

LAURA COMMARE (MÜNCHEN)

## **SOCIAL TAGGING ALS METHODE ZUR OPTIMIERUNG KUNSTHISTORISCHER BILDDATENBANKEN – EINE EMPIRISCHE ANALYSE DES ARTIGO-PROJEKTS**

### **Zusammenfassung**

Das Artigo-Projekt der LMU München nutzt einen Social-Tagging-Ansatz in Form eines Online-Spieles, um den Schlagwortkatalog der Bilddatenbank Artemis zu optimieren. Im Rahmen der vorliegenden Untersuchung wurde eine Teil-Evaluierung des Projekts durchgeführt. Diese sollte im Kern die Frage klären, ob der gewählte Ansatz prinzipiell in der Lage ist, qualitativ hochwertige Daten zu produzieren.

### **1. Einleitung**

<1>

Jedes Jahr wächst die Anzahl der digitalen Daten aller Art um ein Vielfaches an – auch auf wissenschaftlicher Ebene existieren große Ansammlungen von fachspezifischen Daten im Netz.<sup>1</sup> Daher ist heute jede Datenrecherche, um erfolgreich zu sein, auf Suchmaschinen angewiesen, die in der Lage sind, den Zugriff auf Wissensressourcen zu automatisieren.<sup>2</sup> Die Recherche wäre ohne Suchmaschinen höchst voraussetzungsreich. Das gesuchte Objekt müsste in diesem Fall bereits bekannt sein, zudem dessen virtueller Standort. Virtuelle Räume sind jedoch mit einem ungeordneten Durcheinander zu vergleichen,<sup>3</sup> in dem ein Zufallstreffer unwahrscheinlich ist.<sup>4</sup> Aufgrund dieser Schwierigkeiten kommt den diversen Suchmaschinen im virtuellen Raum eine große Bedeutung zu.

<2>

Alle Suchmaschinen funktionieren in der Regel nach dem gleichen Prinzip. Sie bedienen sich eines Schlagwortindexes, dem eine Objektbasis, wie etwa Reproduktionen von Kunstwerken, zugrunde liegt. Anhand der Schlagworte durchsuchen sie den Bestand der Objekte.<sup>5</sup> Als Ergebnis werden diejenigen Objekte präsentiert, an welche im Vorfeld die gesuchten Schlagworte vergeben wurden. Durch eine fachspezifische Einschränkung lässt sich der Aufwand bei der Suche erheblich reduzieren, da in diesem Fall nur solche Ergebnisse präsentiert werden, die einen thematischen Bezug zur Suchanfrage haben und somit ein weiteres Sortieren der Ergebnisse nur in vergleichsweise kleinem Umfang notwendig ist. Das Ziel eines hochwertigen Suchdienstes ist es natürlich, auf eine Anfrage

hin möglichst viele Ergebnisse von hoher Relevanz zu präsentieren. Die Relevanz der durch die Suchfunktion ermittelten Objekte stellt also das wichtigste Kriterium für eine erfolgreiche Suche dar.<sup>6</sup>

<3>

Das eigentliche Augenmerk der in diesem Essay vorgestellten Überlegungen gilt der vertikal eingeschränkten Suche<sup>7</sup> in kunsthistorischen Datenbanken, wobei insbesondere die den Suchmaschinen zugrunde liegenden Schlagwortkataloge in der Analyse betrachtet werden. In der Regel folgen Schlagwortkataloge starren Vorgaben, das heißt gesucht werden kann nur nach vorab festgelegten Informationen. In der Kunstgeschichte betreffen diese Informationen üblicherweise den Künstler, den Bildtitel, die Datierung und den Aufbewahrungsort. Eine ikonographische Verschlagwortung hingegen, die für die Auffindbarkeit von Werken als essentiell betrachtet werden kann, ist in der Regel kaum vorhanden, so dass die Bilddatenbanken nicht exakt durchsucht werden können.<sup>8</sup> Eine tiefergreifende Erschließung der Datenbanken ist aber insofern wünschenswert, da erst dadurch der eigentliche Mehrwert digitaler Archive realisiert wird. Eine ausführlichere, inhaltliche Erfassung der Abbildungen würde sehr viel Zeit in Anspruch nehmen und es bedürfte einer großen Personengruppe, um sämtliche thematischen Facetten einer Reproduktion zu erfassen und passende Begriffe zuzuordnen.<sup>9</sup> An diesem Punkt stellt sich dann allerdings die Frage nach der Finanzierung eines solch aufwendigen Projekts – es scheint auf den ersten Blick nicht ohne Weiteres realisierbar zu sein.<sup>10</sup> Am Institut für Kunstgeschichte der Ludwig-Maximilians-Universität München gemeinsam mit der IT-Gruppe Geisteswissenschaften, dem Institut für Informatik und dem Centrum für Sprachverarbeitung wurde unter der Leitung von Hubertus Kohle jüngst dennoch der Versuch unternommen, ein solches Verschlagwortungs-Projekt im Internet zu realisieren. Im Rahmen des interdisziplinären Forschungsprojektes Play4Science (<http://www.play4science.uni-muenchen.de/>), welches von der Deutschen Forschungsgemeinschaft (DFG) gefördert wird, wurde ein Projekt mit dem Namen *Artigo* (<http://www.artigo.org/>)<sup>11</sup> konzipiert. Hierbei handelt es sich um eine Social Tagging-Anwendung, die es ermöglicht, eine Vielzahl von Menschen an der detaillierten thematischen Verschlagwortung von kunsthistorischen Abbildungen zu beteiligen. Social Tagging bezeichnet die freie, das heißt beliebige, Schlagwortvergabe (Tags) durch eine Gruppe von Menschen, wobei es sich ausdrücklich nicht um Experten handeln muss. Die jeweils vergebenen Tags werden aggregiert und bilden so einen Schlagwortindex.<sup>12</sup>

<4>

Der vorliegende Essay beschäftigt sich mit den im Rahmen des Artigo-Projekts entstehenden Daten. Untersucht werden deren Qualität und die Einflüsse der Social-Tagging-basierten Methode auf diese. Daten von geringer Qualität stellen im wissenschaftlichen Kontext keinen Zugewinn dar. Die Fragestellung der Untersuchung lautet daher: Liefert die Artigo-Anwendung Daten mit wissenschaftlicher Relevanz? Die Frage an sich ist selbstverständlich problematisch, denn: Wann sind Daten objektiv wissenschaftlich relevant? Um dennoch eine Antwort zu entwerfen, wird die wissenschaftliche Bedeutung anhand professionell erzeugter Daten gemessen, die, da sie von Experten erschaffen worden sind, als qualitativ hochwertig gelten können. Demnach lautet die Fragestellung eigentlich: Ist die Artigo-Anwendung in der Lage, Daten zu produzieren, die genauso hochwertig sind wie professionell erzeugte Daten?

<5>

Im folgenden Kapitel 2 wird zunächst ausführlich das Artigo-Projekt vorgestellt.

In Kapitel 3 werden die theoretischen Ansätze erläutert, die den vorgestellten Überlegungen zugrunde liegen. Dieses Kapitel ist unterteilt in zwei Abschnitte: Der erste Abschnitt befasst sich mit der Frage, welche Voraussetzungen auf der Systemebene von Social Tagging-Anwendungen gegeben sein müssen, um einen qualitativ-hochwertigen Schlagwortkatalog zu produzieren. Der zweite Abschnitt behandelt den Tagging-Vorgang selbst, wobei hier der Akteur und das individuelle Vorgehen bei der Verschlagwortung im Mittelpunkt stehen. Aus den theoretischen Überlegungen beider Abschnitte werden Thesen bzw. Leitfragen abgeleitet, die anschließend in Kapitel 4 mit Hilfe verschiedener empirischer Studien überprüft werden. Hierbei wird zuerst ein Überblick über das Studiendesign und die Operationalisierung der Thesen gegeben, um daran anschließend, nach Themen geordnet, die Ergebnisse vorzustellen.

## **2. Das Artigo-Projekt**

<6>

Die Artigo-Anwendung ist als Spiel programmiert. Die Idee ist an ein nahezu identisches Verfahren angelehnt, welches erstmalig von Luis von Ahn umgesetzt wurde. Er entwickelte ein originelles Spiel (<http://www.gwap.com/gwap/gamesPreview/espgame/>), welches es ermöglicht, die bei der Online-Enzyklopädie *Wikipedia*<sup>13</sup> so erfolgreich praktizierte Einbeziehung von Laien in den Entstehungsprozess bei gleichzeitiger Missbrauchskontrolle auf den Bereich der Bildannotation zu übertragen.<sup>14</sup> Das Spiel ist so gestaltet, dass zwei beliebige Spieler anonym im Internet zusammengeschaltet werden, um jeweils Schlagworte

für Bilder zu verteilen. Die Bilder werden ihnen synchron vom Computer zugespielt und stammen aus dem im Netz frei zugänglichen Bilderfundus. Wie bei der Online-Enzyklopädie *Wikipedia* wird auch bei dem Spiel die Qualität der Eingaben kontrolliert.<sup>15</sup> Es wurde hier allerdings eine andere Methode gewählt: Das System speichert nur diejenigen Begriffe als valide Schlagworte ab, welche von mindestens zwei Spielern in identischer Form eingegeben werden. Dies schließt zwar eine falsche Verschlagwortung nicht aus, aber es unterbindet einen Missbrauch der Anwendung, da es ausgeschlossen sein dürfte, dass zwei Personen, beide willkürlich eine Christus-Darstellung mit dem Begriff ›Mickey Mouse‹ oder ›Bananen‹ belegen. Falsche Schlagworte, die dennoch von zwei Spielern eingegeben werden, sind möglicherweise intelligente Fehler, die durchaus von Relevanz sein können. So ist zum Beispiel denkbar, dass zwei Spieler irrtümlich einen Liebermann mit dem Begriff ›Slevogt‹ versehen.

<7>

Im Zuge des Artigo-Projekts wurde, wie oben erwähnt, eine mit dem esp-Game vergleichbare Anwendung entwickelt. Diese weicht jedoch sowohl in ihrer Programmierung als auch in ihrer Datengrundlage von Luis von Ahns Grundidee ab.<sup>16</sup> Die Datengrundlage bilden hierbei die Reproduktionen der Bilddatenbank Artemis der Ludwig-Maximilians-Universität. Das Spiel ist offen zugänglich. Eine Anmeldung ist notwendig, da die Spieler für übereinstimmende Annotationen (›matches‹) Punkte erhalten, welche festgehalten werden, um den Monatsbesten zu ermitteln.<sup>17</sup> Bei einer Anmeldung ist die Angabe jeglicher sozio-demographischer Merkmale freiwillig. Alternativ kann das Spiel aber auch ohne Anmeldung gespielt werden, was insbesondere dort von Vorteil ist, wo Interessierte sich erst einmal mit der Anwendung vertraut machen wollen. Die Spieldauer beträgt jeweils drei Minuten, was den Reiz des Spiels erhöhen soll und zu einer schnelleren Dateneingabe führt. Hierbei wird den Spielern jeweils eine Minute lang dasselbe Bild angezeigt. Die Eingabe selbst ist unkontrolliert. Die Spieler können, um ein Bild zu beschreiben, Begriffe in beliebiger Anzahl frei wählen. So können im erstellten Schlagwortkatalog sowohl Formaspekte als auch semantische Inhaltsaspekte auftauchen. Punkte gibt es für die Begriffe trotzdem nur, wenn der Spielpartner den identischen Begriff eingegeben hat, wobei hier keine zeitliche Koinzidenz notwendig ist. Selbst wenn ein Spieler gleich zu Beginn der Partie den Begriff ›Rubens‹ eingibt und sein Mitspieler diesen Begriff erst 40 Sekunden später wählt, wird dieser als identisch erkannt und als valides Schlagwort abgespeichert.

