

Vom Suchen, Stöbern und Finden.

Information Retrieval am Beispiel der Digitalen Sammlung
des Hans Gross Kriminalmuseums

M a s t e r a r b e i t

zur Erlangung des akademischen Grades
Master of Arts (MA)

an der Karl-Franzens-Universität Graz

vorgelegt von

Christopher POLLIN

am Zentrum für Informationsmodellierung -
Austrian Centre for Digital Humanities

Begutachter: Ass.-Prof. Mag. Dr.phil. Johannes Hubert Stigler

Graz, 2017

Eidesstattliche Erklärung

Ich erkläre ehrenwörtlich, dass ich die vorliegende Arbeit selbständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen nicht benutzt und die den Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen inländischen oder ausländischen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht. Die vorliegende Fassung entspricht der eingereichten elektronischen Version.

Datum

Unterschrift

Vorwort

*"Und sehe, daß wir nichts wissen können!
Das will mir schier das Herz verbrennen."*

- GOETHE, Faust I

So der verzweifelnde Faust, als er feststellte, dass je mehr er die Welt verstand, desto mehr Fragen taten sich für ihn wieder auf. Egal wie intensiv ein Thema analysiert und reflektiert wurde, kam er nie zu einer endgültigen Antwort, einer absoluten Gewissheit, einem absoluten Wissen. Sir Karl Popper zeichnet in diesem Zusammenhang folgendes Bild: Jemand der sich an einem Strand befindet, weiß nicht was im dahinter liegenden gigantischen Ozean verborgen ist. Beim Verfassen dieser Arbeit wurde ich mir der Größe dieses Ozeans bewusst. Besucht wurde nur einen kleinen Küstenbereich. Diese Arbeit, die sich mit dem Zurechtfinden am Küstenbereich beschäftigt, dem Wiederfinden von Information, ist der Versuch - zumindest für mich - einen Küstenabschnitt besser verstehen zu können.

Ich danke recht herzlich meinen Studienkolleginnen und Studienkollegen für die Zusammenarbeit im Studium und im Studienprojekt, Bernadette Biedermann, die mit Herzblut und Fachkompetenz das Projekt begleitet hat, meinen Arbeitskolleginnen und Arbeitskollegen am Zentrum für Informationsmodellierung, für das sehr liebevolle, freundliche und hilfsbereite Arbeitsklima, meinem Betreuer Johannes Stigler für das Vertrauen und die Betreuung in Projekt und Arbeit, meiner Mutter, meinem Bruder und Schwägerin für das Korrekturlesen, sowie meinen Freunden und meiner Familie für die Unterstützung für ein Studium und einen Beruf, den ich immer wieder aufs Neue erklären muss. Etwas, das mir schier das Herz verbrennt. Mit Brennen in der Brust dennoch ein herzliches Danke!

Aus Gründen der Lesbarkeit werden männliche und weibliche Formen alternierend verwendet. Das bedeutet, dass in dieser Arbeit männliche und weibliche Formen in einem ähnlichen Verhältnis zueinander auftreten.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Aufbau der Arbeit	2
1.2	Literaturüberblick	4
2	Perspektiven auf das Phänomen 'Suchen'	7
2.1	Einleitung	7
2.2	Philosophische Perspektive	7
2.3	Gesellschaftliche Perspektive	14
2.4	Fazit	19
3	Suchen: Information Retrieval	20
3.1	Einleitung	20
3.2	Grundbegriffe der Informationswissenschaft	20
3.2.1	Der Ausgangspunkt: Signal und Daten	21
3.2.2	Der Informationsbegriff	24
3.2.3	Der Wissensbegriff	26
3.3	Grundbegriffe des Information Retrieval	29
3.3.1	Informationsbedarf	29
3.3.2	Dokumente	30
3.3.3	Suchanfrage und ihre Formulierung	32
3.3.4	Index und Ranking	34
3.3.5	Ordnungsstrukturen und Wissensrepräsentation	36
3.4	Information Retrieval-Modelle	39
3.4.1	Boolesches Modell	40
3.4.2	Vektorraummodell	41
3.4.3	Probabilistische Modell	43
3.5	Fazit	43
4	Stöbern: Resource Discovery	45
4.1	Einleitung	45
4.2	Ein erweitertes Informationsbedürfnis: Exploratory Search	46
4.3	User-Interaktivität: Als Hilfestellung zur Query-Formulierung	47
4.4	Organisation von Suchergebnissen: Als strukturgebende Methode	50
4.5	Informationsvisualisierung: Als Discovery-Strategie	51
4.6	Semantische Suche: Als Suche nach Bedeutung	52
4.7	Fazit	54

