34
Glück	1,56	1,66	1,47
Überraschung	1,71	1,77	1,65
Spaß	1,75	2,01	1,63

Anmerkung: N = 776 Versuchspersonen; 279.360 Schieberegler-einstellungen

eruiieren möchten, ob konsistentes Indexieren von Emotionen überhaupt möglich ist. Wir sind uns dessen bewusst, dass – mit den Studierenden einer einzigen Universität – weder eine Zufallsstichprobe noch eine repräsentative zielgruppenspezifische Stichprobe vorliegt. Durch die große Zahl an Werten hoffen wir jedoch, zumindest die Tendenzen emotionalen Indexierens von Videos aufspüren zu können.

(5) Zur Auswertung der Rohdaten konnte sowohl auf Excel als auch auf SPSS zurückgegriffen werden, da beide Anwendungen die Analyse der gewonnenen Daten in Bezug auf die formulierten Forschungsfragen ermöglichten.

### Konsistenz der Gefühlsindexierung (Forschungsfrage 4)

Die Standardabweichung ist eine der wichtigsten Streuungsparameter einer diskreten Zufallsvariable. Sie ist in Analogie zu den entsprechenden empirischen Maßzahlen definiert und setzt metrisch skalierte Zufallsvariablen voraus. Wir haben für alle (dargestellten wie empfundenen) Basisemotionen pro Video die Standardabweichung der Voten berechnet und danach pro Gefühl das arithmetische Mittel der Standardabweichungen bestimmt. Grundsätzlich gilt: Je geringer der Wert, desto konsistenter haben die Probanden votiert.

Insgesamt zeigen sich sowohl bei den dargestellten Emotionen (Tabelle 2) als auch bei den empfundenen Gefühlen (Tabelle 3) recht kleine Werte der Standardabweichung und damit hohe Werte der Konsistenz der Einschätzungen der Versuchspersonen. Insbesondere beim Gefühl der Liebe gibt es (bei einer mittleren Standardabweichung von 0,88 bei der dargestellten Emotion und von 0,66 bei der Empfindung) eine sehr hohe Konsistenz, aber auch Ekel, Trauer und Ärger sind offenbar stark konsensfähig. Die Überraschung sorgt für eine Überraschung, denn diese Emotion ist (mit einem Wert von 1,94 beim dargestellten Gefühl) die umstrittenste im gesamten Bewertungsprozess. Auch beim empfundenen Spaß gibt es große Auffassungsunterschiede (die mittlere Standardabweichung beträgt 1,75).

Die kleinen Zahlen der Standardabweichung bedürfen einer Erläuterung. Häufig dominieren bei den Videos nur wenige Emotionen, meist gar nur ein einziges Gefühl. Dies liegt in der Auswahl der Filme begründet, die ja bewusst nach emotionalen Kriterien selektiert worden sind. Damit bekommen jeweils wenige Basisemotionen einen hohen Intensitätswert und die anderen einen Wert nahe Null. Da die Versuchspersonen konsistent mit etwa Null votiert haben, liegt nicht nur das arithmetische Mittel, sondern auch die Standardabweichung bei einem Wert nahe Null.

Ein Nebenergebnis soll hier nicht verschwiegen werden. Eine Trennung der Probanden nach Geschlecht weist interessante Resultate auf: Bei der dargestellten Emotion waren die Frauen fast durchweg uneiniger als die Männer – lediglich bei den Emotionen Trauer, Ärger und Ekel sind sie in ihren Voten konsistenter. Große Unterschiede bemerken wir bei Sehnsucht und Überraschung, hier sind die Männer weitaus konsistenter. Differenzen liegen auch beim Ärger vor, den die weiblichen Versuchspersonen konsistenter den Videos zuschreiben. Auch bei der empfundenen Emotion gibt es geschlechtsspezifische Unterschiede: insbesondere bei Sehnsucht, Glück und Spaß ergibt sich bei den Männern ein weitaus konsistenteres Gefühlserlebnis. Bei der Untersuchung des Gefühls-Tagging bei Bildern von Schmidt und Stock (2009, 871) ergab sich ein ähnliches Bild: Dort indexierten Frauen bei allen Emotionen weniger konsistent als Männer. Eine Erklärung für diese Beobachtung steht noch aus.

Die Forschungsfrage 4 können wir positiv beantworten: In der Tat beurteilen unsere Versuchspersonen Emotionen in Videos konsistent. Damit ist eine zentral wichtige Hürde in Hinblick auf emotionales Retrieval überwunden. Wir können davon ausgehen, dass Nutzer (zumindest bestimmte abgrenzbare

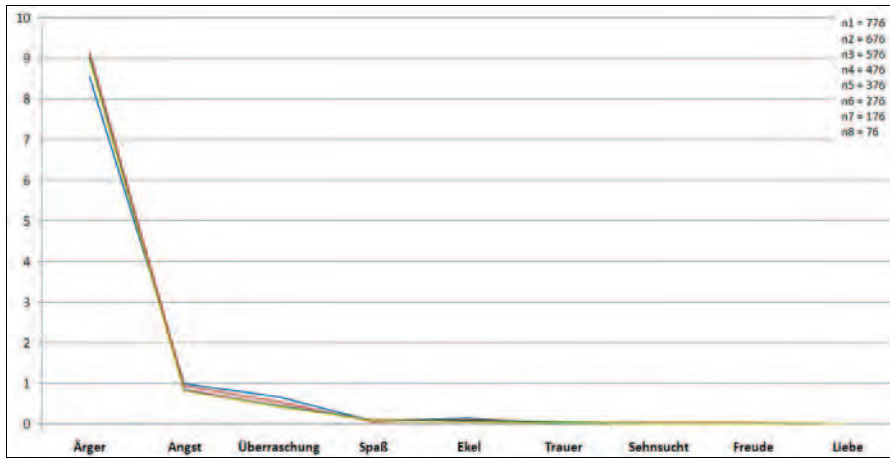


Abbildung 4: Verteilungen der Gefühlsnoten (dargestellte Emotionen) für Video Nr. 20 bei wachsenden Indexierernzahlen.

Tabelle 4: Durchschnittliche SchiebereglerEinstellung für Video Nr. 20 (dargestellte Emotionen) bei wachsenden Indexerzahlen.

Block	Teilnehmerzahl (n)	durchschnittliche SchiebereglerEinstellung	Basisemotion
1	n = 76	6,3	Ärger / dargestellte Emotionen
2	n = 176	6,6	Ärger / dargestellte Emotionen
3	n = 276	6,8	Ärger / dargestellte Emotionen
4	n = 376	6,7	Ärger / dargestellte Emotionen
5	n = 476	6,7	Ärger / dargestellte Emotionen
6	n = 576	6,7	Ärger / dargestellte Emotionen
7	n = 676	6,6	Ärger / dargestellte Emotionen
8	n = 776	6,6	Ärger / dargestellte Emotionen
1	n = 76	0,7	Angst / dargestellte Emotionen
2	n = 176	1,1	Angst / dargestellte Emotionen
3	n = 276	1,0	Angst / dargestellte Emotionen
4	n = 376	0,9	Angst / dargestellte Emotionen
			etc.

Nutzergruppen) den Dokumenten mehr oder minder dieselben Basisemotionen mit zumindest ähnlicher Intensität zuordnen. Ob dieser Sachverhalt global gilt, ist offen. Es ist beispielsweise möglich, dass Nutzer unterschiedlicher Kulturkreise zwar wiederum untereinander konsistent voten, aber – kulturübergreifend – andere Gefühle nennen.

### Stabile Verteilung der Gefühls-Tags (Forschungsfrage 5)

Halpin, Robu und Shepherd (2007) haben in ihrer Untersuchung die Dynamik der Tag-Verteilungen in Folksonomies gemessen. Hierbei wurde versucht festzustellen, ab welcher Zeit sich eine stabile Verteilung von Tags in Folksonomies etabliert. Als Maßeinheit wurden jeweils Monate gewählt. Bezogen auf das kollaborative Tagging ist diese Art der Untersuchung durchaus sinnvoll, da in Folksonomies Nutzer in den Social-Media-Plattformen kontinuierlich Dokumente

taggen. Bei unserem Forschungsprojekt haben die Probanden dagegen nicht zeitverteilt über Monate hinweg, sondern innerhalb einer kurzen Zeitspanne alle zur Verfügung gestellten Videos angesehen und indexiert. Es gilt herauszufinden, ab welcher Anzahl von Nutzern sich eine stabile Verteilungsform herauskristallisiert. Danach ändert sich wohl die Anzahl der SchiebereglerEinstellungen, die aber keine Auswirkung auf den Verlauf der Verteilungskurve hat.