<8>

Das Artigo-Projekt richtet sich, wie auch das esp-Game, vordnehmlich an Laien. Die Teilnehmergruppe wird jedoch nicht in irgendeiner Form ausgewählt, sondern findet sich zufällig und freiwillig zusammen.<sup>18</sup> Grundsätzlich ist es denkbar, dass Laien viele triviale Begriffe produzieren, da ihnen die Fachsprache nicht bekannt ist (beispielsweise ›Rot‹, ›Blau‹, ›Mensch‹, ›Kirche‹). Diese trivialen Begriffe sind durchaus gewünscht, nur sollte der Begriffsfundus auch um nicht-triviale Begriffe erweitert werden. Dieses Problem wurde in einer früheren Version des Artigo-Spiels berücksichtigt: Einmal als valide Schlagworte abgespeicherte Begriffe wurden in roter Schrift unter dem Bild angezeigt. Sie wurden zu so genannten Tabuwörtern – gab ein Spieler eines der Tabuwörter ein, wurde dieses als ungültig gewertet und nicht erneut abgespeichert.<sup>19</sup> Gleichzeitig wurde der Spieler darüber informiert. Die Tabuwörter hatten den Zweck, das Spektrum der verwendeten Schlagworte zu erweitern, indem sie den Spieler dazu anhielten, über alternative Begriffe nachzudenken. Da für die aktuelle Artigo-Version wieder ein Spiel mit Tabuwörtern geplant ist, geht die Analyse der früheren Tabuwörterversion in die Überlegungen des vorliegenden Essays mit ein. Überdies ist das System der alten wie der neuen Version in der Lage, einen Spieler zu simulieren, um so auch bei geringer Anzahl von Spielern zu jedem Zeitpunkt funktionstüchtig zu sein. Insgesamt erzeugt das Programm einen Datensatz valider Schlagworte, der im Folgenden für die weiteren theoretischen und empirischen Betrachtungen als Grundlage dient.

### **3. Theoretische Überlegungen**

<9>

Dieses Kapitel diskutiert zunächst die Voraussetzungen eines qualitativ hochwertigen Datenoutputs auf Systemebene. Es soll hier zunächst geklärt werden, in welcher Umgebung die individuellen Tagging-Prozesse stattfinden müssen, um im Aggregat ein gutes Ergebnis zu produzieren (Kapital 3.1). Kapitel 3.2 widmet sich den einzelnen Akteuren. Die Teilnahmemotivation der Gruppenmitglieder wird hier nicht aufgegriffen. Es wird stattdessen das prozessproduzierte Vokabular der einzelnen Teilnehmer in den Vordergrund gerückt.

#### **3.1 Systemebene – »Weisheit der Vielen«**

<10>

In Tagging-Prozessen werden Schlagworte generiert. Erst in ihrer Summe bilden die Ergebnisse aller stattgefundenen Prozesse einen vollständigen Schlagwortkatalog. An der Entwicklung des Katalogs sind mehrere Menschen beteiligt, die, auch wenn sie nicht direkt kommunizieren können, zusammenwirken. Um zu eruieren wie sich das Ergebnis einer

solchen ›Teamarbeit‹ optimieren lässt, liegt es nahe zunächst einige Überlegungen bezüglich der kollektiven Intelligenz von Gruppen anzustellen.<sup>20</sup>

<11>

Erste Überlegungen zur kollektiven Intelligenz lassen sich bereits 1907 finden. Francis Galton, der Urvater der theoretischen Annahme der ›Weisheit der Vielen‹, stieß bei der Auswertung eines trivialen Schätzwettbewerbs auf das Phänomen. 1906 hatte auf der *West of England Fat Stock and Poultry Exhibition* ein Wettbewerb stattgefunden mit dem Ziel, das Gewicht eines Ochsen korrekt zu schätzen. Galton ging davon aus, dass viele der ca. 800 Teilnehmer über keine Vorkenntnisse auf diesem Gebiet verfügten, demnach also Laien waren. Diese Annahme bestätigten die einzelnen Schätzwerte, welche ein relativ breites Spektrum abbildeten. Erstaunlicherweise lag der aggregierte Durchschnittswert aller Teilnehmer nur um 500g neben dem tatsächlichen Gewicht des Schätzobjekts.<sup>21</sup> Dieses Ergebnis konnte vielfach in ähnlichen Experimenten bestätigt werden.<sup>22</sup> Es zeigt sich sogar öffentlich nachvollziehbar wöchentlich in der Quiz-Sendung *Wer wird Millionär*. Während der Telefonjoker, der als Expertenurteil interpretiert werden kann, in 65 % der Fälle eine richtige Antwort produziert, liegt die Quote des Publikumjokers um 26 % höher.<sup>23</sup> Diesen Ergebnissen ist zu entnehmen, dass zwar die jeweiligen Akteure oft unterschiedliche und unvollständige Informationen besitzen, dieses Wissen jedoch ausreicht, um in der Gruppe ein gutes oder sogar besseres Ergebnis zu erzielen als einzelne Experten. Das bedeutet, dass die oft fehlerhaften Einzelschätzungen im Durchschnitt der Masse neutralisiert werden.<sup>24</sup> Da das Wissen der Masse an keinen bestimmten Fach- oder Themenbereich gebunden ist, findet die kollektive Intelligenz in vielen Forschungsbereichen Einsatz. Beispielsweise werden nach diesem Prinzip auch Wahlergebnisse durch die Auswertung von individuellen Vermutungen in so genannten Wahlbörsen geschätzt.<sup>25</sup> Auch Tim O'Reilly greift in seiner Forderung nach Partizipationsmöglichkeiten für die Nutzer im Web auf den Ansatz der kollektiven Intelligenz zurück.<sup>26</sup> Web 2.0 Anwendungen, wie die Enzyklopädie *Wikipedia*, arbeiten auch tatsächlich mit der Annahme, dass die Masse an Informationen sich durch die Gruppe selbst kanalisiert. Die Daten der Artigo-Anwendung unterliegen denselben gedanklichen Kriterien. Auch hier ist zu vermuten, dass die heterogene Teilnehmergruppe in der Lage ist, individuelle Fehleinschätzungen zu revidieren und einen zusätzlichen Nutzen zu generieren. Bisher ist allerdings keine den Artigo-Ergebnissen entsprechende Datensammlung im wissenschaftlichen Kontext auf die positive Wirkung von kollektiver Intelligenz getestet worden. Das heißt, es ist nicht bekannt, ob im Falle der kunstgeschichtlich relevanten Bildannotation eine Gruppe von nicht professionellen Mitarbeitern intelligente bzw. sinnvolle Ergebnisse produzieren kann. Das Ziel dieser Studie

ist es, im Rahmen der Artigo-Evaluierung Klarheit über diese Frage zu schaffen. Hierzu müssen zunächst die Voraussetzungen betrachtet werden, die gegeben sein müssen, damit das Phänomen der kollektiven Intelligenz auftreten kann. James Surowiecki und Scott Page folgend, führt die Aggregation von Daten nicht bei jeder beliebigen Gruppe zwingend zu besseren Ergebnissen. Surowiecki beschreibt vier Kriterien, die eine Gruppe erfüllen muss, um ›weise‹ Ergebnisse zu erzielen. Erstens muss eine Diversität der Meinungen gegeben sein, zweitens müssen die Meinungen voneinander unabhängig sein, drittens muss die Gruppe die Einzelergebnisse in einem dezentralen Prozess produzieren und viertens müssen diese Einzelergebnisse nach Abschluss ihres Entstehungsprozesses von einem geeigneten Werkzeug gebündelt werden.

<12>

Die Gruppe der Artigo-Nutzer operiert völlig unabhängig voneinander, da jeglicher Austausch aufgrund fehlender Kommunikationsmöglichkeiten unterbunden wird – es kann daher sowohl die Dezentralität des Produktionsprozesses einzelner Ergebnisse als auch die Unabhängigkeit der Meinungen angenommen werden. Auch die Aggregation der Einzelergebnisse kann als gegeben betrachtet werden, da Social-Tagging-Anwendungen per definitionem einen kollektiven Datensatz erzeugen.<sup>27</sup> Das vierte Kriterium, die Diversität der Meinungen, ist von zentraler Bedeutung für das Auftreten kollektiver Intelligenz, da die Ähnlichkeit von Personen, so die theoretische Vermutung, zur Folge hat, dass auch die Ideen, die sie produzieren einander gleichen.<sup>28</sup> Ähnliches gilt für ungenügend ausgeführte Sachverhalte: Aus ein und derselben Perspektive werden mit nur geringer Wahrscheinlichkeit ergänzende Informationen hinzukommen. Auch Fehler können bei einer wissens- oder meinungshomogenen Gruppe nicht oder nur schlecht ausgeglichen werden, da sich die Ergebnisse tendenziell zu ähnlich sind. Scott Page zufolge sind es vor allem die während der Schul- und Ausbildungszeit gemachten Erfahrungen, sowie die dort erworbene Bildung, die Einfluss auf Perspektiven und Interpretationsbildung nehmen.<sup>29</sup> Außerdem hat die persönliche Identität einen starken Einfluss auf die Ausprägung der Perspektiven und Meinungen. Identität wird im weiteren Sinne nach Page durch den kulturellen Hintergrund einer Person, also ihren gesellschaftlich vermittelten Sinnhorizont bedingt. Diese wird überdies durch weitere Faktoren, wie zum Beispiel das Alter, das Geschlecht, die sexuelle Orientierung, die Religionszugehörigkeit, die physischen Fähigkeiten und Attribute oder die ethnische Zugehörigkeit beeinflusst.<sup>30</sup>

<13>

Diesen Überlegungen zur kollektiven Intelligenz ist zu entnehmen, dass eine Gruppe, die bezüglich der genannten Faktoren heterogen ist, in der Lage ist, relevantere Ergebnisse zu produzieren, als eine homogene Gruppe. In den nachfolgend dargestellten Untersuchungen werden insbesondere die Einflussvariablen Geschlecht, Alter und Bildungsabschluss betrachtet.

### **3.2 Akteursebene – Das hierarchische Netzwerkmodell von Quillian**

<14>

Das Wort bildet die Grundeinheit schriftlich basierter Tagging-Systeme. Das von den Nutzern verwendete Vokabular stellt demnach ein wichtiges Kriterium bei der Analyse solcher Systeme dar. Das individuelle Vokabular einer Person spiegelt ihr Verständnis der Realität und ihr Wissen wider; es leitet sich aus ihrem individuellen Sinnhorizont ab. Die besondere Beschaffenheit individueller Wortschätze soll allerdings nicht Gegenstand der folgenden Überlegungen sein, da diese nicht im Einzelnen, sondern nur im Aggregat von Relevanz für die Untersuchung sind. Stattdessen stehen in den nachfolgenden Überlegungen die Voraussetzungen für den Abruf von Wissen, welches sich auch im zur Verschlagwortung verwendeten Vokabular ausdrückt, im Mittelpunkt. Hierzu wird zunächst ein kognitionspsychologisches Modell vorgestellt, welches sich gezielt mit dem Abruf von Wissen auseinandersetzt: das hierarchische Netzwerkmodell von M. Ross Quillian.