5	Prinzipien semantischer Suche am Beispiel der virtuellen Repräsentation des Hans Gross Kriminalmuseums	56
5.1	Einleitung	56
5.2	Projektbeschreibung des Studienprojektes	56
5.2.1	Projektdefinition und Projektziele	56
5.2.2	Das Hans Gross Kriminalmuseum und vorhandenes Quellenmaterial . .	57
5.2.3	Rollen, Work-Packages und Milestones	59
5.3	Technische Grundlagen zur Umsetzung des Studienprojektes	63
5.3.1	Digitale Langzeitarchivierung und digitales Objekt	63
5.3.2	Metadaten und Metadatenstandards	65
5.3.3	X-Technologien	67
5.3.4	Funktionsweise der GAMS-Infrastruktur und Umsetzung des Projekts .	69
5.4	Technische Umsetzung der semantischen Suche und Discovery Möglichkeiten .	73
5.4.1	Einführung: Semantische Technologien	73
5.4.2	Aufbereitung der Karteikarten: TEI/XML nach RDF/XML	76
5.4.3	Abfrage der Karteikarten: SPARQL	79
5.4.4	Resource Discovery der Karteikarten: SPARQL nach HTML	80
6	Fazit und Ausblick	85
7	Anhang	88
7.1	Source Code	88
7.1.1	TEI/XML der Karteikarte o:km.87	88
7.1.2	XSL-Stylesheet zur Transformation der TEI/XML-Karteikarten nach RDF/XML	91
7.1.3	RDF/XML der Karteikarte o:km.87	95
7.1.4	SPARQL zur Abfrage alle Karteikarten	97
7.1.5	XSL-Stylesheet zur Transformation: SPARQL Ergebnis zu HTML Re- präsentation des Resource Discovery	98
7.2	Abkürzungsverzeichnis	104
7.3	Onlineressourcen	105
7.4	Projektspezifische Onlineressourcen	107

1 Einleitung

Wissenschaft behandelt Probleme, formuliert Fragen, bietet Lösungen an und versucht sich einer Wahrheit anzunähern. Sie baut ihre Überlegungen auf Quellen und Datenmaterial auf, damit sie im kritischen Diskurs argumentieren kann, warum sie Theorien und Ergebnisse auf diese Art und Weise formuliert hat. Wissenschaft muss sich der kritischen Diskussion und Überprüfbarkeit stellen und im reflexiven Prozess versuchen sich stetig weiterzuentwickeln.

Die Ansprüche wissenschaftlichen Arbeitens sind hoch und vielfältig. Eine grundlegende Gemeinsamkeit ist die Suche nach etwas. So ist das Suchen nach Literatur für das Verfassen einer wissenschaftlichen Arbeit, wie es sich beispielsweise im Alltag eines Studierenden finden lässt, eine Form einer bewussten Handlung. Für die Studierenden ist der Präsenzbestand einer Fachbibliothek ein erster geeigneter Ort, um mit der **Suche** nach Literatur zu beginnen. Auf dieser Suche nach der richtigen Signatur und dem Platz, an dem sich das Buch im Regal befindet, kann dabei ein Werk entdeckt werden, das neue und bereichernde Perspektiven hinzufügen kann. Da Fachbibliotheken oft nach Themen geordnet sind, könnten sich dort im Regal mehrere interessante und spannende Werke befinden. Bücher werden herausgenommen, aufgeschlagen, durchgeblättert und Buchrücken durchgelesen. Das **Stöbern** hat begonnen und zu fachübergreifenden, vorher nicht bedachten, zusätzlichen Informationsressourcen geführt. Das **Finden** geeigneter Literatur steht am Ende dieses, der Wissenschaft immanenten, Prozesses.

Suchen, Stöbern und Finden sind komplexe Vorgänge, in denen ein Akteur ein Informationsbedürfnis hat und aus diesem Grund eine Suchanfrage (Query) an einen Informationskorporus stellt. Ein Informationskorporus ist eine Ansammlung von Dokumenten, die nach einem gewissen Schema gekennzeichnet (indexiert) wurden, um Suchprozesse zu ermöglichen. Überschneiden sich Suchanfrage und Inhalt des Informationskorporus, steht eine Menge an gefundenen Dokumenten zur Verfügung, die als Antwort zur Befriedigung des Informationsbedürfnisses dienen kann. Diese Treffermenge ist nach einer gewissen Ordnung gereiht (Ranking), die dem Zweck dient, scheinbar wichtigere Dokumente nach vorne, und unwichtigere Dokumente nach hinten zu reihen. Ob damit das Informationsbedürfnis befriedigt wird, zeigt die Reaktion des Suchenden an, der auf Basis dieses Prozesses neue Informationsbedürfnisse entwickelt oder neue Anfragen formuliert, um die Treffermenge nach seinen Bedürfnissen weiter zu konkretisieren. Es liegt in seinem Ermessen zu entscheiden, welche Dokumente er als relevant und welche er als nicht relevant, befindet. Abbildung 1 stellt diesen Retrieval-Prozess dar.