Um nun herauszufinden, ab wie viel Teilnehmern sich eine Stabilität anstellt, werden die Teilnehmerzahlen der jeweiligen Videos in acht Blöcke mit n Teilnehmern unterteilt: 1. Block: n = 76, 2. Block: n = 176 usw. bis zum 8. Block: n = 776. Für jeden dieser Blöcke wurde die durchschnittliche SchiebereglerEinstellung pro Emotion für jedes Video ausgerechnet. Wir wollen das Vorgehen am konkreten Beispiel des Videos Nr. 20 verdeutlichen (Tabelle 4). Die durchschnittliche SchiebereglerEinstellung für Ärger (als dargestellte Basisemotion) beträgt für die ersten (willkürlich selektierten) 76 Teil-

nehmer 6,3, für 176 Personen 6,6 und so weiter bis zum endgültigen Wert bei 776 Probanden von 6,6. Das gleiche Verfahren wiederholt sich für die übrigen Gefühle. Alle Verteilungen der dargestellten wie der empfundenen Emotionen für die 20 Videos (für alle Blöcke an Teilnehmern) sind im Anhang beigefügt.

Bei den Auswertungen wird deutlich, dass eine stabile Verteilung bereits im ersten Block erreicht wird. Größere Ausreißer, die sich vom arithmetischen Mittelwert wegbewegen, werden aufgrund der Menge der Nutzer wieder „korrigiert“. Das heißt: In der Praxis des emotionalen Indexierens benötigen wir keinesfalls sehr großen Mengen an taggenden Nutzern; ganz im Gegenteil, die Verteilung der Basisemotionen „steht“ bereits bei wenigen Dutzenden von Usern. Es ergeben sich drei unterschiedliche Verteilungsformen: *erstens* finden die User keine vorherrschenden Gefühle, so dass sich gar keine Verteilung ergibt (Beispiele: Videos Nr. 7 und 15); *zweitens* zeigen sich Power Law-Verteilungen, bei der ein einziges Gefühl alle anderen dominiert (Beispiele: Videos Nr. 11, 19 und 20, jeweils bei der dargestellten Emotion), drittens ergeben sich invers-logistische Verteilungen, bei denen mehrere Items stark vertreten sind (Beispiele: Video Nr. 5 bei der dargestellten Emotion, Video Nr. 3 bei der empfundenen Emotion). Sowohl das Power Law als auch die invers-logistische Verteilung sind typische Kurvenverläufe bei Rangordnungen informationswissenschaftlicher Gegenstände (Stock, 2006).

### Power Tags bei Gefühlen (Forschungsfrage 6)

Können wir aus den Verteilungen diejenigen Basisemotionen eindeutig separieren, die das jeweilige Video am besten emotional beschreiben? Zur Lösung dieses Problems machen wir vom Konzept der „Power Tags“ Gebrauch (Peters, 2009, 363-372; Peters und Stock, 2007; Peters und Stock, 2008; Stock, 2007b; Stock und Stock, 2008, 172). „Power tags are tags that best describe the resource’s content, or the platform’s focal point of interest, according to Collective Intelligence (...), since they reflect the implicit consensus of the user community“ definiert Peters (2009, 363). Uns interessieren hier nur die Power Tags auf der Ressourcenebene. Peters (2009, 370) betont, dass die Beschränkung einer Suche auf Power Tags die Präzision dieser Recherche erhöht, da der dokumentspezifische „lange Schwanz“ an Tags nicht mehr durchsucht wird.

In Übereinstimmung mit Schmidt und Stock (2009, 871) arbeiten wir mit einem



Tabelle 5: Power Tags bei den dargestellten Basisemotionen.

Video Nr.	Dargestellte Emotion	Mittelwert	Standardabweichung	Verteilungsform	Distanz
1	Überraschung	4,5	3,60	invers-logistisch	1,0
2	keine				
3	Überraschung	7,7	3,26	invers-logistisch	1,9
	Spaß	5,8	2,81		
	Freude	4,0	1,15		
4	Angst	5,1	2,25	invers-logistisch	2,2
5	Spaß	5,7	3,40	invers-logistisch	0,3
	Überraschung	5,4	3,55		
6	Ekel	5,0	3,97	Power Law	3,1
7	keine				-
8	keine				
9	Liebe	8,1	4,04	Power Law	5,5
10	keine				
11	Ärger	8,9	3,94	Power Law	8,1
12	keine				
13	Trauer	7,4	3,17	Power Law	4,4
14	Spaß	4,7	4,02	Power Law	2,5
15	keine				
16	Liebe	7,6	4,22	Power Law	4,7
17	Sehnsucht	6,2	4,75	Power Law	2,8
18	Freude	6,2	3,83	Power Law	2,5
19	Angst	8,9	4,05	Power Law	5,3
20	Ärger	6,6	3,35	Power Law	5,8

Anmerkung: Alle Emotionen mit einer Intensität  $\geq 4$ . Distanz: Abstand zur Intensität der nächsten Emotion. N = 776 Versuchspersonen (für jedes Video).

Tabelle 6: Power Tags bei den empfundenen Basisemotionen.

Video Nr.	Empfundene Emotion	Mittelwert	Standardabweichung	Verteilungsform	Distanz
1	Spaß	5,4	3,39	Power Law	2,7
2	Ekel	6,4	3,59	Power Law	4,8
3	Überraschung	5,7	3,77	invers-logistisch	0,5
	Spaß	5,2	4,96		
4	keine				
5	Spaß	5,0	5,00	invers-logistisch	2,1
	Überraschung	2,9	3,39		
6	Ekel	7,5	4,95	Power Law	6,3
7	keine				
8	keine				
9	keine				
10	keine				
11	keine				
12	Angst	4,5	4,05	Power Law	2,4
13	Trauer	5,0	3,76	Power Law	4,2
14	Spaß	4,5	4,07	Power Law	2,5
15	keine				
16	keine				
17	keine				
18	keine				
19	Angst	5,9	4,14	Power Law	3,4
20	keine				

Anmerkung: Alle Emotionen mit einer Intensität  $\geq 4$ . Distanz: Abstand zur Intensität der nächsten Emotion. N = 776 Versuchspersonen (für jedes Video).

Schwellenwert für die Aufnahme einer Basisemotion als Kandidatin für einen Power Tag von größer oder gleich 4. Für alle Videos, bei denen nicht mindestens ein Gefühl eine SchiebereglerEinstellung von mindestens 4 erreicht, wird „keine Emotion“ notiert. Unabhängig von der Verteilungsform wird eine Gefühlsbenennung nur dann notiert, wenn die Basisemotion eine Intensität von mindestens 4 vorweisen kann. Bei den dargestellten Emotionen wird bei sechs Videos kein herausragendes Gefühl getaggt, bei den Empfindungen passen die Versuchspersonen bei elf Videos.

Da wir nur über insgesamt neun Basisemotionen verfügen, erscheint es sinnvoll, beim Vorliegen einer Power Law-Verteilung (wie in Abbildung 4) alle Terme ab dem zweiten Rangplatz als „langen Schwanz“ anzusehen und dementsprechend abzuschneiden. Eine solche Verteilung folgt der Formel  $f(x) = C / x^a$ , wobei  $x$  der Rangplatz,  $C$  eine Konstante und  $a$  ein Wert zwischen ca. 1 und ca. 2 ist. Wir nehmen zur Sicherheit einen kleinen Wert für  $a$  an (derzeit rechnen wir mit 1). Eine Kurve ist genau dann ein Power Law, wenn der Wert des Items auf Rangplatz 2 gleich oder weniger als die Hälfte des Wertes des Items auf Rang 1 ist. Für alle anderen Fälle wird als Verteilungsform „invers-logistisch“ notiert (dies entspricht nicht der Lehrmeinung in der Literatur – die den Wendepunkt der Verteilung als Schwellenwert definiert (Peters, 2009, 369) –, ist aber leicht anzuwenden). Bei den invers-logistischen Verteilungen werden alle Gefühle notiert, deren Intensitätswert größer/gleich 4 ist. So kann es durchaus passieren (wie bei den dargestellten Basisemotionen von Video 3), dass drei Gefühle indiziert werden. Alle in den Tabellen 5 und 6 genannten Gefühle werden in den Datenfeldern für dargestellte bzw. empfundene Basisemotionen in den Datensätzen der Videodokumente gespeichert und stehen so für eine Recherche nach Gefühlen zur Verfügung.