<15>

Menschliches Wissen wird, laut Quillian, in hierarchisch angelegten Categoriesystemen im Gedächtnis gespeichert.<sup>31</sup> Hierbei unterscheidet er zwischen Ober- und Unterbegriffen, die durch sogenannte isa-Verbindungen<sup>32</sup> verknüpft sind. Durch die isa-Verbindungen kann so ein kognitives Netzwerk aus unterschiedlichen Wissenskategorien entstehen, wodurch das Abrufen ganzer Wissensstrukturen möglich wird. Einzelne Wissensbestandteile – etwa die erinnerten Eigenschaften eines Gegenstandes – sind bei der jeweils zutreffenden Kategorie abgespeichert und können bei Bedarf abgerufen werden. Wird eine Eigenschaft bzw. ein Wissensbestandteil bei einer Oberbegriff-Kategorie abgespeichert, so gilt dieses ebenso für alle darunter angeordneten Kategorien. Die Eigenschaft wird sozusagen nach unten hin vererbt. Allan Collins und M. Ross Quillian haben im Rahmen einer Studie das theoretische Konstrukt Quillians auf die Gedächtnisstruktur über die Eigenschaften von Tieren übertragen.<sup>33</sup>



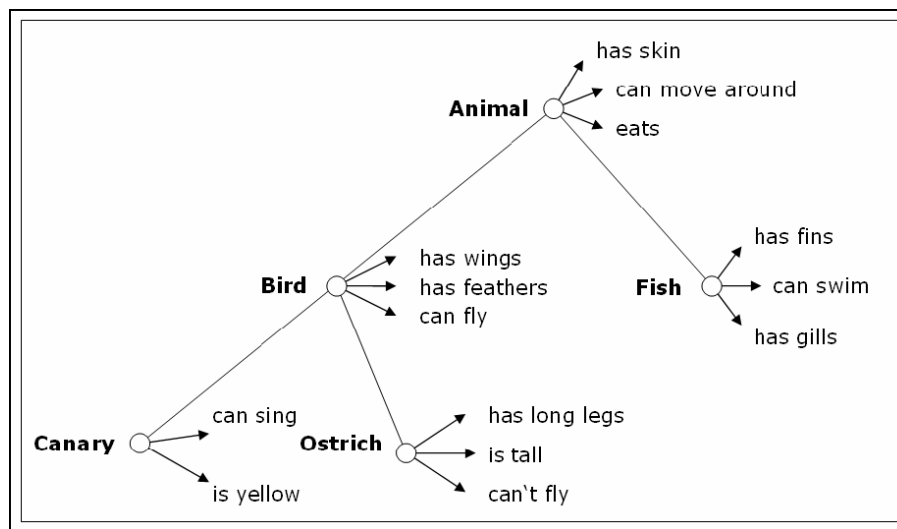


Abbildung 1: Eine hypothetische hierarchische Gedächtnisstruktur

Wie in Abbildung 1 zu erkennen ist, sind nach Collins und Quillian Eigenschaften, die für alle Tiere gelten, direkt in der Hierarchieebene ›Tiere‹ (Animals) abgespeichert. Für alle Unterkategorien folgt daraus, dass diese Eigenschaften hier ebenso gelten. Besondere Eigenschaften oder Ausnahmen sind immer direkt bei derjenigen Kategorieebene memoriert, die sie betreffen. Collins und Quillian baten nun Versuchspersonen im Rahmen eines Experiments, verschiedene Behauptungen zu Eigenschaften von Vögeln auf ihre Richtigkeit zu überprüfen. Sie gingen davon aus, dass die Versuchspersonen die Richtigkeit von Aussagen schneller identifizieren können, wenn die in den Aussagen verknüpften Informationen aus derselben Kategorie (Knoten) stammen, als wenn diese weit voneinander entfernt gespeichert sind. Das Experiment konnte diese Vermutung bestätigen.<sup>34</sup>

<16>

Carol Conrad konnte allerdings in einer weiteren Studie zeigen, dass die Abrufzeiten nicht entscheidend durch die hierarchische Struktur und die Entfernung der gespeicherten Informationen wie bei Quillians Netzwerkmodell beeinflusst werden, sondern vielmehr davon, wie oft eine Information alltäglich erlebt wird.<sup>35</sup> Denn je häufiger zwei Sachverhalte gemeinsam erlebt werden, desto stärker wird die Assoziation zwischen diesen. Collins und Loftus schließen daraus, dass gewisse Wissensbestandteile mehrfach abgespeichert werden können und zwar direkt bei dem jeweiligen Knotenpunkt (Kategorie), der stark mit dem Wissensbestandteil assoziiert wird (beispielsweise ›1968‹ und ›Woodstock‹).<sup>36</sup> Anderson fasst die Überlegungen und Experimente Conrads, Collins und Loftus, sowie Quillians anschaulich zusammen. Er kommt zu folgenden drei Thesen: Erstens werden Informationen

über ein Konzept – die häufig in dieser Kombination auftreten – direkt am Knoten des Konzepts gespeichert. Zweitens werden Information und Konzept umso stärker assoziiert, je häufiger sie in Kombination auftreten. Drittens haben Informationen, die nicht direkt mit einem Konzept an einem Knoten gemeinsam gespeichert sind, längere Abrufzeiten als solche, die direkt verknüpft sind.<sup>37</sup>

<17>

In einer Social Tagging-Anwendung muss der Nutzer permanent Wissen abrufen, um Schlagworte zu vergeben. Im Falle des Artigo-Spiels wird der Nutzer sogar dazu angehalten, eine größere Anzahl an Begriffen nacheinander zu demselben Objekt abzurufen. Den dargelegten Überlegungen folgend ist zu erwarten, dass die Spieler, von ihrem Anfangsbegriff ausgehend, zuerst andere verwandte bzw. häufig assoziierte Begriffe nennen. Für das Artigo-Projekt bedeutet das, dass die von Teilnehmern generierten Daten nicht nur von den in Kapitel 3.1 genannten sozio-demographischen Merkmalen abhängen, sondern ebenso durch das Vorgehen bei der Annotation beeinflusst werden können.

<18>

Ein bestimmtes Vorgehen ist in der Regel an einen Stimulus geknüpft, auf welchen ein Akteur reagiert. Stimuli stellen im Rahmen der Artigo-Anwendung sowohl die permanenten visuellen Impulse der zu annotierenden Abbildungen dar, als auch die, in der früheren Version vorhandenen, Tabuwörter, da diese zu weiteren Überlegungen anregen können. Anschließend an das hierarchische Netzwerkmodell von Quillian lässt sich ein weiteres kognitionspsychologisches Modell nennen, das Überlegungen zum Handlungsreiz einbezieht: Das assoziative Priming bezeichnet einen Vorgang bei dem einer bzw. eine Mehrzahl von Knoten, die mit gesuchtem Wissen oder Begriffen assoziiert werden, durch externen Einfluss aktiviert werden.<sup>38</sup> Durch diese externe Aktivierung sind die mit dem aktivierten Knoten verknüpften Verbindungen leichter abzurufen. Meyer und Schvaneveldt konnten diesen Effekt in einem Experiment nachweisen. Versuchsteilnehmer wurden hierbei gebeten zu entscheiden, ob bestimmte Wortpaare, die ihnen vorgelegt wurden, lexikalisch korrekt sind, wobei ihre Reaktionszeit gemessen wurde.<sup>39</sup> Es zeigte sich, dass Wortpaare, welche eine semantische Verknüpfung aufwiesen, deutlich schneller als korrekt identifiziert werden konnten als solche, denen eine semantische Verknüpfung fehlte.<sup>40</sup> Der Priming-Effekt erleichtert demnach den Abruf von Wissen durch die Aktivierung von assoziiertem Wissen. Außerdem kann die Aktivierung bestimmter Knoten auch den Zugriff zu sehr speziellem oder ungewöhnlichem Wissen ermöglichen, welches ohne eine Priming-

Aktivierung im Voraus nicht möglich gewesen wäre, wenn etwa die Assoziation für die Person nicht stark genug ausgeprägt ist. Das assoziative Priming kann also als eine Art Denkanstoß beschrieben werden.

<19>

Je nach individuellem Wissen zu einem in der Artigo-Anwendung ausgesandtem Impuls (Abbildung oder Tabuwort), wird der Stimulus unterschiedliche Assoziationen hervorrufen. Es ist daher zu erwarten, dass Experten und Laien in ihren Assoziationen bezüglich eines bestimmten Stimulus differieren. Zunächst scheint diese Aussage trivial, ein letztes kognitionspsychologisches Modell soll darlegen, warum gerade dieser Unterschied von Relevanz für die Frage nach einem hochwertigen Schlagwortkatalog ist: der Fächereffekt.

Der Fächereffekt bezeichnet nach Anderson zunächst die verlängerte Abrufzeit von Wissen, die aufgrund einer zunehmenden Zahl von verknüpften Informationen (bzw. Knoten) entsteht.<sup>41</sup> Die Abrufzeit der Informationen verlängert sich, wenn eine Vielzahl von Verbindungen zu anderen Knoten vorliegt, da gleichzeitig von einem Knoten ausgehend nur eine begrenzte Anzahl anderer Knoten aktiviert werden kann.<sup>42</sup> Das heißt, dass ein aktivierter Knoten, der normalerweise alle mit ihm verknüpften Knoten aktivieren würde, in seiner Potenz durch eine große Anzahl von Verknüpfungen gehemmt wird. Anderson konnte diesen Effekt 1974 und 1976 in Experimenten nachweisen.<sup>43</sup> Im Rahmen dieser Versuche wurden die Teilnehmer aufgefordert, eine Reihe von Fakten zu lernen. Diese waren entweder mit nur einer Assoziation verknüpft oder mit mehreren. Wie schon bei Schvaneveldts Versuchsaufbau zum assoziativen Priming wurde auch hier die Reaktionszeit gemessen, die die Probanden benötigten, um die gelernten Informationen abzurufen. Anderson konnte zeigen, dass die Reaktionszeit mit der Anzahl an verknüpften Assoziationen zunahm.<sup>44</sup> Das bedeutet jedoch nicht, dass grundsätzlich der Abruf von Informationen erschwert wird, wenn die Aktivierung von einem Knoten ausgeht, der selbst viele Verknüpfungen enthält. Vielmehr wird nur der Abruf der Informationen von bestimmten Knoten gehemmt.<sup>45</sup> Die Hemmung mancher Informationen ist aus einer rein logischen Perspektive leicht nachzuvollziehen, wenn beispielsweise die umfangreichen Mengen an Wissensbestandteilen von Experten betrachtet werden. Häufig werden Teilinformationen als offensichtlich betrachtet und daher nicht mehr näher thematisiert. Die Intensität der bewusst erlebten Assoziation sinkt damit, was zur Folge hat, dass dieses Teilwissen bei der Aktivierungskette übergangen wird. Eine Hemmung des Abrufs bestimmter Informationen ist allerdings nur dann gegeben, wenn das Wissen ausschließlich von einem Knoten aus aktiviert werden kann – also sonst keine weiteren eigenen Verknüpfungen aufweist. Ein

umfangreiches Wissen über ein Konzept kann demnach den Zugriff auf spezielle verknüpfte Informationen stören.