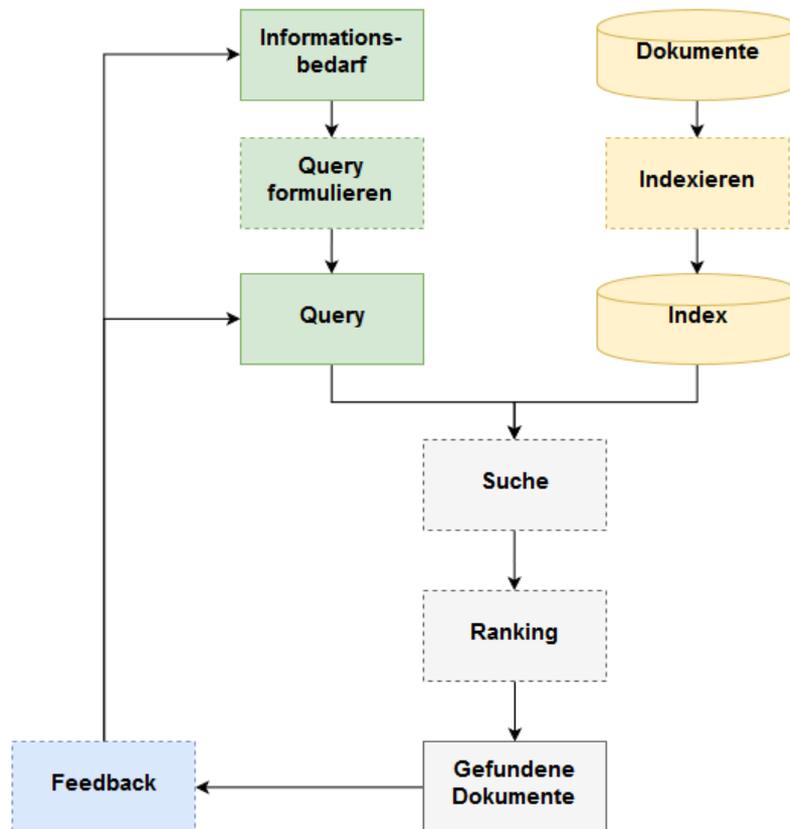


Abbildung 1: Modell des Retrieval-Prozess

Bis es zu einem Ergebnis kommt, müssen eine Reihe an Voraussetzungen erfüllt sein. Akteur und Informationskorpus müssen miteinander 'kommunizieren' um Anfragen, sowie Antworten formulieren können.

1.1 Aufbau der Arbeit

In dieser Arbeit sollen die Bausteine für das IR ausführlich erörtert und in einen praxisbezogenen Kontext gesetzt werden. Ziel der Arbeit ist es der Frage nach zu gehen, inwieweit Information Retrieval und Resource Discovery für geisteswissenschaftliches Datenmaterial in einer digitalen Sammlung im World Wide Web umgesetzt werden kann. Folgende Fragestellungen beschreiben den Aufbau der Arbeit und orientieren sich gleichzeitig am in Abbildung 1 dargestellten Modell des Retrieval-Prozesses.

- **Welchen Stellenwert haben Retrieval-Prozesse für Gesellschaft und Wissenschaft?**

In **Kapitel 2** wird der Versuch gewagt das Thema 'Suchen' aus einer philosophischen und gesellschaftlichen Perspektive heraus zu betrachten, um sich dem Konzept Suchen und seiner Auswirkungen auf Wissenschaft und Gesellschaft annähern zu können. In diesem Zusammenhang sind die Begriffe *Wirklichkeit*, *Wahrnehmung*, *Erkenntnis* und *Wahrheit* von zentraler Bedeutung. Unter Berücksichtigung diese Begriffe wird *Information als Wirtschaftsgut*, *KUHLEN's Wissensökologie* und *Google als Informationsmonopol* un-

tersucht und der ökonomische, gesellschaftliche und soziale Stellenwert von Informationsressourcen und ihrer Verteilung herausgearbeitet.

- **Welche Theorien, Technologien und Modelle liegen dem Retrieval-Prozess zu Grunde?**

Diese Fragestellung wird in **Kapitel 3** behandelt. Es widmet sich dem 'klassischen' Zugang zu theoretischen, methodischen und technologischen Aspekten des IR. In diesem Zusammenhang werden Grundbegriffe der Informationswissenschaft: *Daten*, *Information* und *Wissen* definiert. Weiters werden die 'klassischen' Fachbegriffe und Verfahren des IR, sowie die gängigen IR-Modelle beschrieben.

- **Welche weiteren Möglichkeiten des IR im Sinne des Stöberns und Entdeckens hat ein Akteur im digitalen Raum?**

Das **Kapitel 4** beschäftigt sich mit IR im Sinne des Resource Discovery (RD) und beschreibt die erweiterte Form dieses Informationsbedürfnisses und gibt *User-Interaktivität*, *Organisation von Suchergebnissen*, *Informationsvisualisierung* und *semantische Suche* als Beispiel für Discovery-Strategien im World Wide Web an.

- **Wie kann Resource Discovery konkret umgesetzt werden?**

Aufbauend auf den behandelten Themen wird die praktische Arbeit im Abschlussprojekt, in **Kapitel 5**, des EuroMACHS-Masterstudiengangs¹ des Durchgangs 2014 bis 2016 dokumentiert und näher ausgeführt. Dies umfasst die Aufgabenbereiche der Projektplanung, bestehend aus Projektplan und Pflichtenheft, einer Beschreibung der Sammlung des Hans Gross Kriminalmuseums, der Implementierung der Webpräsenz, sowie der IR- bzw. RD-Funktionalitäten, unter Verwendung semantischer Technologien bzw. der Prinzipien semantischer Suche, im Wechselspiel aus theoretischen Grundlagen für Technologien, sowie der GAMS-Infrastruktur und praktischer Anwendung.

Erst das Zusammenspiel der genannten Bausteine ermöglicht effizientes Suchen und damit Informations(rück)gewinnung, oder wie '*die Wissenschaft, die Technik und der Praxisbereich des Suchens und Findens von Information*'² im Englischen heißt: **Information Retrieval IR**).

¹Europe, Digital Media, Arts and Cultural Heritage Studie, <http://euromachs.uni-graz.at>

²STOCK, Wolfgang G.: Information retrieval: Informationen suchen und finden. Band 1, Oldenbourg Verlag, 2007, S.2.