Die Standardabweichungen bei der Bewertung sowohl der dargestellten als auch der empfundenen Top-Emotionen sind ziemlich hoch – zumindest im Vergleich zu den Durchschnittswerten in den Tabellen 2 und 3. So ein Ergebnis deutet darauf hin, dass die Teilnehmer sich nicht besonders einig darüber sind, wie hoch jedes dominante Gefühl tatsächlich bewertet werden soll, obwohl prinzipiell Konsens darüber herrscht, welche Emotion vorherrschend ist (und nur diese Beobachtung ist für unsere Zwecke wichtig). Ist für einen Nutzer ein Wert von, sagen wir, 6 Ausdruck einer starken Intensität, so schiebt ein anderer den Regler auf 10, um das Gleiche auszusagen.

## Korrelationen zwischen den Basisemotionen (Forschungsfrage 7)

Welche Basisemotionen treten gemeinsam auf? Gibt es typische Gefühlskonstellationen? Zur Beantwortung dieser Fragen errechnen wir die Korrelationskoeffizienten (nach Pearson, zweiseitig) für alle untersuchten Gefühle.

Sind Gefühle ansteckend (Wild, Erb und Bartels, 2001)? Wenn ja, müssten die in den Videos dargestellten Emotionen auch von den Versuchspersonen empfunden werden. Bei den Power Tags gibt es in der Tat breite Übereinstimmungen zwischen dargestellten und empfundenen Gefühlen: Bei Video 1 ist es der Spaß, bei Videos 3 und 5 sind es Überraschung und Spaß, bei 6 Ekel, bei 13 Trauer und bei Video Nr. 19 die Angst. Die Korrelationen zwischen den gleichen Gefühlen, also zwischen als „dargestellt“ und als „empfunden“ eingeschätzten gleichen Basisemotionen, zeigt Tabelle 7. Insbesondere bei Trauer und Liebe scheinen diese Gefühle in der Tat ansteckend zu sein: Dargestellte Trauer bzw. dargestellte Liebe korrelieren bei vielen Betrachtenden mit erlebter Trauer bzw. mit Liebesgefühl. Dies gilt aber auch (und dies wegen der diversen Null-Bewertungen) für den Fall, dass ein nicht dargestelltes Gefühl mit demselben nicht empfundenen Gefühl korreliert. Wir sollten

Tabelle 7: Sind Gefühle „ansteckend“?

Basisemotion	Korrelation (dargestellt – empfunden)
Trauer	+0,47
Liebe	+0,47
Spaß	+0,39
Ekel	+0,39
Angst	+0,38
Überraschung	+0,38
Ärger	+0,32
Sehnsucht	+0,24
Freude	+0,18

die Ergebnisse in diesem Abschnitt also nicht überbewerten.

Von den vielen signifikanten Korrelationen (alle mit einer Irrtumswahrscheinlichkeit von kleiner als 0,01) sei hier nur auf die auffälligsten eingegangen (Tabelle 8). Bei der Betrachtung aller Datensätze treten die dargestellte Freude und der dargestellte Spaß, sowie die empfundene Freude und der empfundene Spaß häufig gemeinsam auf: Die Korrelationskoeffizienten liegen hier bei +0,476 und bei +0,386. Beide Emotionen scheinen bei den Umfrageteilnehmern oft gleich verwandt zu werden, weshalb die Schieberegler hier sehr ähnlich bewegt wurden.

Negative Korrelationen bestehen vorwiegend zwischen positiven und negativen Gefühlen. Dargestellte Angst steht demnach im Gegensatz zu dargestellter Freude (-0,16) und zu dargestelltem Spaß (-0,16), empfundene Angst zu dargestellter Freude (-0,15) und dargestelltem Spaß (-0,15). Analoge Gegensatzpaare liegen bei Freude und Trauer sowie bei Spaß und Trauer vor.

Die Basisemotionen Sehnsucht und Überraschung zeigen keine durchgehend klaren Korrelationen zu positiven bzw. negativen Gefühlen. Überraschung korreliert (sogar stark) positiv mit dargestelltem (+0,35) und empfundenem Spaß (+0,32), aber negativ beispielsweise mit dargestellter Liebe (-0,12) und dargestellter Trauer (-0,08). Sehnsucht korreliert positiv z.B. mit empfundener Liebe, aber negativ etwa mit (dargestelltem wie empfundenem) Ärger.

Betrachtet man nur die Werte der männlichen Versuchsteilnehmer, so finden sich die gleichen Auffälligkeiten wie zuvor, wobei diese in diesem Fall zum Teil allerdings noch stärker ausgeprägt sind. Die Korrelation zwischen dargestellter Freude und dargestelltem Spaß beträgt hierbei sogar +0,513. Bei den weiblichen Teilnehmern sind die Korrelationen im Schnitt weniger hoch.

Positive Emotionen korrelieren oft positiv mit anderen positiven Emotionen und negativ mit den negativen Emotionen. Das gleiche gilt umgekehrt auch für negative Emotionen. Allerdings gibt es auch Emotionen wie Sehnsucht und Überraschung, die stark mit positiven, aber auch mit negativen Emotionen korrelieren. Hier scheint es auf den Kontext anzukommen, der bewirkt, dass diese Emotionen eher positiv oder eher negativ eingeschätzt werden.

## Fazit und Ausblick

### Wir wollen abschließend unsere Forschungsfragen beantworten.

(Ff 1): Ist zufriedenstellendes Retrieval für Filme im Web derzeit praktisch einsetzbar? Contentbasiertes Videoretrieval befindet sich nach wie vor in den Kinderschuhen und ist in der Praxis derzeit nicht nutzbringend einsetzbar. Es gibt zwar Experimente, content-based Emotionen in Videos zu erkennen, von zufriedenstellenden Lösungen sind wir noch weit entfernt. Begriffsorientiertes Videoretrieval arbeitet entweder mit kontrolliertem Vokabular oder mit nutzergenerierten Tags. Erschließt ein einziger Indexer ein Video oder liegen narrow bzw. extended narrow Folksonomies vor, stellt sich das – hier sehr massiv auftretende – Problem der Indexierungsinkonsistenz. Für die Bewältigung der Massen an Videos im Web kommt ausschließlich der Einsatz von Folksonomies infrage. Der optimale Weg wäre demnach die Anwendung von broad Folksonomies. Diese werden aber von derzeit betriebenen Videoservices (wie YouTube) nicht verwendet. Ein zufriedenstellendes Videoretrieval im Web ist somit heutzutage nicht anzutreffen. In der Praxis brauchbares emotionales Videoretrieval gibt es überhaupt nicht.

Tabelle 8: Korrelationen zwischen den Basisemotionen.

		Alle Korrelationen																		
		Dargestellte Emotionen									Empfundene Emotionen									
		Liebe	Freude	Spaß	Überr.	Sehns.	Trauer	Ärger	Ekel	Angst	Liebe	Freude	Spaß	Überr.	Sehns.	Trauer	Ärger	Ekel	Angst	
Dargestellte Emotionen	Liebe	1																		
	Freude	0,18	1																	
	Spaß	-0,1	0,48	1																
	Überraschung	0,01	0,23	0,35	1															
	Sehnsucht	0,24	-0,02	-0,12	-0,06	1														
	Trauer	0,04	-0,14	-0,15	-0,12	0,34	1													
	Ärger	-0,12	-0,11	-0,1	-0,07	-0,12	-0,09	1												
	Ekel	-0,1	-0,06	-0,05	0,18	-0,1	-0,1	-0,07	1											
Empfundene Emotionen	Angst	-0,09	-0,16	-0,16	0,14	0,01	0,08	-0,04	0,23	1										
	Liebe	0,47	0,08	-0,07	-0,04	0,13	0,01	-0,08	-0,07	-0,07	1									
	Freude	0,15	0,18	0,2	0,12	-0,02	-0,1	-0,07	-0,06	-0,11	0,15	1								
	Spaß	-0,11	0,17	0,39	0,3	-0,12	-0,13	0	-0,05	-0,08	-0,06	0,39	1							
	Überraschung	-0,12	0,08	0,23	0,33	-0,1	-0,08	-0,05	0,09	0,11	-0,07	0,16	0,32	1						
	Sehnsucht	0,2	0	-0,1	-0,08	0,24	0,13	-0,09	-0,08	-0,07	0,29	0,03	-0,09	-0,09	1					
	Trauer	0,01	-0,12	-0,13	-0,1	0,23	0,47	-0,07	-0,09	0,03	0,06	-0,09	-0,12	-0,09	0,17	1				
	Ärger	-0,07	-0,04	-0,02	-0,03	-0,06	-0,05	0,32	-0,04	-0,02	-0,05	-0,07	-0,07	-0,03	-0,04	-0,02	1			
Ekel	-0,12	-0,12	-0,11	0,02	-0,13	-0,13	-0,07	0,39	0,08	-0,08	-0,1	-0,11	0,08	-0,1	-0,1	0,02	1			
Angst	-0,1	-0,15	-0,14	0,05	-0,1	-0,05	0,02	0,18	0,38	-0,06	-0,11	-0,11	0,12	-0,06	-0,01	0,03	0,24	1		

Anmerkung: N = 776 Versuchspersonen; 279.360 SchiebereglerEinstellungen; Korrelationen nach Pearson (zweiseitig).