<20>

Das theoretische Konstrukt des Fächereffekts besagt also, dass Personen mit umfangreichem Wissen in ihrem Informationsabruf durch dieses behindert werden. Damit wird sowohl die These der kollektiven Intelligenz als auch die Einbeziehung von Laien gestützt. Die Überlegungen zum Fächereffekt ergänzen außerdem die in Kapitel 3.1 formulierten Annahmen zum Einfluss bestimmter sozio-demographischer Merkmale auf die Verschlagwortung. So wird im Rahmen dieses Essays nicht nur analysiert, ob eine möglichst heterogene Gruppenzusammensetzung bessere Ergebnisse liefert als eine homogene, sondern auch welche Rolle fachspezifisches Wissen bei der Bildannotation spielt.

#### **4. Studiendesign**

<21>

Wie in Kapitel 3 diskutiert, befasst sich der hier präsentierte Ausschnitt der Artigo-Evaluation mit dem Einfluss von Expertenwissen, sowie dem Einfluss der Gruppenzusammensetzung auf die Ergebnisse der Artigo-Anwendung. Außerdem werden weitere mögliche Einflüsse untersucht, die durch spielimmanente Faktoren, wie Tabuwörter und die Möglichkeit Punkte zu erzielen, hervorgerufen werden. Das Datenmaterial, welches der Analyse zugrunde liegt, wurde mittels drei unabhängiger Erhebungen, sowie der Auswertung der Artigo-Originaldaten, gewonnen.

<22>

Die Kernstudie dieser Evaluation stellt eine Simulation der Artigo-Anwendung dar, wobei zusätzlich zu der Vergabe von Schlagworten zu verschiedenen Abbildungen auch sozio-demographische Merkmale und Angaben zu fachspezifischem Wissen erhoben wurden. Die Datenerhebung fand mit Hilfe eines Online-Fragebogens statt, wobei die Teilnehmer anonym blieben. Diese wurden hierbei gebeten, vier Abbildungen jeweils sieben Begriffe zuzuordnen. Im Folgenden wird die Auswertung zu zwei stilistisch unabhängigen Abbildungen präsentiert – es handelt sich dabei um *Napoleon überquert die Alpen am großen St. Bernard* von Jacques Louis David und *Die großen Badenden* von Paul Cézanne.



Abbildung 2: Jacques Louis David, Napoleon überquert die Alpen am großen St. Bernard, 1801, Öl auf Leinwand, 260 x 221 cm, Musée National Château de Malmaison.



Abbildung 3: Paul Cézanne, Die Großen Badenden, 1906, Öl auf Leinwand, 208 x 249 cm, Philadelphia Museum of Art.

An der Artigo-Simulationsstudie nahmen 168 Personen teil, sodass für beide Bilder jeweils 1176 Tags generiert wurden, die in die Auswertung mit eingeflossen sind. Bei den beiden weiteren Erhebungen handelt es sich um Expertenbefragungen. Die erste Befragung mit Hilfe eines Fragebogens hatte zum Ziel, die Einschätzung von neun Experten zur Relevanz von häufig genannten Schlagworten zu einer Abbildung zu erfassen. Unter Experten werden hier Personen verstanden, die sich entweder in der Abschlussphase eines kunstgeschichtlichen Studiums befinden oder dieses bereits beendet haben. Die abgefragten Schlagworte wurden aus den Online-Fragebögen zur Artigo-Simulation gewonnen. Die zweite Expertenbefragung richtete sich speziell an Artigo-Spieler. In einem qualitativen Leitfadenterview wurden vier Spieler dazu befragt, ob und aus welchem Grund sie bestimmte Begriffe häufig verwenden und welche Rolle für sie Tabuwörter bei der Wahl ihrer Begriffe spielen. Die Angaben der Spieler beziehen sich hierbei auf die ältere Artigo-Version.

<23>

Aus den Daten der Artigo-Simulationsstudie und der Artigo-Originalanwendung konnten für die vorliegende Arbeit unterschiedliche quantitative Messvariablen gebildet werden, die sich für eine statistische Überprüfung der Leitthesen eignen. Eine Messvariable bezeichnet in diesem Zusammenhang das Merkmal einer Person oder eines Objekts, etwa das Alter. In dieser Studie werden lediglich deskriptive Statistiken ausgewertet, deren Ziel es ist, einfache empirische Daten durch Tabellen und Grafiken anschaulich darzustellen, sowie durch verschiedene Kenngrößen statistisch zu beschreiben. Mit deskriptiven statistischen Methoden kann beispielsweise zweidimensional der Zusammenhang zwischen zwei Messvariablen erfasst und über verschiedene Maßzahlen kenntlich gemacht werden. Solche so genannten Zusammenhangsmaße können anhand von Häufigkeits- und Mittelwertsvergleichen berechnet werden, wobei getestet wird, ob Unterschiede zwischen den einzelnen Variablen-Ausprägungen der untersuchten Träger bestehen. Ein hohes Durchschnittseinkommen von Männern und ein niedriges Durchschnittseinkommen von Frauen ließen demnach beispielsweise auf einen Zusammenhang zwischen dem Geschlecht und dem Einkommen schließen. Dieser müsste anschließend allerdings in weiteren fundierten Analysen erforscht werden. Durch einige statistische Maßzahlen, etwa den Phi- oder den Lambda-Koeffizienten, kann die Stärke eines Zusammenhangs gemessen werden. Überdies wird durch die Statistikprogramme für den jeweiligen Zusammenhang ein Signifikanzniveau ausgegeben, das anzeigt, mit welcher Wahrscheinlichkeit ein Zusammenhang nicht zufällig zustande gekommen ist, sondern tatsächlich vorliegt. Das Signifikanzniveau dient insbesondere dazu, von der gezogenen Stichprobe auf die bekannte Grundgesamtheit zu schließen. Da es sich bei den hier durchgeführten Studien jedoch nicht

um eine zufällige Stichprobenziehung aus einer zuvor definierten Grundgesamtheit handelt, kann dieser inferenzstatistische Aspekt vernachlässigt werden. Stattdessen dient das Signifikanzniveau, das sich aus der Zusammenhangsstärke in Abhängigkeit der untersuchten Teilnehmerzahl berechnen lässt, hier näherungsweise als rein beschreibendes Gültigkeitsmaß für einen Zusammenhang. Gerechnet wurde hierbei mit einem Signifikanzniveau von 0,01.<sup>46</sup>

## 5. Ergebnispräsentation

<24>

Zuerst soll im Rahmen des vorliegenden Essays besprochen werden, ob eine heterogene Gruppe in der Lage ist, bessere, d.h. im Falle eines Schlagwortindexes vielfältigere Ergebnisse zu produzieren. Anschließend wird die Begriffswahl von Laien und Experten hinsichtlich der Relevanz der vergebenen Schlagworte verglichen. Die in der Artigo-Simulationsstudie von den Probanden verwendeten Schlagworte wurden kategorisiert. Bei allen Auswertungsvorgängen wurde dabei auf drei verschiedene Kategoriensysteme Bezug genommen – eine thematische Zusammenfassung der Begriffsnennungen, eine nach dem Fokus geordnete Zusammenstellung der Begriffe (›Vorder-«, ›Mittel-«, ›Hintergrund‹) und ein Kategoriensystem, das zwischen Begriffen unterscheidet, die eine Interpretationsleistung voraussetzen und solchen die rein beschreibend sind.

### 5.1 Einflüsse von sozio-demographischen Merkmalen auf die Qualität der Artigo-Daten

<25>

Für das Geschlecht ergeben sich bei keiner der drei Kategorien signifikante Einflüsse. Weder für J.L. Davids *Napoleon überquert die Alpen am großen St. Bernard* noch für P. Cézannes *Die großen Badenden*. Die kumulierten relativen Häufigkeiten der Nennungen sind für alle Kategorien sehr ähnlich. Obwohl das Geschlecht für die thematische Wahl der Begriffe keine Rolle zu spielen scheint, zeigen sich dennoch Unterschiede bei der Begriffswahl von Frauen und Männern. Jeder Studienteilnehmer hat pro Abbildung sieben Begriffe vergeben. Insgesamt haben an der Studie 66,1 % Frauen und 33,9 % Männer teilgenommen; von Frauen wurden insgesamt 777 und von Männern 399 Begriffe vergeben, wobei einige Begriffe mehrfach auftraten. Für Davids *Napoleon überquert die Alpen am großen St. Bernard* wurden insgesamt 424 unterschiedliche Schlagworte vergeben. Hiervon wurden 218 nur von Frauen und 116 nur von Männern genannt, die restlichen 90 Schlagworte wurden sowohl von Männern als auch von Frauen zur Beschreibung der Bilder gebraucht, wie etwa der Begriff ›Napoleon‹ oder ›Reiter‹. Die Verwendung spezifischer Begriffe in Abhängigkeit vom Geschlecht wird im folgenden Venn-Diagramm (Abbildung 4) visualisiert.



Abbildung 4 – Die Verwendung einzelner Begriffe in Abhängigkeit des Geschlechts, Quelle: Eigene Erhebungen.

Auf den ersten Blick wirkt es, als hätten Frauen eine größere Anzahl an unterschiedlichen Begrifflichkeiten verwendet als Männer. Normiert man allerdings die Begriffszahlen auf die Gruppengröße, ergibt sich, dass Männer und Frauen im Durchschnitt ähnlich viele neue, also noch nicht genannte Begriffe nutzen. Für Paul Cézannes *Die großen Badenden* ergibt sich ein äquivalentes Ergebnis. Hier wurden insgesamt 475 verschiedene Begriffe vergeben. Frauen verwendeten davon 269 und Männer 122. Gemeinsam waren ihnen 42 Begriffe. Auch hier benutzten die Geschlechter im Durchschnitt ähnlich viele neue Begriffe (Frauen: 2,42 / Männer: 2,10).

<26>

Insgesamt lässt sich demnach zeigen, dass das Geschlecht zwar keinen Einfluss auf die thematische Begriffswahl, die besondere Beachtung einzelner Bildelemente oder die in den Begriffen enthaltene Interpretationsleistung hat, jedoch tragen sowohl Männer als auch Frauen mit ihrem spezifischen Wortschatz positiv zur Entstehung eines heterogenen Schlagwortkataloges bei.

<27>

Die Schulbildung hat nur auf die Interpretationsleistung, die den Begriffen zugrunde liegt, einen schwach signifikanten Einfluss: Personen mit niedriger Schulbildung nennen häufiger inhaltliche Begriffe und weniger äußere beschreibende Schlagworte als Personen mit höherer Bildung. Die Unterschiede in der Interpretationsleistung zeigen sich auch dann, wenn auf das Fachwissen kontrolliert wird, also nur Personen betrachtet werden, die zwar ein Abitur besitzen, jedoch keinen kunsthistorischen Hintergrund haben.<sup>47</sup> Dennoch kann



hier ein verzerrender Einfluss durch die Stichprobenziehung nicht ausgeschlossen werden: Es ist zu erwarten, dass vor allem solche Personen an dieser kunsthistorischen Studie teilgenommen haben, die sich auch für den Themenbereich interessieren und womöglich eine stärkere Affinität zur Kunst besitzen als der Durchschnitt. Ferner liegen auch nur sehr wenige Fälle mit niedriger Bildung in der gesamten Stichprobe vor – die große Mehrheit der Befragten besitzt Abitur – so dass kaum gültige Aussagen für alle Bildungsgruppen getroffen werden können. Die folgende Tabelle 4 gibt einen Überblick der Bildungsabschlüsse der Teilnehmer.

| Abschlussart** | Häufigkeit | Prozent | Kumulierte<br>Prozente |
|----------------|------------|---------|------------------------|
| Keiner         | 2          | 1,2%    | 1,2%                   |
| Realschule     | 10         | 6,0%    | 7,1%                   |
| Gymnasium      | 147        | 87,5%   | 94,6%                  |
| Sonstige       | 9          | 5,4%    | 100,0%                 |
| Gesamt*        | 168        | 100,0%  |                        |

\* Experten sind in dieser Auszählung noch nicht ausgeschlossen. Insgesamt haben 38 Personen mit kunsthistorischer Ausbildung an der Studie teilgenommen.