1.2 Literaturüberblick

An dieser Stelle ist eine Auswahl der wichtigsten Literatur für diese Arbeit zusammengefasst. STOCK liefert mit seinen zwei Bänden *Wissensrepräsentation*³ und *Information Retrieval*⁴ zwei Standardwerke, die einen umfangreichen und konzeptuellen Überblick über den Themenbereich IR und die **Informationswissenschaft** an sich liefern. In diesen Ausführungen wird auf Klassiker der Informationswissenschaft wie KUHLEN⁵ oder WERSIG⁶ verwiesen, die sich ausführlich mit den theoretischen Grundlagen beschäftigt haben. GAUS⁷ konzentriert sich darauf Information zu strukturieren, um Erschließungsprozesse im Sinne der Dokumentations-, und Ordnungslehre zu verbessern. FAVRE-BULLE⁸ ergänzt mit seiner ausführlichen Ausarbeitung der Begriffstrios Daten-Information-Wissen, diese theoretischen Auseinandersetzungen.

Im Bereich des **Information Retrieval** liefern die bereits genannten Arbeiten von STOCK und POETZSCH⁹, sowie FERBER¹⁰ eine Einführung in Grundlagen, Methoden und fachbezogene Aspekte des Information Retrieval. Technische Grundlagen hinsichtlich IR und Suchmaschinen lassen sich sehr ausführlich bei BAEZA-YATES et al.¹¹ finden. Die Thematik wird im Sammelband von GÖKER et DAVIES¹² vertieft, indem nicht-textbasiertes Information Retrieval, mobiles Suchen oder beispielsweise die semantisch Suche erörtert werden.

Da die praktische Umsetzung in die Richtung einer **semantischen Suche** steuert bzw. semantische Technologien verwendet werden, bieten sich die Arbeiten von VIRGILIO, GUERRA et VELEGRAKIS¹³, sowie LINCKELS et MEINEL¹⁴ an, da sie sich direkt mit semantischen Suchen im World Wide Web und in digitalen Bibliotheken beschäftigen. Als Grundlage zum

³STOCK, Wolfgang G./STOCK, Mechtild: Wissensrepräsentation: Informationen auswerten und bereitstellen. Band 2, Oldenbourg Verlag, 2008.

⁴STOCK: Information retrieval: Informationen suchen und finden.

⁵KUHLEN, Rainer: Information - Informationswissenschaft. In KUHLEN, Rainer/SEMAR, Wolfgang/STRAUCH, Dietmar (Hrsg.): Grundlagen der praktischen Information und Dokumentation: Handbuch zur Einführung in die Informationswissenschaft und -praxis. 6. Auflage. Berlin and Bosten, 2013.

⁶WERSIG, Gernot: Information-Kommunikation-Dokumentation. Ein Beitrag zur Orientierung der Informations- und Dokumentationswissenschaft. Band 5, Pullach bei München, 1974.

⁷GAUS, Wilhelm: Dokumentations- und Ordnungslehre: Theorie und Praxis des Information Retrieval. Springer-Verlag, 2006.

⁸FAVRE-BULLE, Bernard: Information und Zusammenhang: Informationsfluß in Prozessen der Wahrnehmung, des Denkens und der Kommunikation. Springer, 2001.

⁹POETZSCH, Eleonore: Information Retrieval: Einführung in Grundlagen und Methoden. Berlin-Brandenburg, 2005.

¹⁰REGINALD, Ferber: Information Retrieval-Suchmodelle und Data-Mining-Verfahren für Textsammlungen und das Web. Heidelberg, 2003.

¹¹BAEZA-YATES, Ricardo/RIBEIRO-NETO, Berthier et al.: Modern information retrieval. The concept and technology behind search. 2. Auflage. 2011.

¹²GOKER, Ayse/DAVIES, John: Information retrieval: searching in the 21st century. Chichester, 2009.

¹³DE VIRGILIO, Roberto/GUERRA, Francesco/VELEGRAKIS, Yannis: Semantic search over the web. Springer Science & Business Media, 2012.

¹⁴LINCKELS, Serge/MEINEL, Christoph: E-librarian service: user-friendly semantic search in digital libraries. Springer Science & Business Media, 2011.

Semantic Web kann KASHYAP, BUSSLER et MORAN¹⁵ herangezogen werden. Arbeiten zur Nutzung semantischer Technologien für IR-Systeme wurden auch von HILDEBRAND et al.¹⁶, GRAUPMANN¹⁷ und DENGEL¹⁸ veröffentlicht.

Arbeiten, die sich mit **Resource Discovery** beschäftigen, findet sich bei RÜGER¹⁹ oder GESER²⁰. JORDANOUS²¹ beschreibt die Umsetzung von RD-Funktionalitäten unter Hilfenahme semantische Technologien: Dabei geht er von Datensätzen aus die in TEI/XML formalisiert wurden.