(Ff 2): Was sind Basisemotionen? Voruntersuchungen zum emotionalen Retrieval (Lee und Neal, 2007, sowie Schmidt und Stock, 2009) sowie das psychologische Schrifttum zu Emotionen legen nahe, dass es mindestens fünf (Freude, Trauer, Angst, Ärger, Ekel), wahrscheinlich aber eher neun (zusätzlich: Überraschung, Sehnsucht, Liebe, Spaß) fundamentale Gefühle gibt. Emotionen haben eine Intensität, können also als schwach oder stark wahrgenommen werden.

(Ff 3): Wie können wir Basisemotionen indexieren, um sie im emotionalen Retrieval praktisch einzusetzen? Zur Abhilfe dieser misslichen Lage im Videoretrieval schlagen wir den Einsatz von kontrolliertem Vokabular (Bezeichnungen der Basisemotionen) und Schieberegler (zur Abschätzung der Intensität der Emotion) im Rahmen einer broad Folksonomy zur Indexierung des emotionalen Gehalts von Videos vor.

(Ff 4): Wie konsistent sind nutzerorientierte Verfahren beim Taggen von Emotionen in Videos? Die Konsistenz der Nutzervoten, gemessen an der Standardabweichung vom Mittelwert, ist derart hoch (ca. 1 bis 2 bei einer Skala von 0 bis zehn), dass wir von einer genügenden Übereinstimmung der Indexer ausgehen können. Einige Gefühle – allem voran Liebe – sind stark konsensfähig.

(Ff 5): Wie viele unterschiedliche Nutzer braucht man, bis sich eine stabile Verteilung der Gefühls-Tags ergibt? Gibt es überhaupt stabile Verteilungen bei Gefühls-Tags? Es gibt in der Tat stabile Verteilungen der Gefühle pro Video. Diese Verteilungen zeigen sich bereits bei geringen Nutzerzahlen (< 100) und bleiben in der Form stabil, egal, wie viele weitere User das Video taggen. Wir können drei typische Verteilungsformen separieren: keine Emotion (und damit keine Verteilung) vorhanden, Power Law-Verteilung und invers-logistische Verteilung.

(Ff 6): Können wir aus den Tagverteilungen Power Tags separieren, die den dargestellten bzw. empfundenen Gefühlen entsprechen? Bei der Verteilung der Basisemotionen nach Power Law ergibt sich genau ein Gefühl, das dem Dokument als emotionaler Zugangspunkt zugeordnet wird. Bei invers-logistischen Verteilungen sind es mehrere Gefühle (zwei oder drei), die dem Video zukommen. Wir müssen zwischen dargestellten und empfundenen Emotionen unterscheiden, da die Nutzer diese auch unterschiedlich beurteilen.

(Ff 7): Korrelieren gewisse Emotionen miteinander? Bei den Korrelationen unterscheiden wir zwischen der Analyse gleicher Emotionen bei Darstellung und

Empfindung sowie der Analyse der Zusammenhänge zwischen unterschiedlichen Gefühlen. „Ansteckende“ Basisemotionen (also solche mit hohen Korrelationswerten zwischen dargestellt und empfunden) sind vor allem Trauer und Spaß. Gefühlsübergreifend positiv korrelieren besonders ausgeprägt Freude und Spaß, negativ beispielsweise Angst mit Freude und mit Spaß.

**Ausblick.** Wir können anhand der Resultate unseres groß angelegten Indexierungsexperiments mit hoher Wahrscheinlichkeit annehmen, dass Nutzer in der Lage sind, (vorgegebene) Basisemotionen bei Videos konsistent zu erkennen. Um stabile Verteilungen der Gefühle pro Video zu erhalten, benötigen wir nur recht kleine Zahlen (kleiner hundert) an beteiligten Usern. Diese Ergebnisse stehen im Einklang mit den Resultaten beim emotionalen Bilderindexieren (Schmidt und Stock, 2009). Dies ermutigt uns, als nächsten Schritt im EmIR-Forschungsprogramm die Entwicklung eines Prototyps eines Retrievalsystems für emotionale Dokumente – als „Media Emotion Search (Memose)“ – in Angriff zu nehmen.

**Danksagung.** Alexander Kochanek, Natalia Mamerow und Michael Mamerow halfen bei der Vorbereitung und Durchführung der empirischen Untersuchung.

## Literatur

- Aigrain, P., Zhang, H., & Petkovic, D. (1996). Content-Based Representation and Retrieval of Visual Media: A State-of-the-Art Review. In: *Multimedia Tools and Applications*, 3(4), 179-202.
- ARD/ORF/ZDF (2008). Regelwerk Mediendokumentation: Fernsehen. Richtlinien für die Formalbeschreibung und Inhaltserschließung von Fernsehproduktionen (Version 1.0, Stand: 9. April 2008).
- Arnold, M.B. (1960). *Emotion and Personality*. New York: Columbia University Press.
- Bailenson, J.N., Pontikakis, E.D., Mauss, I.B., Gross, J.J., Jabon, M.E., Hutcherson, C.A., Nass, C., & John, O. (2008). Real-time Classification of Evoked Emotions Using Facial Feature Tracking. In: *International Journal Human-Computer Studies*, 66(5), 303-317.
- Boehner, K., DePaula, R., Dourish, P., & Sengers, P. (2007). How Emotion is Made and Measured. In: *International Journal of Human-Computer Studies*, 65(4), 275-291.
- Bornwasser, M., & Mummendey, A. (1983). Ärger. In: Euler, H., & Mandl, H. (Eds.): *Emotionspsychologie. Ein Handbuch in Schlüsselbegriffen* (S. 156-162). München: Urban & Schwarzenberg.
- Burch, D. (2008). How Much of a Typical Video Online Is Actually Watched? [www.tubemogul.com/research/report/18](http://www.tubemogul.com/research/report/18) [10.5.2009]
- Chan, C.H., & Jones, G.J.F. (2005). Affect-Based Indexing and Retrieval of Films. In: *Proceedings of the 13th Annual ACM International Conference on Multimedia* (S. 427-430). New York: ACM.
- Chen, L., Chen, G.C., Xu, C.Z., March, J., & Benford, S. (2008). EmoPlayer: A Media Player for Video Clips with Affective Annotations. In: *Interacting with Computers*, 20(1), 17-28.
- Colombo, C., Del Bimbo, A., & Pala, P. (1999). Semantics in Visual Information Retrieval. In: *IEEE Multimedia*, 6(3), 38-53.