\*\* Die Stichprobe enthält keine Personen mit Hauptschulabschluss.

Tabelle 1 – Deskriptive Häufigkeiten der Schulabschlüsse, Quelle: Eigene Erhebungen.

Werden die beiden niedrigeren Schulbildungskategorien (›Keiner‹ und ›Realschule‹) in eine Kategorie zusammengefasst, so verschwinden die signifikanten Unterschiede. Wie beim Geschlecht zeigt sich überdies auch bei der Schulbildung kein Effekt auf die thematische Begriffswahl oder die besondere Beachtung einzelner Bildelemente. Auch bei Cézannes *Die Großen Badenden* lassen sich keinerlei signifikante Effekte der Schulbildung nachweisen. Aufgrund der geringen Varianz bei den Schulabschlüssen wird auf die Analyse des spezifischen, zur Verschlagwortung genutzten, Vokabulars verzichtet.

<28>

Insgesamt lässt sich für die Bildung als Einflussvariable auf die Datenheterogenität also keine eindeutige Aussage treffen. Es ist jedoch zu erwarten, dass sich bei einer Vergrößerung der Stichprobe über alle Bildungsgruppen Effekte zeigen. Hier könnten

weitere Studien zur Rolle der Bildung bei der Wahrnehmung von Bildern neue Erkenntnisse liefern.

<29>

Als letztes sozio-demographisches Merkmal wurde der Einfluss des Alters der Personen auf die Ergebnisheterogenität untersucht. Für die beiden Kategorien ›Interpretationsleistung‹ und ›Begriffsfokus‹ ergeben sich erneut keine Einflüsse für beide Gemälde. Hinsichtlich der thematischen Kategorie ergeben sich bei Davids *Napoleon überquert die Alpen am großen St. Bernard* allerdings signifikante Unterschiede bei den gruppierten Altersklassen. Die Altersklassen reichen von 0 bis 23 Jahre (Junge Personen), 24-40 Jahre (Personen im mittleren Alter) und 41-80 Jahre (Ältere Personen). Folgende Überlegungen sind für die Einteilung ausschlaggebend: Personen unter 24 Jahren sind meist noch in der Schule oder befinden sich am Anfang ihres Studiums und haben demzufolge vergleichsweise wenig Lebenserfahrung; Personen von 24 bis 40 Jahren durchleben die Endphase ihres Studiums und die Anfangsphase ihres Arbeitslebens; Personen über 40 Jahren sind in der Regel in der Arbeitswelt etabliert und verfügen über einen größeren Erfahrungsschatz als junge Personen.<sup>48</sup>

<30>

Das Alter hat bei der thematischen Begriffswahl einen leicht signifikanten Einfluss: ältere Personen wählten beispielsweise hinsichtlich des Gemäldes von David prozentual häufiger ein Schlagwort, das in variierenden Formen den Namen Napoleons enthielt, während jüngere Personen weit häufiger Begriffe aus den Kategorien ›Krieg/Schrecken‹ und ›Macht/Stärke‹ nannten. Betrachtet man wiederum die Verwendung einzelner Begriffe, so zeigt sich, wie schon beim Geschlecht, dass auch die unterschiedlichen Altersgruppen verschiedene Einzelbegriffe verwenden, um die beiden Gemälde zu beschreiben und damit zur Entstehung eines heterogenen Schlagwortkatalogs beitragen.

<31>

Insgesamt zeigten sich bei keinem der drei untersuchten Merkmalen (Geschlecht, Bildung, Alter) große Unterschiede in der thematischen Begriffswahl, der besonderen Beachtung einzelner Bildelemente oder der in den Begriffen enthaltenen Interpretationsleistung. Es kann aufgrund der Betrachtung der Einzelbegriffe allerdings vermutet werden, dass, wie durch die Annahmen zur Weisheit der Vielen erwartet, geschlechts- und altersheterogene Gruppen vielfältigere Begriffe verwenden als homogene Gruppen und damit besser geeignet sind, um umfassende Schlagwortkataloge zu produzieren.

## 5.2 Einflüsse von Experten- und Laienwissen auf die Qualität der Artigo-Daten

<32>

Ergänzend zu den vorangegangenen Ergebnissen soll der Einfluss von fachspezifischem Wissen auf die Ergebnisheterogenität der Artigo-Daten bzw. deren Simulation diskutiert werden. Es wurden für die Analyse der Daten zwei verschiedene Expertenmaße verwendet: Das erste Maß stuft Personen als Experten ein, die ein kunstgeschichtliches Studium absolvieren oder absolviert haben. Die Studienteilnehmer werden in die Kategorien ›Kunstgeschichtsstudium‹, ›Studium‹ und ›Kein Studium‹ unterteilt. Das zweite Expertenmaß stellt einen Index aus den Angaben der Studienteilnehmer zur eigenen Kompetenz und Beschäftigung mit fachbezogenen Aktivitäten dar. Die eigenen Angaben wurden auf Skalen von 1 (nie/schwach) bis 4 (oft/stark) gemessen und die individuellen Werte addiert. Diese Summe als Abbild der Fachkenntnis wurde um einen zusätzlichen Faktor gewichtet: Handelte es sich bei der betroffenen Person um einen Kunsthistoriker nach Definition des ersten Expertenmaßes, so wurde sein Summenwert mit der Zahl 1,5 multipliziert. Die Skala des sich hieraus ergebenden Gesamtindex reicht von 3 bis 21, wobei 21 die höchste Ausprägung der Fachkenntnis darstellt.

<33>

Vergleicht man die Schlagworte von Personen mit unterschiedlicher Fachkenntnis, so zeigen sich eindeutige Untersuchungsergebnisse. Zunächst unterscheidet sich die thematische Begriffswahl von Experten und Laien – für beide Expertenmaße sind die Unterschiede hoch signifikant. Die nachfolgende Abbildung zeigt an sechs Kategorien zu Davids *Napoleon überquert die Alpen am großen St. Bernard* exemplarisch, wo sich eindeutige Unterschiede ergeben haben. Gerade die Nennung des korrekten Künstlers ist hoch mit dem Vorhandensein von Fachwissen korreliert. Eine falsche Datierung wurde für dieses Bild hingegen ausschließlich von Laien angegeben, wobei Personen mit nicht-kunsthistorischem Hochschulhintergrund seltener eine falsche Datierung genannt haben. Erstaunlicherweise zeigen sich auch große Unterschiede bei der Farbbeschreibung des Bildes: Während die Kategorie mit Hinweisen auf rote Farbgestaltung häufiger von Fachleuten genannt wurde, wurde im Gegensatz dazu die Kategorie mit Hinweisen auf eine dunkle Farbgestaltung verstärkt von Personen ohne fachspezifisches Wissen genannt. Auch bei den Kategorien ›Macht/Stärke‹ und ›Krieg/Schrecken‹ ist der Zusammenhang offensichtlich. Diese beiden Kategorien enthalten hauptsächlich Begriffe, die nicht tatsächlich abgebildet sind, sondern eine Interpretationsleistung erfordern. Dies legt den Schluss nahe, dass sich Laien bei der Bildannotation stärker mit den Eindrücken, die eine Abbildung vermittelt, beschäftigen als Experten. Hinsichtlich des zweiten Expertenmaßes zeigen sich äquivalente Ergebnisse.

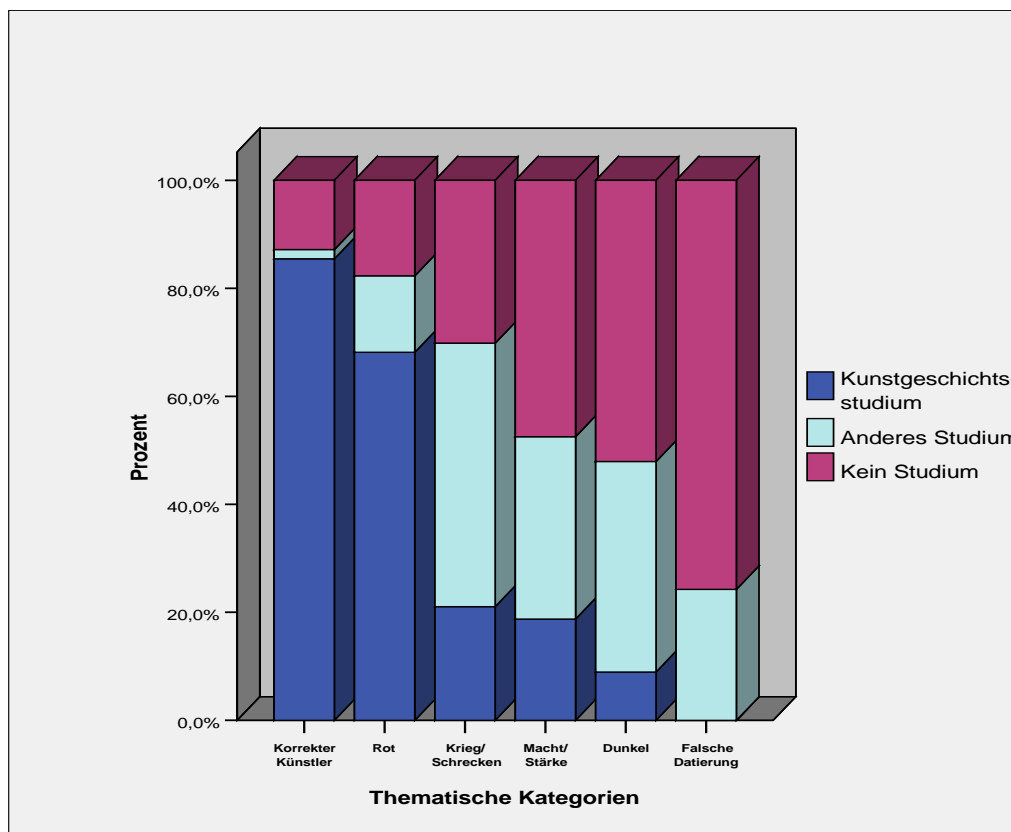


Abbildung 5 – Prozentuale Nennungshäufigkeit bestimmter Begriffe in Abhängigkeit des fachspezifischen Wissens (Expertenmaß 1), Quelle: Eigene Erhebungen.