Die Darstellung von Information und die Umsetzung von **User Interfaces** ist ein weiterer zentraler Aspekt der Arbeit. Für einen vertiefenden Einblick in die praktische und theoretische Auseinandersetzung mit User Interfaces und Usability empfiehlt sich HEARST²², SHNEIDERMAN²³. Grundlegendes zur Darstellung von Information findet sich bei TUFTE²⁴, das Themenfeld der **Informationsvisualisierung** bei CARD et. al²⁵ und FORTMÜLLER.²⁶

Hinsichtlich einer **philosophischen und gesellschaftlichen Perspektive** auf die Thematik kann der Sammelband von SPINK und ZIMMER²⁷, KUHLEN's Theorie der Wissensökologie²⁸ und die Arbeiten im Sammelband von BECKER und STALDER²⁹, die sich mit politischen Fragen

¹⁵VIPUL, Kashyap/BUSSLER, Christoph/MORAN, Matthew: The Semantic Web-Semantics for Data and Services on the Web. Berlin and Heidelberg, 2008.

¹⁶HILDEBRAND, Michiel/OSSENBRUGGEN, Jacco Ronald van/HARDMAN, Lynda: An analysis of search-based user interaction on the semantic web. 2007.

¹⁷GRAUPMANN, Jens: Semantische Suche: intelligente Suchmaschinen durch innovative und zukunftsweisende Konzepte und Technologien. Saarbrücken, 2007.

¹⁸DENGEL, Andreas: Semantische Technologien: Grundlagen - Konzepte - Anwendungen. Heidelberg, 2012.

¹⁹RÜGER, Stefan: Multimedia Resource Discovery. In GOKER, Ayse/DAVIES, John (Hrsg.): Information retrieval: searching in the 21st century. Chichester, 2009.

²⁰GESER, Guntram: Resource Discovery - Position Paper: Putting the Users First. In GESER, Guntram/PEREIRA, John (Hrsg.): DigiCULT Thematic Issue 6 – Resource Discovery Technologies for the Heritage Sector. Salzburg and Glasgow, 2004.

²¹JORDANOUS, Anna: Enhancing information retrieval and resource discovery from data using the Semantic Web. In Emerging Trends and Technologies in Libraries and Information Services (ETTLIS), 2015 4th International Symposium on. IEEE 2015.

²²HEARST, Marti: Search User Interfaces. Cambridge, 2009.

²³SHNEIDERMAN, Ben: Designing the user interface: strategies for effective human-computer interaction. Pearson Education India, 2010.

²⁴TUFTE, Edward: Envisioning Information. Cheshire, 2008.

²⁵CARD, S.K./MACKINLAY, J.D./SHNEIDERMAN, B.: Readings in Information Visualization: Using Vision to Think. San Francisco, 1999.

²⁶FORTMÜLLER, Walter: Informationsvisualisierung: Analyse und Evaluierung ausgewählter Applikationen zur interaktiven visuellen Exploration von Datenbeständen. Graz, 2011.

²⁷SPINK, Amanda/ZIMMER, Michael: Web search: Multidisciplinary perspectives. Band 14, Springer Science & Business Media, 2008.

²⁸KUHLEN, Rainer: Wissensökonomie und Wissensökologie zusammen denken. In HELFRICH, Silke (Hrsg.): Commons: für eine neue Politik jenseits von Markt und Staat. Bielefeld, 2012.

²⁹BECKER, Konrad/STALDER, Felix: DEEP SEARCH: Politik des Suchens jenseits von Google. Bundeszentrale für politische Bildung, 2010.

des IR auseinandersetzen, herangezogen werden. NIEDERMAIR³⁰ sieht in der (meta) informationellen Autonomie ein wichtiges Element für eine funktionierende Gesellschaft. MASON³¹, BAUERNHANSL et al.³² und PILLER³³ beschäftigen sich mit dem ökonomischen Stellenwert von Information und Wissen und dem Wandel der Industriegesellschaft hin zu einer Informationsgesellschaft.

Die Ausführungen zur philosophischen Auseinandersetzung basieren zentral auf den Grundlagenwerken der Philosophie von WUCHTERL³⁴ und SALAMUN³⁵, sowie Arbeiten von POPPER³⁶. Diese Literatur wurde für die Themenkomplexe Wirklichkeit, Wahrnehmung, Erkenntnis und Wahrheit herangezogen.

Projektbezogene Inhalte zum **Hans Gross Kriminalmuseum** wurden von BACHHIESEL³⁷ und KOCHER³⁸ veröffentlicht. Für weitere Grundlagen zur **GAMS-Infrastruktur**³⁹ empfiehlt sich folgende Literatur: für Metadaten(Standards) JEHN⁴⁰, für Langzeitarchivierung BRÜBACH⁴¹ und KEITEL⁴².

Für weitere Begriffsdefinitionen, besonders bei Standards, Webtechnologien oder Programmiersprachen, wird auf die Webseite des World Wide Web Consortium⁴³ und auf die einschlägigen Wikipedia-Artikel, die eine hohe Qualität aufweisen, verwiesen.

³⁰NIEDERMAIR, Klaus: Gefährden Suchmaschinen und Discovery-Systeme die informationelle Autonomie? Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare, 67 2014, Nr. 1.

³¹MASON, Paul: Postkapitalismus: Grundrisse einer kommenden Ökonomie. Berlin, 2016.

³²BAUERNHANSL, Thomas/TEN HOMPEL, Michael/VOGEL-HEUSER, Birgit: Industrie 4.0 in Produktion, Automatisierung und Logistik: Anwendung· Technologien· Migration. Wiesbaden, 2014.

³³PILLER, Frank Thomas: Einführung: Informationsrevolution und industrielle Produktion. In Mass Customization. Wiesbaden, 2000.