- Darwin, C. (1872). *The Expression of the Emotions in Man and Animals*. London: John Murray.
- Del Bimbo, A. (1999). *Visual Information Retrieval*. London: Academic Press.
- Ekman, P., & Friesen, W.V. (1971). Constants Across Cultures in the Face and Emotion. In: *Journal of Personality and Social Psychology*, 17, 124-129.
- Ekman, P., Friesen, W.V., & Ellsworth, P. (1982). What Emotion Categories or Dimensions Can Observers Judge from Facial Behavior? In: Ekman, P. (Ed.): *Emotion in the Human Face* (S. 39-55). New York: Cambridge University Press.
- Enser, P. (2008a). Visual Image Retrieval. In: *Annual Review of Information Science and Technology*, 42, 3-42.
- Enser, P. (2008b). The Evolution of Visual Information Retrieval. In: *Journal of Information Science*, 34(4), 531-546.
- Enser, P., Sandom, C., Hare, J., & Lewis, P. (2007). Facing the Reality of Semantic Image Retrieval. In: *Journal of Documentation*, 63(4), 465-481.
- Fehr, B., & Russell, A. (1984). Concept of Emotion Viewed from a Prototype Perspective. In: *Journal of Experimental Psychology: General*, 113, 464-486.
- Frijda, N. H. (1986). *The Emotions*. New York: Cambridge University Press.
- Geisler, G., & Burns, S. (2007). Tagging Videos: Conventions and Strategies of the YouTube Community. In: *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries* (S. 480). New York: ACM.
- Goodrum, A.A. (2000). Image Information Retrieval: An Overview of Current Research. In: *Informing Science*, 3(2), 63-67.
- Gupta, A., & Jain, R. (1997). Visual Information Retrieval. In: *Communications of the ACM*, 40(5), 71-79.
- Halpin, H., Robu, V., & Shepherd, H. (2007). The Complex Dynamics of Collaborative Tagging. In: *Proceedings of the 16th International World Wide Web Conference (WWW'07)*, Banff, Canada, 2007 (S. 211-220). New York: ACM.
- Hanjalic, A., & Xu, L. (2005). Affective Video Content Representation and Modeling. In: *IEEE Transactions on Multimedia*, 7(1), 143-154.
- Hoffmann, N. (1983). Trauer. In: Euler, H., & Mandl, H. (Eds.): *Emotionspsychologie. Ein Handbuch in Schlüsselbegriffen* (S. 183-188). München: Urban & Schwarzenberg.
- Izard, C. E. (1994). *Die Emotionen des Menschen. Eine Einführung in die Grundlagen der Emotionspsychologie*. Weinheim: Psychologie Verlags Union.
- Izard, C. E. (1977). *Human Emotions*. New York: Plenum Press.
- Izard, C. E. (1971). *The Face of Emotion*. New York: Appleton-Century-Crofts.
- James, W. (1884). What is an Emotion? In: *Mind*, 9, 188-205.
- Jörgensen, C. (1998). Attributes of Images in Describing Tasks. In: *Information Processing & Management*, 34(2/3), 161-174.
- Jörgensen, C. (2007). Image Access, the Semantic Gap, and Social Tagging as a Paradigm Shift. In: *Proceeding of the 18th Workshop of the American Society for Information Science and Technology Special Interest Group in Classification Research*, Milwaukee, WI.
- Kim, H., & Kim, Y. (2009). A Two-Step Model for Video Key-Frame Determination. In: *Proceedings of the 72nd ASIS&T Annual Meeting. Thriving on Diversity – Information Opportunities in a Pluralistic World*.
- Kim, M.G., & Park, Y.T. (2003). Information Retrieval Agents Using Emotional Features. In: Palau Salvador, C.E. (Ed.): *Proceedings of the Second IASTED International Conference on Communication Systems and Networks*. Anaheim: Acta Press.
- Kleinginna, P.R., & Kleinginna, A.M. (1981). A Categorized List of Emotion Definitions, with Suggesti-

ons for a Consensual Definition. In: *Motivation and Emotion*, 5(4), 345-379.

Laine-Hernandez, M., & Westman, S. (2006). Image Semantics in the Description and Categorization of Journalistic Photographs. In: *Proceedings of the 69th Annual Meeting of the American Society for Information and Technology. Information Realities: Shaping the Digital Future of All*.

Lee, H.J., & Neal, D. (2007). Toward Web2.0 Music Information Retrieval: Utilizing Emotion-Based, User-Assigned Descriptors. In: *Proceedings of the 70th ASIS&T Annual Meeting. Joining Research and Practice: Social Computing and Information Science*.

Markey, K. (1984). Interindexer Consistency Tests: A Literature Review and Report of a Test of Consistency in Indexing Visual Materials. In: *Library & Information Science Research*, 6, 155-177.

Markkula, M., & Sormunen, E. (2000). End-User Searching Challenges Indexing Practices in the Digital Newspaper Photo Archive. In: *Information Retrieval*, 1(4), 259-285.

McDougall, W. (1908/1960). *An Introduction to Social Psychology*. Boston, Mass.: Methuen.

McDougall, W. (1926). *An Outline of Abnormal Psychology*. Boston, Mass.: Luce.

Melenhorst, M., Grootveld, M., van Setten, M., & Veenstra, M. (2008). Tag-Based Information Retrieval of Video Content. In: *Proceedings of the 1st International Conference on Designing Interactive User Experiences for TV and Video* (S. 31-40). New York: ACM.

Meyer, W.U., Reisenzein, R., & Schützwohl, A. (2001). *Einführung in die Emotionspsychologie*. Band I: Die Emotionstheorien von Watson, James und Schachter. Bern: Verlag Hans Huber.

Ortony, A., & Turner, T.J. (1990). What's Basic About Basic Emotions? In: *Psychological Review*, 97, 315-331.

Panksepp, J. (1982). Toward a General Psychobiological Theory of Emotions. In: *The Behavioral and Brain Sciences*, 5, 407-467.

Peters, I. (2009). *Folksonomies. Indexing and Retrieval in Web 2.0*. Berlin: De Gruyter Saur. (Knowledge and Information. Studies in Information Science).

Peters, I., & Stock, W.G. (2007). *Folksonomy and Information Retrieval*. In: *Proceedings of the 70th ASIS&T Annual Meeting. Joining Research and Practice: Social Computing and Information Science*.

Peters, I., & Stock, W.G. (2008). *Folksonomies in Wissensrepräsentation und Information Retrieval*. In: *Information – Wissenschaft und Praxis*, 59(2), 77-90.

Petrelli, D., & Auld, D. (2008). An Examination of Automatic Videoretrieval Technology on Access to the Contents of an Historical Video Archive. In: *Program: Electronic Library and Information Systems*, 42(2), 115-136.

Picard, R.W. (1995). *Affective Computing*. M.I.T Media Laboratory Perceptual Computing Section Technical Report, 321. Cambridge, Mass.: MIT.

Plutchik, R. (1980). A General Psychoevolutionary Theory of Emotion. In: *Emotion: Theory, Research, and Experience*, Vol. 1: Theories of Emotion (S. 3-33). New York: Academic.

Rohde, U., & Lustig, S. (2007). Die takeGas-Studie von DoubleClick und der TOMORROW FOCUS AG. DoubleClick: Digital Advertising: [http://emea.doubleclick.com/DE/downloads/pdfs/070827\\_Berichtsband%20takeGAS-Studie\\_final.pdf](http://emea.doubleclick.com/DE/downloads/pdfs/070827_Berichtsband%20takeGAS-Studie_final.pdf) [10.5.2009]

Rosch, E. (1975). Cognitive Reference Points. In: *Cognitive Psychology*, 7(4), 532-547.

Salway, A., & Graham, M. (2003). Extracting Information about Emotions in Films. In: *Proceedings of the 11th ACM International Conference on Multimedia* (pp. 299-302). New York: ACM.

Schmidt, S., & Stock, W.G. (2009). Collective Indexing of Emotions in Images. A Study in Emotional Information Retrieval. In: *Journal of the American Society for Information Science and Technology*, 60(5), 863-876.

Schmidt-Atzert, L. (1996). *Lehrbuch der Emotionspsychologie*. Stuttgart: Kohlhammer.

Schwarzer, R. (1983). Angst und Furcht. In: Euler, H., & Mandl, H. (Eds.), *Emotionspsychologie. Ein Handbuch in Schlüsselbegriffen* (S. 147-154). München: Urban & Schwarzenberg.

Schweins, K. (1997). *Methoden zur Erschließung von Filmsequenzen*. Köln: FH Köln. (Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft; 5).

Schwender, C. (2001). *Medien und Emotionen*. Wiesbaden: Deutscher Universitäts-Verlag.

Smeaton, A.F. (2007). Techniques Used and Open Challenges to the Analysis, Indexing and Retrieval of Digital Video. In: *Information Systems*, 32(4), 545-559.

Smeaton, A.F., Wilkins, P., Worring, M., de Rooij, O., Chua, T.S., & Luan, H. (2008). Content-Based Videoretrieval: Three Example Systems from TRECVID. In: *International Journal of Imaging Science and Technology*, 18(2-3), 195-201.

Smith, G. (2008). *Tagging. People-Powered Metadata for the Social Web*. Berkeley, Cal.: New Riders.

Snoek, C.G.M., & Worring, M. (2009). Concept-Based Videoretrieval. In: *Foundations and Trends in Information Retrieval*, 4(2), 215-322.

Sokolowski, K. (2002). *Emotion*. In: Müsseler, J. & Prinz, W. (Eds.): *Allgemeine Psychologie* (S. 337-384). Heidelberg, Berlin: Spektrum Akademischer Verlag.

Soleymani, M., Chanel, G., Kierkels, J.J.M., & Pun, T. (2008). Affective Ranking of Movie Scenes Using Physiological Signals and Content Analysis. In: *Proceedings of the 2nd ACM Workshop on Multimedia Semantics* (pp. 32-39). New York: ACM.