<34>

Insgesamt zeigt sich in den statistischen Analysen also, dass für einige thematische Kategorien zwar große Unterschiede zwischen Laien und Experten existieren, dies jedoch nicht auf alle Kategorien zutrifft. Die Vermutung, dass Laien sich stärker mit den Eindrücken, die eine Abbildung vermittelt, beschäftigen, bestätigt sich auch in den Messungen zum Gebrauch von Schlagworten mit erforderlicher Interpretationsleistung (Tabelle 2), da Begriffe, die Eindrücke widerspiegeln, Interpretationsleistung erfordern. Wie in der folgenden Tabelle zu sehen ist, ist die prozentuale Verwendung von Begriffen, die eine Interpretationsleistung erfordern, bei Laien größer als bei Personen mit einem Kunstgeschichtsstudium. Die Ergebnisse sind allerdings nicht signifikant, so dass aus der Häufigkeitsverteilung lediglich Tendenzen abgelesen werden können. Es zeigt sich auch, dass tendenziell von Kunsthistorikern stärker auf tatsächlich abgebildete Elemente eingegangen wird, die deskriptiv beschrieben werden. Es zeigt sich ferner in den prozentualen Häufigkeiten, dass Experten öfter fehlerhafte Aussagen hinsichtlich Künstler, Datierung und Bildtitel treffen. Anderes zeigt sich hingegen am Beispiel von Cézannes *Die*

*großen Badenden*, wie in der nächsten Tabelle (Tabelle 3) zu sehen ist. Hier vergeben Experten häufiger Begriffe, die eine Interpretationsleistung erfordern. Die meisten Begriffe, die sich auf tatsächlich Abgebildetes beziehen, werden hier von Studenten bzw. Absolventen anderer Fachbereiche vergeben. Die meisten Fehler machen Personen ohne Hochschulstudium, dicht gefolgt von Experten. Diese Unterschiede sind leicht signifikant, doch lassen sich hier keine Richtungszusammenhänge feststellen, so dass die Differenzen zwischen den Gruppen nicht einwandfrei interpretierbar sind.

| Interpretationsleistung      |        | Expertenmaß 1                    |                    |                 | Gesamt |
|------------------------------|--------|----------------------------------|--------------------|-----------------|--------|
|                              |        | Kunst-<br>geschichts-<br>studium | Anderes<br>Studium | Kein<br>Studium |        |
| <b>Ja</b>                    | Anzahl | 145                              | 286                | 249             | 680    |
|                              | in %   | 54,5%                            | 58,4%              | 59,3%           | 57,8%  |
| <b>Nein</b>                  | Anzahl | 100                              | 171                | 146             | 417    |
|                              | in %   | 37,6%                            | 34,9%              | 34,8%           | 35,5%  |
| <b>Sonstige<br/>(Fehler)</b> | Anzahl | 21                               | 33                 | 25              | 79     |
|                              | in %   | 7,9%                             | 6,7%               | 6,0%            | 6,7%   |
| <b>Gesamt</b>                | Anzahl | 266                              | 490                | 420             | 1176   |
|                              | in %   | 100,0%                           | 100,0%             | 100,0%          | 100,0% |

Tabelle 2 – Die Interpretationsleistung des Schlagwortes in Abhängigkeit des Fachwissens: Expertenmaß 1, Davids ›Napoleon überquert die Alpen am Großen Sankt Bernhard‹, Quelle: Eigene Erhebungen.

| Interpretationsleistung      |        | Expertenmaß 1                    |                    |              | Gesamt |
|------------------------------|--------|----------------------------------|--------------------|--------------|--------|
|                              |        | Kunst-<br>geschichts-<br>studium | anderes<br>Studium | Kein Studium |        |
| <b>Ja</b>                    | Anzahl | 163                              | 291                | 240          | 694    |
|                              | in %   | 61,3%                            | 59,4%              | 57,1%        | 59,0%  |
| <b>Nein</b>                  | Anzahl | 85                               | 182                | 149          | 416    |
|                              | in %   | 32,0%                            | 37,1%              | 35,5%        | 35,4%  |
| <b>Sonstige<br/>(Fehler)</b> | Anzahl | 18                               | 17                 | 31           | 66     |
|                              | in %   | 6,8%                             | 3,5%               | 7,4%         | 5,6%   |
| <b>Gesamt</b>                | Anzahl | 266                              | 490                | 420          | 1176   |
|                              | in %   | 100,0%                           | 100,0%             | 100,0%       | 100,0% |

Tabelle 3 – Die Interpretationsleistung des Schlagwortes in Abhängigkeit des Fachwissens:  
Expertenmaß 1, Cézannes ›Die Großen Badenden‹, Quelle: Eigene Erhebungen.

<35>

Verwendet man das zweite Expertenmaß zur Analyse von Unterschieden in der Interpretationsleistung, so zeigen sich für beide Gemälde keine signifikanten Unterschiede. Insgesamt kann daher hier von nicht näher erklärbaren Tendenzen zur Verwendung von Begriffen mit oder ohne inhaltliche Interpretationsleistung bei Laien und Experten gesprochen werden, wobei an dieser Stelle nochmals darauf verwiesen sei, dass sich bei den beiden Gemälden gegensätzliche Ergebnisse zeigen. Hier könnte beispielsweise der Bekanntheitsgrad von Davids Gemälde verzerrend gewirkt haben. Die Ergebnisse könnten allerdings auch darauf hinweisen, dass die Wahl der Schlagworte grundsätzlich von der vorliegenden Abbildung abhängt. In diesem Fall scheint es besonders ratsam, nicht nur Gruppen mit homogenem Fachwissen zur Verschlagwortung heranzuziehen, sondern wiederum hinsichtlich ihrer sozio-demographischen Merkmale heterogene Gruppen, die in der Lage sind, vielfältigere Ergebnisse mit und ohne Interpretationsleistung zu erzeugen.

<36>

Hinsichtlich des Kategoriensystems, das die Foci der Begriffe bewertet, konnten bei beiden Beispielen und Expertenmaßen keine nennenswerten signifikanten Unterschiede festgestellt werden.

<37>

In Bezug auf die Kategoriensysteme kann also gesagt werden, dass sich nur für die thematische Begriffswahl Einflüsse des Expertenwissens nachweisen lassen. Die Ergebnisse der beiden Beispiele variieren allerdings, so dass angenommen werden muss, dass weitere Faktoren auf das Annotationsverhalten wirken. Als Ergänzung zur Analyse der thematischen Begriffswahl wurden ebenfalls die Einzelbegriffe von Experten und Laien untersucht (Abbildung 6). Tatsächlich zeigt sich, dass bei dem Gemälde von David von den insgesamt 424 ergebnen Begriffen den beiden Gruppen nur 66 gemein sind. Experten haben hierbei insgesamt 64 verschiedene Begriffe vergeben und Laien 293. Bei dieser Betrachtung wurde ausschließlich das Expertenmaß 1 verwendet. Da die Stichprobe mehr Laien (130 Personen) als Experten (38 Personen) enthält, ergibt sich eine durchschnittliche Verwendung neuer Begriffe von 1,68 für Experten und 2,5 für Laien. Es scheint, als würden sich die Schlagworte von Personen mit fachspezifischem Wissen häufiger wiederholen als die von Laien, was womöglich auf die ähnliche Vorgehensweise bei der wissenschaftlichen Betrachtung eines Bildobjektes zurückgeführt werden kann.



Abbildung 6 – Die Verwendung einzelner Begriffe in Abhängigkeit des Fachwissens, Quelle: Eigene Erhebungen.

Für Cézannes *Die großen Badenden* ergeben sich ähnliche Ergebnisse. Von insgesamt 475 verschiedenen Begriffen haben Laien und Experten hier 55 gemein. Damit haben die Experten insgesamt 72 verschiedene Wörter verwendet und die Laien 348. Normiert auf die Anzahl der Personen ergibt das einen Zuwachs von 2,68 noch nicht genannten Begriffen pro Person bei den Laien und von 1,89 neuen Begriffen pro Person bei den Experten. Es zeigt sich demnach auch hier, dass die Laien im Schnitt auf ein breiteres Vokabular zurückgreifen, um das Bild zu beschreiben, während die Experten weniger Begriffe verwenden. Auch hier wird deutlich, dass eine heterogene Auswahl von Personen bei der Verschlagwortung für die Ergebnisvielfalt eine wichtige Rolle spielt, da sich die verwendeten Begriffe sowohl zwischen als auch innerhalb der Gruppen stark unterscheiden, wobei sich insbesondere die Gruppe der Laien durch die Verwendung einer größeren Begriffsvielfalt auszeichnet.

### **5.3 Einflüsse spielimmanenter Faktoren auf die Qualität der Artigo-Daten**

<38>

Da Tabuwörter, wie es in der älteren Artigo-Version der Fall war, die wiederholte Eingabe von bereits vorhandenen Begriffen unterbinden, kann davon ausgegangen werden, dass sie zu einer größeren Begriffsvielfalt führen und damit einen positiven Einfluss ausüben. Es ist allerdings zu klären, ob die Qualität der hierdurch provozierten heterogenen Ergebnisse ausreichend hoch ist, um davon ausgehen zu können, dass die Verwendung der Tabuwörter einen Zugewinn darstellt. Zur Beantwortung dieser Kernfrage wurden Leitfaden-Interviews mit vier Artigo-Spielern geführt. Ihre Angaben wurden anschließend mit den prozessproduzierten Ergebnissen der Artigo-Anwendung verglichen. Um die Vergleichbarkeit der Aussage mit den restlichen Ergebnissen dieser Arbeit zu gewährleisten, wurden als Beispiele dieselben Gemälde gewählt wie in der Daten-Simulation. Trotz der geringen Zahl der Befragten kann bei Übereinstimmung der Ergebnisse davon ausgegangen werden, dass andere Spieler ähnlich empfinden.

<39>

Zunächst werden die positiven Aspekte der Tabuwörter betrachtet. Alle vier Spieler hatten den Eindruck, dass Tabuwörter sie generell dazu anregen, hochwertigere Begriffe einzugeben (z.B. ›Monet‹ – ›Impressionismus‹ oder ›Seurat‹ – ›Pointilismus‹). Allerdings gaben auch zwei der Spieler explizit an, dass sie die Tabuwörter nicht ganz lesen. Das heißt, es kann davon ausgegangen werden, dass nicht die umfassende Gesamtheit der Tabuwörter die entscheidende Rolle für die Eingabe von hochwertigeren Begriffen spielt, sondern dass eher vereinzelte Wörter einen Einfluss ausüben. Ferner kann aus den



Aussagen der Interviewpartner geschlossen werden, dass die Tabuwörter dazu anregen, nach Synonymen für die bereits gematchten Begriffe zu suchen (›düster‹ – ›bedrohlich‹).

<40>

Negative Einflüsse der Tabuwörter auf die Qualität der Ergebnisse können zum Beispiel entstehen, wenn diese zu fehlerhaften Eingaben oder wenig sinnvollen Details führen. Alle vier Befragten gaben an, dass sie häufig Begriffe verwenden, bei denen sie nicht mit Sicherheit sagen können, ob diese tatsächlich zutreffen, besonders im Hinblick auf Ortsbezeichnungen, Epochen und Künstler. Dieses Problem tritt jedoch auch unabhängig von den Tabuwörtern auf.

<41>

Anhand der Originaldaten des Artigo-Spiels lassen sich die Aussagen der Spieler nicht eindeutig nachweisen. Tendenziell zeigt sich jedoch, dass mit steigender Tabuwörterzahl auch die Anzahl an weniger wertvollen Detailbeschreibungen steigt, wobei angemerkt werden muss, dass nicht jede Detailbeschreibung sinnlos ist. Gemeint ist hier, wie auch von einem der Spieler beschrieben, die Verwendung von Begriffen, die für den Bildinhalt offensichtlich irrelevant sind, da sie kein valides Unterscheidungsmerkmal darstellen, wie etwa die Bezeichnungen ›Nase‹, ›Augenbrauen‹ oder ›Mund‹ bei einem Porträt. Die Spieler versuchen demnach Punkte mit Begriffen zu erzielen, die wenig Aussagekraft besitzen, anstatt zum nächsten Bild weiter zu schalten. Dies kann einen negativen Einfluss auf die Qualität der Begriffssammlung zur Folge haben, da es zur Verschlagwortung einer Vielzahl unnötiger Begriffe kommt, was das eigentliche Ziel der Artigo-Anwendung verfehlt.