³⁴WUCHTERL, Kurt: Lehrbuch der Philosophie. Bern and Stuttgart, 1998.

³⁵SALAMUN, Kurt: Was ist Philosophie? Tübingen, 2009.

³⁶POPPER, Karl R: Auf der Suche nach einer besseren Welt: Vorträge und Aufsätze aus dreißig Jahren. München, 1984.

³⁷BACHHIESL, Christian: Räuber, Mörder, Sittenstrolche. 37 Fälle aus dem Kriminalmuseum der Karl-Franzens-Universität Graz. 2. Auflage. Graz, 2004.

³⁸KOCHER, Gernot: Das K.k kriminalistische Universitätsinstitut in Graz. In BACHHIESL, Christian/BACHHIESL, Sonja Maria/LEITNER, Johann (Hrsg.): Kriminologische Entwicklungslinien. Eine interdisziplinäre Synopsis. Wien and Berlin, 2014.

³⁹GAUS: Dokumentations- und Ordnungslehre: Theorie und Praxis des Information Retrieval.

⁴⁰JEHN, Mathias: Metadatenstandards im Bereich der digitalen LZA. In NEUROTH, Heike et al. (Hrsg.): nestor-Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Göttingen, 2010 (URL: http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_476.pdf).

⁴¹BRÜBACH, Nils: Das Referenzmodell OAIS. In NEUROTH, Heike et al. (Hrsg.): nestor-Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Göttingen, 2010 (URL: http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_438.pdf).

⁴²KEITEL, Christian; SCHOGER, Astrid (Hrsg.): Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644. Berlin and Wien and Zürich, 2013.

⁴³W3C, <https://www.w3.org>, 05.03.2017.

2 Perspektiven auf das Phänomen 'Suchen'

2.1 Einleitung

Die Thematik des Suchens ist vielschichtig und ist Bestandteil unterschiedlichster Lebensbereiche und Disziplinen. Sie setzt sich aus vielen Fragmenten zu einem komplexen Ganzen zusammen. Dieser Abschnitt der Arbeit soll als eine Annäherung verstanden werden, um die Bedeutung des IR für Wissenschaft und Gesellschaft hervorzuheben. Dies umfasst eine philosophische und eine gesellschaftliche Annäherung, in denen versucht wird, mit Hilfe facheinschlägiger Literatur, einen Überblick zu erhalten, Definitionen zu formulieren und Fragen aufzuwerfen, die für die vorliegende Arbeit wichtig erscheinen.

Die *Wahrnehmung der Wirklichkeit* und die daraus gewonnenen *Erkenntnisse* und die Überprüfung ihres *Wahrheitsgehalts* spielen im Retrieval-Prozess eine wichtige Rolle. Vor allem der Zugang und die Ideen von Sir Karl POPPER prägen diesen Diskurs der philosophischen Perspektive in Kapitel 2.2. Seine und die Auseinandersetzungen andere Autoren mit diesen Fragen prägen wiederum die Debatte zur gesellschaftlichen Perspektive des IR, in der die ökonomischen, sozialen und politischen Auswirkungen auf die Gesellschaft in Kapitel 2.3 dargestellt werden sollen. Hier sei vor allem KUHLEN genannt, der sich mit diesen Fragestellungen auseinandergesetzt hat.

2.2 Philosophische Perspektive

Der Mensch sucht. Suchen ist eine spezielle Form des menschlichen Handelns mit dem Ziel ein konkretes Bedürfnis zu stillen. Dieses Handeln erstreckt sich von der alltäglichen Suche nach dem Haustürschlüssel, über die Recherche nach Informationsressourcen im World Wide Web, bis zu abstraktem und transzendtem Suchen, wie der Suche nach dem Sinn des Lebens. Wir suchen nach Anerkennung, nach Information, nach Glück. Suchen dürfte eine Form des Handelns sein, die dem Menschen immanent ist. Das Modell des Retrieval-Prozesses, als wissenschaftliches Ergebnis, basiert auf grundlegenden philosophischen Denkmustern, die im Folgenden diskutiert werden sollen.

*"Jedes Fragen ist ein Suchen. Jedes Suchen hat ein vorgängiges Geleit aus dem Gesuchten her. Fragen ist erkennendes Suchen des Seienden in seinem Daß- und Sosein."*⁴⁴,

formulierte Martin HEIDEGGER. Als Philosoph beschäftigt sich HEIDEGGER mit der Frage nach dem 'Sein'. Er verstand Suchen, als Erkenntnis erzeugende Interaktion mit dem 'Sein' in Form einer Frage. Kein Suchprozess kann ohne eine Frage existieren, sowie keine Frage ohne eine Vorgeschichte formuliert werden kann. Suchen entsteht somit nicht aus dem Nichts. Aus diesem Zitat von HEIDEGGER lassen sich vier philosophische Fragen ableiten, die zentral für

⁴⁴Vgl. HEIDEGGER, Martin: Sein und Zeit. 19. Auflage. Tübingen, 2006, S.5.

wissenschaftliches Arbeiten und daraus entstehenden Retrieval-Prozesse sind: Was ist **Wirklichkeit**, das 'Daß- und Sosein'? Was versteht man unter **Wahrnehmung** und **Erkenntnis**, als 'erkennendes Suchen des Seienden'? Was ist **Wahrheit**, also 'Evaluierung der Antwort auf die Frage'?