Stock, W.G. (2006). On Relevance Distributions. In: *Journal of the American Society for Information Science and Technology*, 57(8), 1126-1129.

Stock, W.G. (2007a). *Information Retrieval. Informationen suchen und finden*. München, Wien: Oldenbourg.

Stock, W.G. (2007b). *Folksonomies and Science Communication. A Mash-Up of Professional Science Databases and Web 2.0 Services*. In: *Information Services & Use*, 27, 97-103.

Stock, W.G., & Stock, M. (2008). *Wissensrepräsentation. Informationen auswerten und bereitstellen*. München: Oldenbourg.

Tomkins, S.S. (1962). *Affect, Imagery, Consciousness, Vol. I: The Positive Affects*. New York: Springer Publishing.

Tomkins, S.S. (1963). *Affect, Imagery, Consciousness, Vol. II: The Negative Affects*. New York: Springer Publishing.

Van Velsen, L., & Melenhorst, M. (2009). Incorporating User Motivations to Design for Video Tagging. In: *Interacting with Computers*, 21(3), 221-232.

Vendrell Ferran, I. (2008). *Die Emotionen. Gefühle in der realistischen Phänomenologie*. Berlin: Akademie Verlag.

Watson, J. B. (1930). *Behaviorism*. Chicago, Ill.: University of Chicago Press.

Wild, B., Erb, M., & Bartels, M. (2001). Are Emotions Contagious? Evoked Emotions While Viewing Emotionally Expressive Faces: Quality, Quantity, Time Course and Gender Differences. In: *Psychiatry Research*, 102(2), 109-124.

Weller, K., Peters, I., & Stock, W.G. (2010). *Folksonomy. The Collaborative Knowledge Organization System*. In: Dumova, T., & Fiordo, R. (Eds.): *Handbook of Research on Social Interaction Technologies and Collaborative Software: Concepts and Trends* (S. 132-146). Hershey, PA: Information Science Reference.

Weiner, B., & Graham, S. (1984). An Attributional Approach to Emotional Development. In: Izard, E., Kagan, J., & Zajonc, R. (Eds.): *Emotions, Cognition, and Behavior* (S. 167-191). New York: Cambridge University Press.

Wilhelm, K. (2008). *Zwischenstation Sehnsucht*. In: *MaxPlanckForschung*, Nr. 3, 66-72.

Xu, M., Chia, L., & Jin, J. (2005). Affective Content Analysis in Comedy and Horror Videos by Audio Emotional Event Detection. In: *IEEE International Conference on Multimedia and Expo (ICME 2005)*.

Information Retrieval, Film, inhaltliche Erschließung, empirische Untersuchung, Folksonomy, Schieberegler, emotionales Retrieval, EMLR, Basisemotion, Gefühl, Tagging, Tag-Verteilung, Power Tag


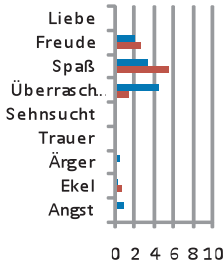
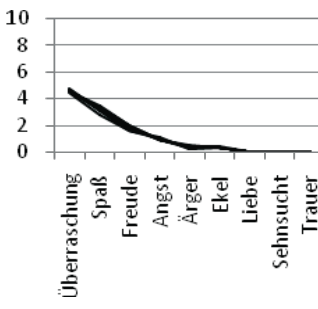
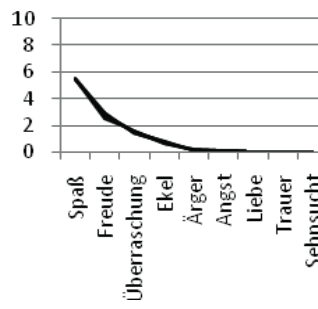

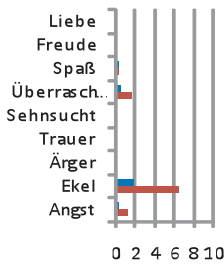
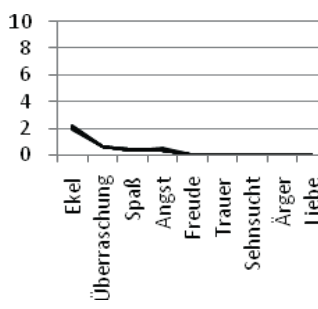
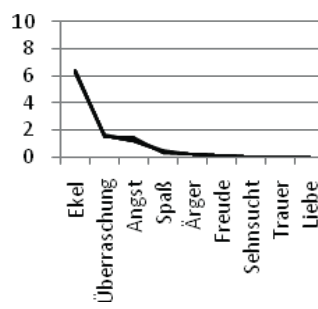

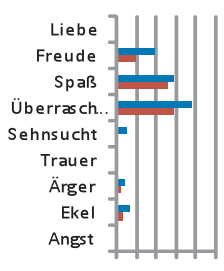
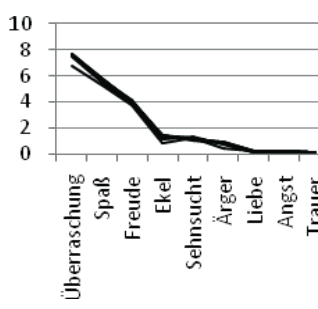
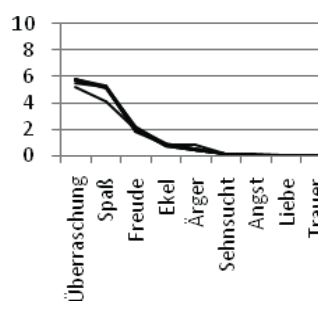

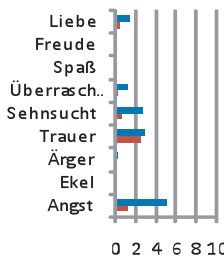
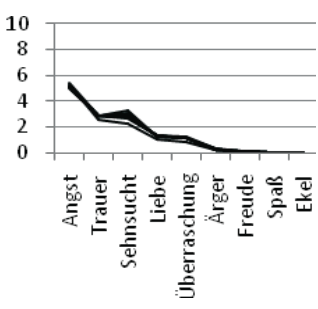
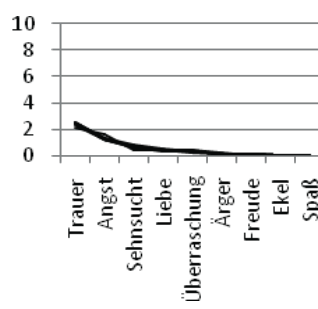

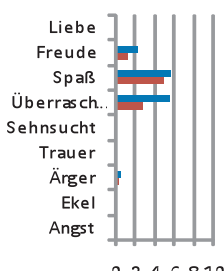
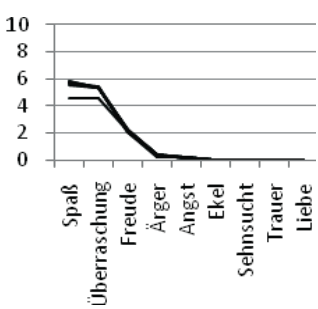
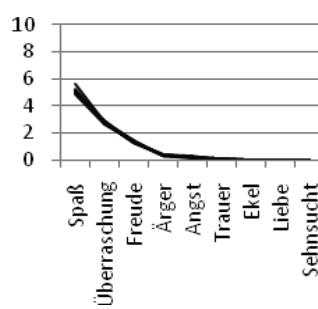
## DAS AUTORENTEAM


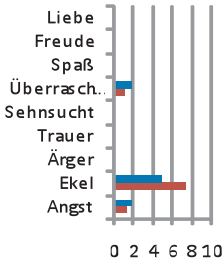
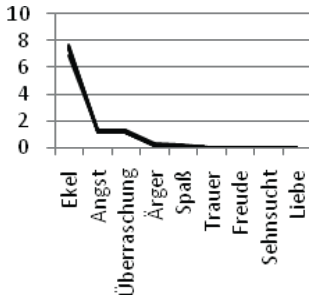

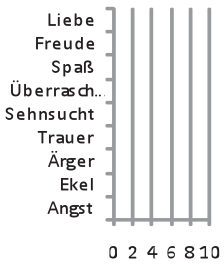
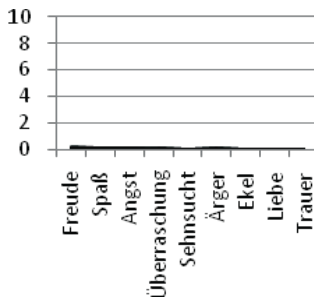
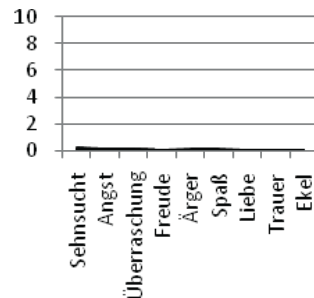