Andererseits kann auch davon ausgegangen werden, dass durch Tabuwörter hochwertige Begriffe generiert werden, die wiederum die Qualität des Schlagwortkataloges anheben. Die hochwertigen Begriffe fallen hier stärker ins Gewicht als die weniger hochwertigen Details, da die unnötig abgespeicherten Schlagworte die Suchfunktion nicht stören, sofern sie keine falschen inhaltlich-sachlichen Informationen enthalten. Ein solcher Zusammenhang zwischen falschen Begriffen und Tabuwörtern wurde in den Interviews nicht thematisiert.

<42>

Insgesamt kann davon ausgegangen werden, dass die Tabuwörter einen positiven Einfluss auf die Qualität des Schlagwortkatalogs ausüben. Dieser Einfluss ist allerdings nicht sonderlich stark, da die Tabuwörter in der Regel nicht vollständig von Spielern gelesen werden und sie außerdem den unerwünschten Nebeneffekt haben, überflüssigen ›Daten-Müll‹ zu produzieren. Es zeigt sich, dass es in der Artigo-Datensimulation, in der keine

Punkte vergeben werden, nicht mit steigender Anzahl von Tagging-Prozessen auch zu einer steigenden Anzahl an Detailbeschreibungen kommt. Daraus lässt sich schließen, dass die Verwendung wenig sinnvoller Details (beispielsweise detailgenaue Menschenbeschreibungen oder sehr kleine Farbflächen) in hohem Maße mit der Artigo-Anwendung korreliert. Ferner wird durch die Systematik der Tabuwörter die Erfassung der Häufigkeiten der Schlagworte verhindert.

<43>

Als weiterer spielimmanenter Faktor, der Einfluss auf den Datenoutput nehmen kann, wurde in den Interviews auf den Spielcharakter der Artigo-Anwendung eingegangen. Alle Befragten gaben an, dass sie bestimmte Worte mehrfach wieder verwendeten, um Punkte zu erzielen. Beispielhafte Begriffe, die häufig benutzt wurden, stellen vor allem Farben oder Hintergrundbeschreibungen dar – wie etwa die Begriffe ›Himmel‹, ›Wolke‹ und ›Landschaft‹. Überdies werden Menschen sehr detailliert beschrieben, mit Worten wie ›Augen‹, ›Nase‹ und ›Mund‹. Auch Begriffe, welche die Art der Abbildung beschreiben, wurden häufig verwendet, zum Beispiel ›Fotografie‹ oder ›Gemälde‹.

<44>

Für beide Bilder (Cézanne und David) lassen sich im Datensatz die von den Spielern benannten wiederholt auftretenden Begriffe nachweisen. Vor allem die Daten zu Davids Gemälde zeigen dies deutlich.<sup>49</sup> Betrachtet man zusätzlich andere Abbildungen des Datensatzes, bestätigt sich dies erneut. Die von den Spielern genannten Begriffe sind überdies beinahe deckungsgleich mit den insgesamt am häufigsten vergebenen Worten. Es kann demnach davon ausgegangen werden, dass sich für die Artigo-Anwendung eine Art Gruppenwortschatz herausbildet, auch wenn keine direkte Kommunikation zwischen den Teilnehmern stattfindet. Dieser scheint sich jedoch nicht deswegen entwickelt zu haben, weil bestimmte Begrifflichkeiten besonders sinnvoll erscheinen, sondern da die Spieler erfahrungsgemäß für diese Punkte erhalten.

<45>

Insgesamt lässt sich im Hinblick auf den Spielcharakter der Artigo-Anwendung schließen, dass die Möglichkeit Punkte zu erzielen einen negativen Einfluss auf die Gesamtqualität des Schlagwortkatalogs haben kann. Dies ist zu erwarten, da die Spieler aufgrund des Punktesystems einen starken Anreiz erhalten, Begriffe zu wählen, die wahrscheinlich gematched werden. Nichtsdestotrotz stellt der Spielcharakter einen wesentlichen Anreiz zur Projektteilnahme dar.

## 6. Fazit und Ausblick

<46>

In dieser Untersuchung konnte gezeigt werden, dass die Artigo-Anwendung in der Lage ist, Daten zu produzieren, die genauso hochwertig sind wie professionell, also von Experten erzeugte Daten. Allerdings gilt es einige Besonderheiten zu beachten: Zum einen ist es zur Verschlagwortung wichtig, eine hinsichtlich ihrer sozio-demographischen Merkmale und ihrem fachspezifischem Wissen möglichst heterogene, große Teilnehmergruppe heranzuziehen. Dadurch kann es gelingen, einen umfassenden und qualitativ hochwertigen Schlagwortkatalog zu erstellen.

<47>

Im Rahmen der Untersuchungen wurden jedoch auch einige nachteilige Aspekte der Artigo-Anwendung festgestellt. Die Spieler werden zum Beispiel aufgrund des Spiel-Charakters der Anwendung dazu animiert, Falscheingaben zu tätigen, um Punkte zu erhalten. Die in der Anwendung implementierten Tabuwörter bewirken des Weiteren, wie gezeigt werden konnte, in Kombination mit der ›Punktejagd‹ der Spieler, einen Anstieg in der Verwendung wenig sinnvoller Detailbeschreibungen. Demnach muss eingeschränkt werden, dass der entstehende Schlagwortkatalog zwar genauso gute Daten produzieren kann wie ein professioneller Erzeuger, jedoch gleichzeitig auch ›Daten-Müll‹ mitproduziert. Während dies bei wenig sinnvollen Details keine Rolle spielt, da diese Schlagworte zwar abgespeichert werden, aber keine weiteren Störungen hervorrufen, stellt die Abspeicherung von falschen Informationen ein gravierendes Problem dar. Stellt ein Nutzer eine Suchanfrage, kann er aufgrund eines falschen Schlagwortes eine inhaltlich fehlerhafte Antwort erhalten. Die Qualität der Daten sinkt damit merklich ab. Dieses Problem könnte beispielsweise umgangen werden, wenn der hier verwendete Typ der Set-Model-Folksonomie (Schlagwortkatalog), der die Nennungshäufigkeiten nicht speichert, durch eine Bag-Model-Folksonomie ersetzt wird, welche Häufigkeiten unterstützt.<sup>50</sup> Das bedeutet, dass anhand der abgebildeten Nennungshäufigkeit die Relevanz (Kapitel 5) der diversen Schlagworte für die gesuchte Abbildung identifiziert werden könnte. Wie gezeigt wurde, werden gerade Begriffe von großer Relevanz besonders häufig genannt und Begriffe mit niedriger Relevanz eher selten. Fehlerhafte Eingaben, da sie wenig Relevanz besitzen, würden, dieser Annahme folgend, seltener verwendet und könnten dadurch identifiziert werden (etwa falsche Künstlernamen oder Ortsbezeichnungen). Auch die umfangreich vorhandenen und wenig sinnvollen Detailbeschreibungen können kontextabhängig zu Problemen führen, beispielsweise dann, wenn tatsächlich nach einer Detaildarstellung, etwa von einem Auge gesucht wird (etwa

René Magritte, *Der falsche Spiegel*, 1935). Hier könnte wiederum die Erfassung der Nennungshäufigkeiten eine Lösung darstellen.

<48>

Zusammenfassend kann aus den Überlegungen dieses Essays geschlossen werden, dass es möglich und sinnvoll ist, eine Social Tagging-Anwendung im Internet zu nutzen, um kunsthistorisches Wissen zu sammeln und hieraus einen Schlagwortkatalog zu erstellen. Das Artigo-Projekt stellt hierbei einen ersten Schritt dar, um das vorhandene kunsthistorische Wissen im Bereich der Bildannotation zu bündeln und zugänglich zu machen.

## **Bildnachweis**

Abbildung 1: Eine hypothetische hierarchische Gedächtnisstruktur, Quelle: Held, Christoph/ Cress, Ulrike: Social Tagging aus kognitionspsychologischer Sicht, in: Gaiser, Birgit/ Hampel, Thorsten/ Panke, Stefanie (Hrsg.): Good Tags – Bad Tags. Social Tagging in der Wissensorganisation, Münster 2008, S. 37-49, S. 41; nach Collins, Allan M./ Quillian, M. Ross: Retrieval Time from Semantic Memory, in: Journal of Verbal Learning and Verbal Behavior, 1969/8, S. 240-247.

Abbildung 2: Jaques Louis David, Napoleon überquert die Alpen am großen St. Bernard, 1801, Öl auf Leinwand, 260 x 221 cm, Musée National Château de Malmaison, Quelle: Artemis – Bilddatenbank des Instituts für Kunstgeschichte der Ludwig-Maximilians-Universität München.

Abbildung 3: Paul Cézanne, Die Großen Badenden, 1906, Öl auf Leinwand, 208 x 249 cm, Philadelphia Museum of Art, Quelle: Artemis – Bilddatenbank des Instituts für Kunstgeschichte der Ludwig-Maximilians-Universität München.

Abbildung 4: Die Verwendung einzelner Begriffe in Abhängigkeit des Geschlechts, Quelle: Eigene Erhebungen.

Abbildung 5: Prozentuale Nennungshäufigkeit bestimmter Begriffe in Abhängigkeit des fachspezifischen Wissens (Expertenmaß 1), Quelle: Eigene Erhebungen.

Abbildung 6: Die Verwendung einzelner Begriffe in Abhängigkeit des Fachwissens, Quelle: Eigene Erhebungen.

Tabelle 1: Deskriptive Häufigkeiten der Schulabschlüsse, Quelle: Eigene Erhebungen.

Tabelle 2: Die Interpretationsleistung des Schlagwortes in Abhängigkeit des Fachwissens: Expertenmaß 1, Davids *Napoleon überquert die Alpen am Großen Sankt Bernard*, Quelle: Eigene Erhebungen.

Tabelle 3: Die Interpretationsleistung des Schlagwortes in Abhängigkeit des Fachwissens: Expertenmaß 1, Cézannes *Die Großen Badenden*, Quelle: Eigene Erhebungen.