Wirklichkeit

Ausgangspunkt dieser Überlegung ist die Frage danach, was die Wirklichkeit ist. Die **Ontologie** ist die wissenschaftliche Disziplin der theoretischen Philosophie, die sich mit der Struktur der Wirklichkeit auseinandersetzt.⁴⁵ Der ontologische Realismus, der ontologische Idealismus und der Kantianismus werden in diesem Zusammenhang unterschieden. Der ontologische Realismus geht davon aus, dass die Welt unabhängig von Gedanken und Sprachprozessen, sprich ohne die menschliche Existenz, existieren würde. Dem gegenüber steht der ontologische Idealismus, der die Wirklichkeit als Produkt menschlicher Wahrnehmung und Verarbeitung sieht. Der Kantianismus wiederum versucht beide Aspekte zusammenzuführen, indem zwischen einer realen Wirklichkeit und einer subjektiven Wahrnehmung dieser Wirklichkeit unterschieden wird. Die Wirklichkeit besteht aus Dingen, die wahrgenommen werden können und dem Wissen, welches sich auf die Wahrnehmung der Dinge beschränkt.⁴⁶

Ein anderer Zugang findet sich bei **POPPER**. Zum einen versteht er als Wirklichkeit alle materiellen Körper, die existieren und von Menschen mit ihren Sinnen wahrgenommen werden können. Für POPPER ist dies die Körperwelt, bestehend aus belebten und unbelebten Körpern. Dazu gehören auch Phänomene wie physikalische Kräfte. Dann sind solche Dinge wirklich, die nicht berührt oder gemessen werden können, die aber erlebbar sind. POPPER bezeichnet das als bewusste und unbewusste Erlebnisse der Menschen. Dies sind Gedanken, Gefühle, Empfindungen, die jeder Mensch für sich wahrnehmen kann. Eine dritte Welt sieht POPPER in den geistigen, geplanten oder gewollten Produkten menschlicher Geisteskraft. Hier sind Ideen oder Probleme angesiedelt. Als Welt versteht er Teilstücke, die in Wechselwirkung miteinander stehen und so die Wirklichkeit ergeben. Der Mensch ist in der Lage diese Wirklichkeit zu gestalten. Der Traum vom Fliegen, aus der Welt des Bewusstseins (Welt 2), wurde im Laufe der Geschichte mittels Idee aus der Welt der objektiven Gedankeninhalte (Welt 3) erfüllt und hat sich in der Körper- bzw. Außenwelt (Welt 1) manifestiert, da man heutzutage mittels Flugzeug von A nach B gelangen kann.⁴⁷

Suchen im Sinne des IR umfasst alle drei Welten. Ideen aus der Welt 3 werden in materiellen Objekten 'gespeichert', um sie anderen Menschen zur Verfügung zu stellen. Um zu diesen Ideen und ihrer Repräsentation in Welt 1 - z.B. der Verschriftlichung einer Idee in einem Buch

⁴⁵Unterscheidet sich vom Informatikbegriff 'Ontologie' (siehe 5.4.1), Vgl. Artikel 'Ontologie' In: SANDKÜHLER, Hans Jörg et al.: Europäische Enzyklopädie zu Philosophie und Wissenschaften. 1994.

⁴⁶Vgl. HERMANS, Jan: Ontologiebasiertes Information Retrieval für das Wissensmanagement. Band 39, Berlin, 2008, S.2-3.

⁴⁷Vgl. POPPER: Auf der Suche nach einer besseren Welt: Vorträge und Aufsätze aus dreißig Jahren, S.16-40.

- zu gelangen, bedarf es eines Suchprozesses. Das Verwerten dieser Ideen und der Erhalt dergleichen durch einen Kommunikationsprozess und das dadurch gestillte - oder nicht gestillte - Bedürfnis kann als ein individuelles Erlebnis aufgefasst werden.

Es stellt sich jedoch die Frage, was genau zu diesen Welten hinzugerechnet werden kann und inwieweit die menschliche Erkenntnisfähigkeit die Wirklichkeit richtig verarbeitet. In diesem Zusammenhang lässt sich die Frage formulieren inwieweit IR-Technologien, wie beispielsweise eine Suchmaschine im Web, nun in der Lage sind die Wirklichkeit abzubilden? Es ist die Aufgabe einer Suchmaschine die Wirklichkeit nach gewissen Kriterien einzugrenzen, ansonsten müssten die Suchergebnisse alles, das in der Wirklichkeit vorkommt, ausgeben. Eine Suchmaschine gibt also per Definition ein reduziertes Bild der Wirklichkeit zurück. Doch dieses Bild ist verzerrt. Weder kann ein Akteur die perfekte Suchanfrage formulieren, noch gibt es eine allwissende Suchmaschine, die im gesamten Suchraum der Wirklichkeit nach Treffern suchen könnte. Für den Menschen selbst steht nur ein Teil der Wirklichkeit zur Verfügung, den er wahrnehmen kann. Die Idee des infiniten Regress, beschreibt eine unendliche Reihe von Ursachen, die sich selbst bedingen.⁴⁸ Ein Beispiel dafür ist die Frage danach was vor dem Beginn des Universums war. Auf die Begründung wiederum folgt die Frage danach, was vor dieser Begründung war. Dieses Beispiel soll zeigen, dass der Mensch Fragen stellen bzw. Suchanfragen formulieren kann, die außerhalb der wahrnehmbaren Wirklichkeit des Menschen liegen und aus denen, außer dem kreativen freien Denken, keine konkreten Erkenntnisse abgeleitet werden können. Mit einer Idee was Wirklichkeit nun sein könnte, ist es nun notwendig die Begriffe Wahrnehmung und Erkenntnis weiter auszuführen.