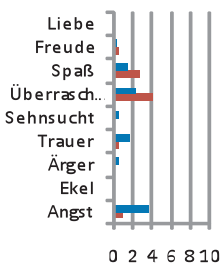
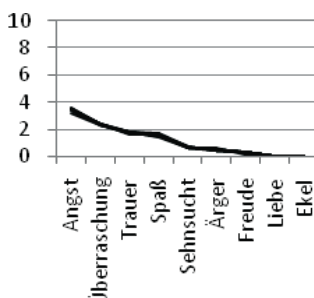
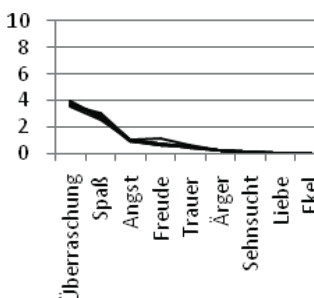

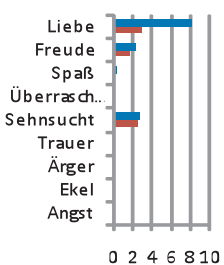
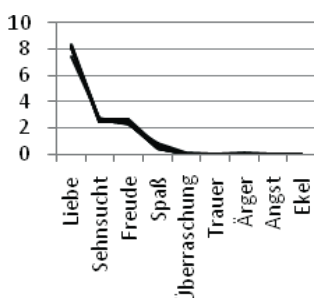
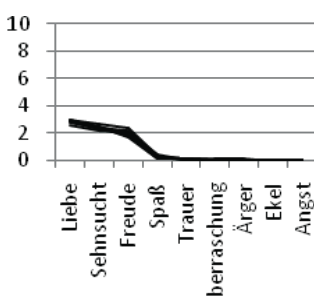

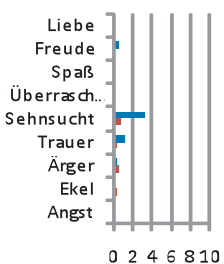
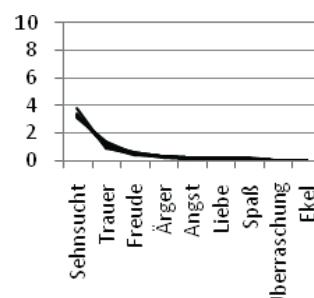
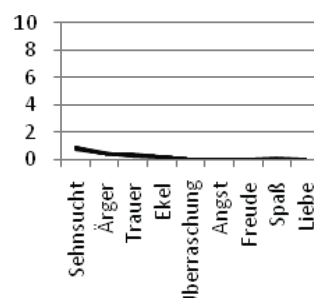
Evelyn Dröge, Susanne Finkelmeyer, Daniel Guschauski, Kerstin Juchem, Kathrin Knautz, Cynthia Krzmyk, Daniel Miskovic, Jan Schiefer, Evrim Sen, Julia Verbina und Nils Werner sind Master- oder Promotionsstudierende der Informationswissenschaft an der Heinrich-Heine-Universität Düsseldorf.


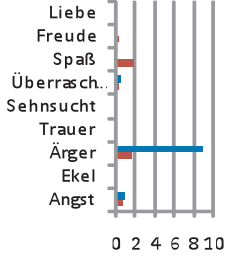
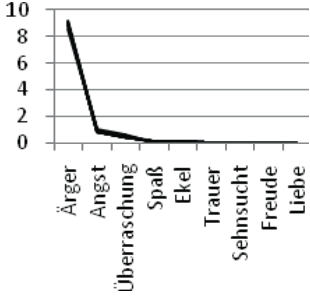
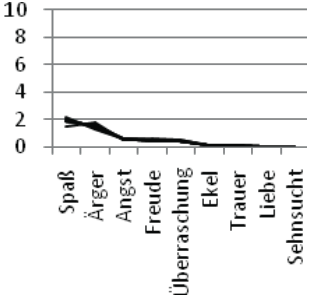

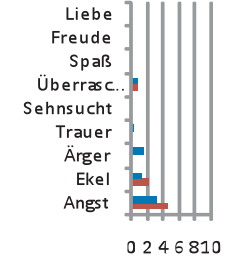
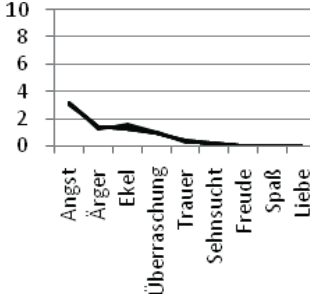
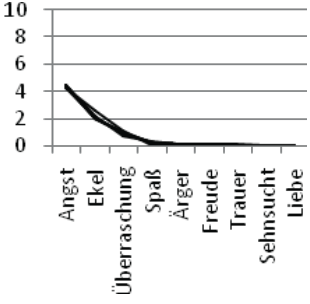

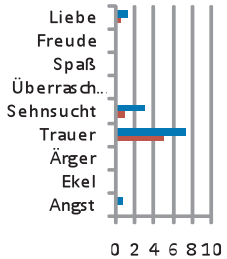
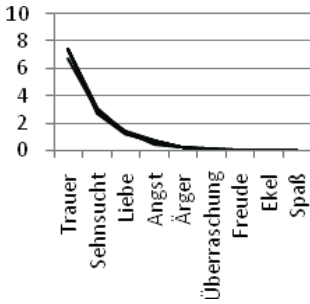
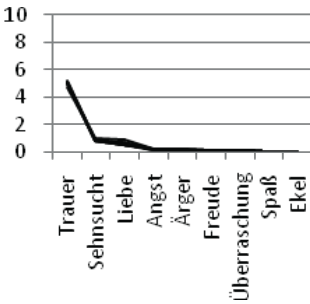

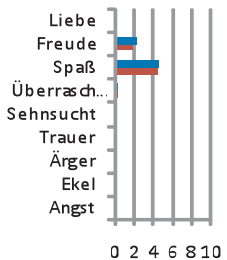
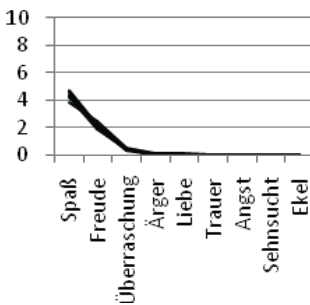
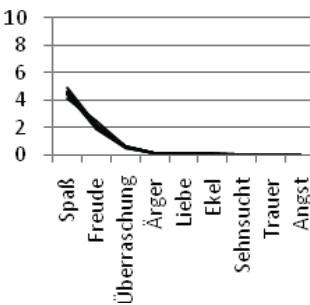

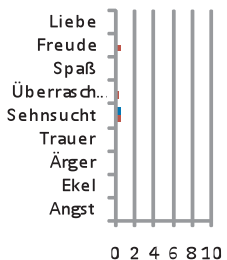
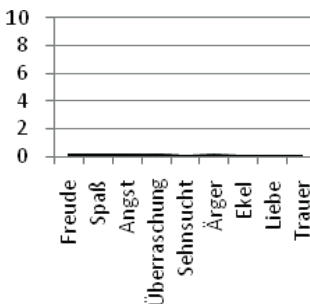
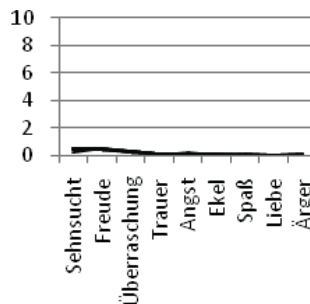
Wolfgang G. Stock ist Leiter der Abteilung für Informationswissenschaft in Düsseldorf.