- 
- 1 Für eine kurze Beschreibung relevanter Literaturdatenbanken im Bereich Kunstgeschichte siehe Kohle, Hubertus/ Kwastek, Katja: *Computer, Kunst und Kunstgeschichte*, Köln 2003, S. 23-26.
  - 2 Röttgers, Arndt: *Die Dienste des Internet*, in: Kohle, Hubertus (Hg.): *Kunstgeschichte digital*, Berlin 1997, S. 160-166, hier S. 165f.
  - 3 Vgl. Weinberger, David: *Das Ende der Schublade. Die Macht der neuen digitalen Unordnung*, München 2008.
  - 4 Mandl, Thomas: *Linkanalyse und alternative Verfahren zur Qualitätsverwertung im Web* *Information Retrieval* 12/2005, Online abrufbar unter: <http://www.datenbank-spektrum.de/pdf/dbs-12-16.pdf> (zuletzt abgerufen am 27.08.2010)
  - 5 Battelle, John: *The Search – how Google and its rivals rewrote the rules of business and transformed our culture*, New York 2005, S. 19.
  - 6 Vgl. Müller-Wienbergen, Felix: *Spezialisiertes Suchen im Web*, in: Grob, Heinz Lothar/ Vossen, Gottfried (Hrsg.): *Entwicklungen im Web 2.0 aus technischer, ökonomischer und sozialer Sicht*, Institut für Wirtschaftsinformatik der Westfälischen Wilhelms-Universität Münster 2007, S. 141-152, hier S. 142ff.
  - 7 Herwig, Sebastian: *Universelles Suchen im Web – Eine technische, ökonomische und soziale Betrachtung*, in: Grob, Heinz Lothar/ Vossen, Gottfried (Hrsg.): *Entwicklungen im Web 2.0 aus technischer, ökonomischer und sozialer Sicht*, Institut für Wirtschaftsinformatik der Westfälischen Wilhelms-Universität Münster 2007, S.129-140, S.129, vgl. auch Perez, Juan Carlos: *What is vertical search?* 18.01.2006, Online abrufbar unter: <http://www.infoworld.com/t/platforms/what-vertical-search-057>.
  - 8 Für eine knappe Begründung siehe Kohle/ Kwastek 2003, S. 51-55.

- 9 von Ahn, Luis: Human Computation, Google TechTalks, 26.07.2006 [Video], Online abrufbar unter: <http://video.google.com/videoplay?docid=-8246463980976635143#> (Zuletzt abgerufen am: 03.02.2011).
- 10 Kohle, Hubertus: Bemerkungen zu einigen kunsthistorisch relevanten Adressen im Internet, in: Kohle, Hubertus (Hg.): Kunstgeschichte digital, Berlin 1997, S. 167-170, hier S. 170; siehe auch von Ahn 2006, o. S.
- 11 Vgl. Kohle, Hubertus/ Ehleiter, Regine: LIEBLINK: ARTIGO, 26.08.2008 [Interview Transkript], Online abrufbar unter: <http://www.art-magazin.de/szene/9564.html>.
- 12 Alby, Tom: Web 2.0. Konzepte, Anwendungen, Technologien., München 2007, S. 117.
- 13 <http://de.wikipedia.org/>
- 14 Putz, Nele: Artigo Social Image Tagging: Ein Kunstgeschichtsspiel zur Optimierung von Bilddatenbanken, 2008, Online abrufbar unter: <http://www.akmb.de/web/pdf/herbst2008/Putz.pdf>; und bzgl. Wikipedia vgl. Giles, Jim: Internet encyclopaedias go head to head, in: Nature 438/15, 2005, S. 900-901.
- 15 Unter anderem Änderungsbenachrichtigung für den Erstautor und regelmäßige Kontrolle missbrauchgefährdeter Artikel; Für Genaueres siehe: Feldkamp, Jens: Qualitätsaspekte im User-Generated Content, in: Grob, Heinz Lothar/ Vossen, Gottfried (Hrsg.): Entwicklungen im Web 2.0 aus technischer, ökonomischer und sozialer Sicht, Institut für Wirtschaftsinformatik der Westfälischen Wilhelms-Universität Münster 2007, S. 39-49.
- 16 Kohle/Ehleiter 2008, o. S.
- 17 Ebd.
- 18 Eine solche unbeeinflusste Gruppe kann natürlich auch Experten enthalten.
- 19 Die Eingabe selbst wird zwar abgespeichert, aber der Begriff kann nicht erneut gematcht werden und ist somit für den weiteren Verlauf des Spiels irrelevant.
- 20 Vgl. u.a. Surowiecki, James: Die Weisheit der Vielen. Warum Gruppen klüger sind als Einzelne und wie wir das kollektive Wissen für unser wirtschaftliches, soziales und politisches Handeln nützen können, München 2005.
- 21 Galton, Francis: Vox Populi, Nature 75, 07.03.1907, S. 450-451, hier S. 450; Galton, Francis : The Ballot-Box, Nature 75, 28.03.1907, S. 509-510, hier S. 509.
- 22 Für eine Sammlung von Beispielen siehe Surowicki 2005, S. 23-47; Page, Scott E.: The Difference. How the power of Diversity creates better Groups, Firms, Schools, and Societies, Princeton 2007, S. 1-4 und 177-179.; Berg, Joyce/ Forsythe, Robert/ Nelson, Forrest/ Rietz, Thomas: Results from a Dozen Years of Election Futures Markets Research, IEM Working Paper, Iowa City 2000, Online abrufbar unter: [http://www.biz.uiowa.edu/iem/archive/BFNR\\_2000.pdf](http://www.biz.uiowa.edu/iem/archive/BFNR_2000.pdf) (Zuletzt abgerufen am: 07.10.2010), o. S., Forsythe, Robert/ Nelson, Forrest/ Neumann, George R./ Wright, Jack: Anatomy of an Experimental Political Stock Market, in: The American Economic

- Review, 1992/82, Nr. 5, S. 1142-1161, Wolfers, Justin/ Zitzewitz, Eric: Prediction Markets, in: Journal of Economic Perspectives, 2004/18, Nr. 2, S. 107-126.
- 23 Surowiecki 2005, S. 23f.
- 24 Scott Page zu Folge stellt dies, unter bestimmten Bedingungen, eine statistische Tatsache dar. Genauerer Vgl. Page 2007, S. 180-196.
- 25 Surowiecki 2005, S. 40-43; Berg et al. 2000, o. S., Forsythe et al. 1992, S. 1142-1161; Wolfers/ Zitzewitz 2004, 107-126.
- 26 O'Reilly, Tim: What is Web 2.0? Design Patterns and Business Models for the Next Generation of Software, 30.09.2005, Online abrufbar unter: <http://www.oreilly.de/artikel/web20.html>.
- 27 Golder, Scott A./ Hubermann, Bernardo A.: Usage patterns of collaborative tagging systems, in: Journal of information Science, 2006/32, Nr. 2, S. 198-208., hier S. 198.
- 28 Surowiecki 2005, S. 54.; Page 2007, S. 48.
- 29 Page 2007., S. 302f.
- 30 Ebd., S. 306.
- 31 Quillian, M. Ross: Semantic Memory., in: Minsky, Marvin (Hrsg.): Semantic Information Processing, Cambridge 1968, S. 227-270., hier S. 234-240; Vgl. auch Fiske, Susan T./ Linville, Patricia W.: What does the schema concept buy us?, in: Personality and Social Psychology Bulletin, 1980/6, Nr. 4, S. 543-557, hier S. 543; für eine Übersicht der unterschiedlichen theoretischen Ansätze zur Schema-Theorie bzw. Hierarchischen Netzwerktheorie siehe Mandl, Heinz/ Friedrich, Helmut F./ Hron, Aemilian.: Theoretische Ansätze zum Wissenserwerb, in: Mandl, Heinz/ Spada, Hans (Hrsg.): Wissenspsychologie, München/Weinheim 1988, S. 123-160, hier S.124.; Schwarz, Norbert: Theorien konzeptgesteuerter Informationsverarbeitung in der Sozialpsychologie, in: Frey, Dieter/ Irle, Martin (Hrsg.): Theorien der Sozialpsychologie, Bern 1985, S. 269-292, hier s. 272.
- 32 Lässt sich aus dem Englischen ableiten: is a = ist ein. Klasse von Verbindungen in einem semantischen Netzwerk oder einer Schemarepräsentation, die die Oberbegriff-Unterbegriff-Relation anzeigt. Wenn ein Merkmal nicht direkt bei einem Konzept gespeichert ist, kann es von einem übergeordneten Konzept abgerufen werden. (Ebd.)
- 33 Collins, Allan M./ Quillian, M. Ross: Retrieval Time from Semantic Memory, in : Journal of Verbal Learning and Verbal Behavior, 1969/8, S. 240-247., hier S. 240f.
- 34 Ebd., S. 246f.
- 35 Conrad, Carol: Cognitive Economy in Semantic Memory, in: Journal of Experimental Psychology, 1972/92, Nr. 2, S. 149-154., hier S. 154.
- 36 Collins, Allan M./ Loftus, Elisabeth F.: A Spreading-Activation Theory of Semantic Processing, in: Psychological Review, 1975/82, Nr. 6, S. 407-428., hier S. 422.

37 Ebd.

38 Kiefer, Markus: Bewusstsein, in: Müsseler, Jochen (Hrsg.): Allgemeine Psychologie, Berlin 2008, S. 154-188, hier S. 169; Held, Christoph/ Cress, Ulrike: Social Tagging aus kognitionspsychologischer Sicht, in: Gaiser, Birgit/ Hampel, Thorsten/ Panke, Stefanie (Hrsg.): Good Tags – Bad Tags. Social Tagging in der Wissensorganisation, Münster 2008, S. 37-49, hier S. 42; Anderson, John R.: Kognitive Psychologie, Heidelberg 2001, S. 209.

39 Meyer, David E./ Schvaneveldt, Roger W.: Facilitation in Recognizing Pairs of Words: Evidence of a Dependence between Retrieval Operations, in: Journal of Experimental Psychology, 1971/90, Nr. 2, S.227-234., hier S. 227f.; Solso, Robert L.: Kognitive Psychologie, Heidelberg 2005, S. 114; Kroeber-Riel, Werner/ Weinberg, Peter/ Gröppel-Klein, Andrea: Konsumentenverhalten, München 2009, S. 364f.

40 Meyer/ Schvaneveldt 1971/90, S. 233.

41 Anderson 2001, S. 210.

42 Ebd., S. 209.

43 Anderson, John R.: Verbatim and Propositional Representation of Sentences in Immediate and Long-Term Memory, in: Journal of Verbal Learning and Verbal Behavior, 1974/13, S.149-162., hier S. 161f.; Anderson, John R.: Language, memory, and thought, Hillsdale 1976., hier S. 253-293 – insbesondere S. 290f.

44 Anderson 2001, S. 211.

45 Ebd.

46 Weitere Ausführungen zu einfachen zweidimensionalen statistischen Verfahren finden sich unter anderem bei Jann oder fortgeschritten bei Fahrmeir et al. (Jann, Ben: Einführung in die Statistik, München 2005; Fahrmeir, Ludwig/ Künstler, Rita/ Pigeot, Iris/ Tutz, Gerhard: Statistik. Der Weg zur Datenanalyse, Berlin 2002.)

47 Experten (gemessen an der Variable Kunstgeschichtsstudium) werden in diesem Fall von der Betrachtung ausgeschlossen, da zu erwarten ist, dass sie aufgrund ihres Fachwissens die Ergebnisse der Gruppe mit Abschlussart „Abitur“ verzerren.

48 Es wurde hierbei vor allem berücksichtigt, dass es sich bei der Stichprobe vorwiegend um Personen handelt, die sich noch im Studium befinden oder ein solches bereits abgeschlossen haben.

49 Unter anderen gelistete Begriffe: »Beine«, »Bein«, »Arme«, »Arm«, »Hand«, »Finger«, »Kopf«, »Haare«, »Himmel«, »Wolken«, »Gemälde«, »Porträt«, »grau«, »weiß«, »schwarz« und »braun«.

50 Marlow, Cameron/ Naaman, Mor/ Boyd, Danah/ Davis, Marc: HT06, tagging paper, taxonomy, Flickr, academic article, to read., in: ACM Press (Hrsg.): HYPERTEXT '06: Proceedings of the seventeenth conference on Hypertext and hypermedia New York, 2006, S. 31-39., hier S. 34.