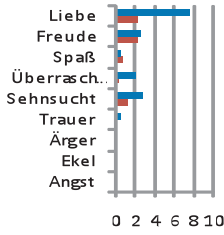
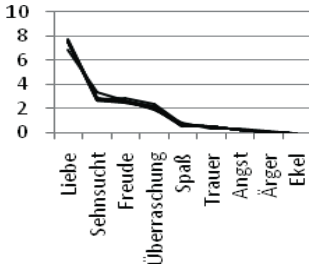
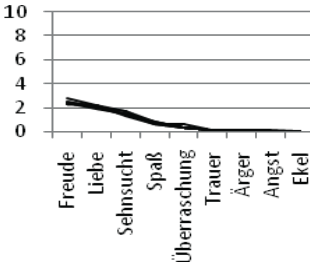

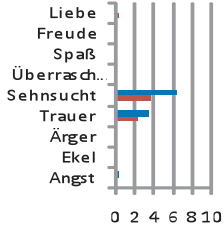
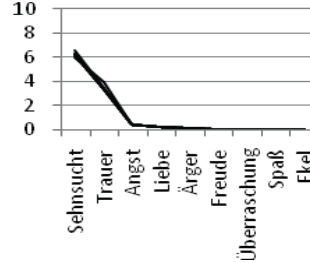
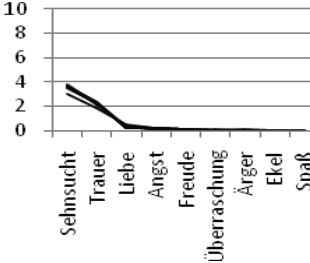

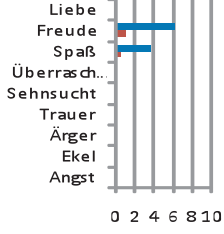
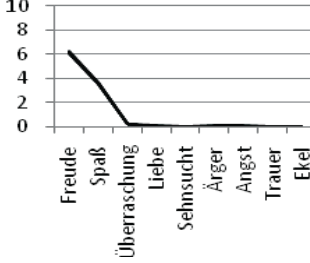
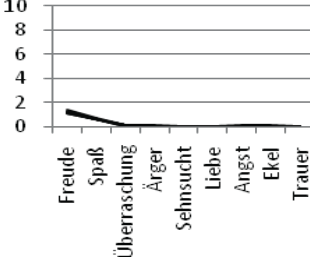

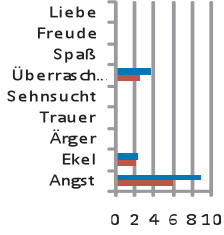
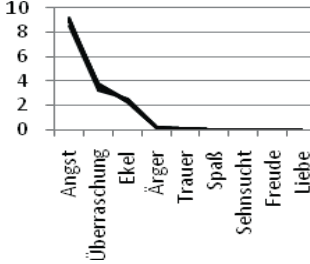
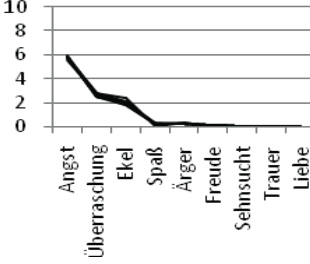
Kontakt:  
Kathrin.Knautz@uni-duesseldorf.de

Anhang

VIDEO-DETAILS	GESAMT-ERGEBNIS DER UMFRAGE	DARGESTELLTE EMOTIONEN	EMPFUNDENE EMOTIONEN
<p><b>VIDEO 1</b>                      Länge: 23 Sek.                      Musik: ja                      Quelle: Youtube  <a href="https://www.youtube.com/watch?v=6HyDgu4UNTk">youtube.com/watch?v=6HyDgu4UNTk</a></p> 			
<p><b>VIDEO 2</b>                      Länge: 17 Sek.                      Musik: nein                      Quelle: Youtube  <a href="https://www.youtube.com/watch?v=FCMJwKI4xU">youtube.com/watch?v=FCMJwKI4xU</a></p> 			
<p><b>VIDEO 3</b>                      Länge: 24 Sek.                      Musik: nein                      Quelle: Youtube  <a href="https://www.youtube.com/watch?v=Zb_F9Xg2X7s">youtube.com/watch?v=Zb_F9Xg2X7s</a></p> 			
<p><b>VIDEO 4</b>                      Länge: 23 Sek.                      Musik: ja                      Quelle: Youtube  <a href="https://www.youtube.com/watch?v=iF_J9rLXuTA">youtube.com/watch?v=iF_J9rLXuTA</a></p> 			
<p><b>VIDEO 5</b>                      Länge: 9 Sek.                      Musik: nein                      Quelle: Youtube  <a href="https://www.youtube.com/watch?v=YTFY5NG67aQ">youtube.com/watch?v=YTFY5NG67aQ</a></p> 			

VIDEO-DETAILS	GESAMT-ERGEBNIS DER UMFRAGE	DARGESTELLTE EMOTIONEN	EMPFUNDENE EMOTIONEN
<p><b>VIDEO 6</b>                      Länge: 16 Sek.                      Musik: ja                      Quelle: Youtube  <a href="https://www.youtube.com/watch?v=tDV5-wxw7Tc">youtube.com/watch?v=tDV5-wxw7Tc</a>  </p>			
<p><b>VIDEO 7</b>                      Länge: 8 Sek.                      Musik: nein                      Quelle: Youtube  <a href="https://www.youtube.com/watch?v=fEnx1Yc_0Y">youtube.com/watch?v=fEnx1Yc_0Y</a>  </p>			
<p><b>VIDEO 8</b>                      Länge: 20 Sek.                      Musik: ja                      Quelle: Youtube  <a href="https://www.youtube.com/watch?v=0GgYumfE2GY">youtube.com/watch?v=0GgYumfE2GY</a>  </p>			
<p><b>VIDEO 9</b>                      Länge: 21 Sek.                      Musik: ja                      Quelle: Youtube  <a href="https://www.youtube.com/watch?v=YqU8QLxQFLO">youtube.com/watch?v=YqU8QLxQFLO</a>  </p>			
<p><b>VIDEO 10</b>                      Länge: 11 Sek.                      Musik: nein                      Quelle: Youtube  <a href="https://www.youtube.com/watch?v=q2PZLwbicQQ">youtube.com/watch?v=q2PZLwbicQQ</a>  </p>			

VIDEO-DETAILS	GESAMT-ERGEBNIS DER UMFRAGE	DARGESTELLTE EMOTIONEN	EMPFUNDENE EMOTIONEN
<p><b>VIDEO 11</b>                      Länge: 21 Sek.                      Musik: nein                      Quelle: Youtube  <a href="https://youtube.com/watch?v=jdhkUL-5F14">youtube.com/watch?v=jdhkUL-5F14</a>  </p>			
<p><b>VIDEO 12</b>                      Länge: 21 Sek.                      Musik: nein                      Quelle: Youtube  <a href="https://youtube.com/watch?v=jMjnH8H5tJ8">youtube.com/watch?v=jMjnH8H5tJ8</a>  </p>			
<p><b>VIDEO 13</b>                      Länge: 17 Sek.                      Musik: nein                      Quelle: Youtube  <a href="https://youtube.com/watch?v=yPoJ6tYw_pE">youtube.com/watch?v=yPoJ6tYw_pE</a>  </p>			
<p><b>VIDEO 14</b>                      Länge: 21 Sek.                      Musik: nein                      Quelle: Youtube  <a href="https://youtube.com/watch?v=ZLif6N7LA4s">youtube.com/watch?v=ZLif6N7LA4s</a>  </p>			
<p><b>VIDEO 15</b>                      Länge: 16 Sek.                      Musik: ja                      Quelle: Youtube  <a href="https://youtube.com/watch?v=u8U32srGqQc">youtube.com/watch?v=u8U32srGqQc</a>  </p>			

VIDEO-DETAILS	GESAMT-ERGEBNIS DER UMFRAGE	DARGESTELLTE EMOTIONEN	EMPFUNDENE EMOTIONEN
<p><b>VIDEO 16</b>                      Länge: 21 Sek.                      Musik: nein                      Quelle: Youtube  <a href="https://www.youtube.com/watch?v=MujfX70w42w">youtube.com/watch?v=MujfX70w42w</a></p> 			
<p><b>VIDEO 17</b>                      Länge: 22 Sek.                      Musik: ja                      Quelle: Youtube  <a href="https://www.youtube.com/watch?v=kZa7OYWPFpc">youtube.com/watch?v=kZa7OYWPFpc</a></p> 			
<p><b>VIDEO 18</b>                      Länge: 5 Sek.                      Musik: ja                      Quelle: Youtube  <a href="https://www.youtube.com/watch?v=aCDgdbywRZI">youtube.com/watch?v=aCDgdbywRZI</a></p> 			
<p><b>VIDEO 19</b>                      Länge: 21 Sek.                      Musik: ja                      Quelle: Youtube  <a href="https://www.youtube.com/watch?v=1tEtA28T39M">youtube.com/watch?v=1tEtA28T39M</a></p> 			
<p><b>VIDEO 20</b>                      Länge: 21 Sek.                      Musik: ja                      Quelle: Youtube  <a href="https://www.youtube.com/watch?v=sBkLhQ1FMb4">youtube.com/watch?v=sBkLhQ1FMb4</a></p> 