Wahrnehmung und Erkenntnis

Der Mensch und alle seine Fähigkeiten kommen aus der Natur. Es ist kein Zufall, dass der Mensch so ist, wie er ist, sondern das Ergebnis der Evolution und der Welt (oder den Welten, wenn man an POPPER denkt) um ihn herum. Den Menschen kann man als **kognitiven Agenten** bezeichnen. Darunter versteht man Lebewesen oder intelligente Maschinen, die über Erkenntnisfähigkeit verfügen, also auf Grund von empfangenen und verarbeiteten Informationsflüssen, Rückschlüsse auf die Wirklichkeit ziehen können und dementsprechend bewusst oder unbewusst handeln.⁴⁹ WERSIG geht von einer objektiven Außenwelt aus, die ein Organismus (bzw. kognitiver Agent) nur erahnen, erfahren und in ihr interagieren, aber sie nie in ihrer Gesamtheit verstehen kann. Über Perzeptionen wird die Außenwelt (wahrnehmbare Wirklichkeit) erfasst und in einer operationalen Funktionseinheit gespeichert. Mit den bereits gespeicherten Informationen und der internen Motivation werden Handlungen in der Welt getätigt. Der Organismus arbeitet mit einem *'internen Außenweltmodell'*, einem offenem Set interner Regeln,

⁴⁸Infinites Regress, https://de.wikipedia.org/wiki/Infinites_Regress, 25.01.2017.

⁴⁹Vgl. FAVRE-BULLE: Information und Zusammenhang: Informationsfluß in Prozessen der Wahrnehmung, des Denkens und der Kommunikation, S.100.

Konzepten und Strukturen, um Informationen effektiv verarbeiten zu können.⁵⁰ Die grundlegenden IR-Funktionen des Menschen, mit denen er mit der Außenwelt interagieren kann, könnte man als operative Methoden des Organismus nach WERSIG verstehen.

Die Verarbeitung von Information durch einen kognitiven Agenten und das daraus neu geschaffene Wissen beschreibt die Erkenntnisfähigkeit des Menschen. Sie ist ein Produkt menschlicher Lebensumwelt und die evolutionäre Antwort, um in dieser Welt überleben zu können. Noch verfügen Maschinen nicht über diese Fähigkeiten, da künstliche Intelligenz bis heute nicht erreicht wurde. Jedoch können intelligente Maschinen auch als kognitive Agenten definiert werden, wenn sie auf Basis ihrer Wahrnehmung (Messgeräte) eigenständige Erkenntnisse über die Wirklichkeit auf Basis einer intelligenten Recheneinheit erzeugen können. Somit kann gesagt werden, dass Wahrnehmung zweckmäßig ist. Der Mensch kann mit seinen Sinnen das wahrnehmen, das für ihn auch von Relevanz ist. Das Auseinanderdriften von Kontinenten kann man nicht mit dem Auge erkennen, da es ein Prozess von Millionen von Jahren ist. Die Geschwindigkeit dieses Prozesses ist zu niedrig, wohingegen die Lichtgeschwindigkeit für die menschliche Wahrnehmung viel zu hoch ist. Der Mensch kann nur das Ergebnis, nicht aber den Prozess an sich wahrnehmen. Um dies doch zu erreichen muss die Wahrnehmung mit technischen Hilfsmitteln erweitert werden. Die menschliche Wahrnehmung alleine spiegelt also nicht die gesamte, objektive Wirklichkeit wider. Von der Erde aus wahrgenommen wirkt es so auch, als ob sich die Sonne um die Erde dreht und die Erde der Mittelpunkt des Sonnensystems sei. Die Sonne geht im Osten auf und im Westen unter. Betrachtet man dieses System aber aus einem anderen Punkt oder nutzt Möglichkeiten der Welt 3, beispielsweise Geometrie und Mathematik, kann man erkennen, dass nicht die Sonne sich um die Erde dreht, sondern die Erde um die Sonne. Die geistigen Produkte aus Welt 3 erlauben es dem Menschen seine Möglichkeiten der Wahrnehmung zu vergrößern. Er ist in der Lage Thesen über die Wirklichkeit zu formulieren, diese mit anderen zu diskutieren, sie mit Experimenten und Erfahrungen abzugleichen und daraus neue Erkenntnisse über die Wirklichkeit zu gewinnen.⁵¹

Der epistemologische Aspekt beschreibt die Frage, in welcher Beziehung sich Erkenntnis und Objekt der Erkenntnis befindet. Die wissenschaftliche Diskussion unterscheidet einen epistemologischen Realismus, der eine objektive Erkenntnis einer unabhängigen Wirklichkeit vertritt, und den Konstruktivismus, der von einer subjektiven Erkenntnis ausgeht. Im Konstruktivismus wird die Beziehung zwischen Erkenntnis und dem Objekt der Erkenntnis durch den kognitiven Agenten konstruiert.⁵²

Die Erkenntnisfähigkeit des Menschen wird durch die Möglichkeiten, die das IR bietet, erweitert. Die Erkenntnisse, die aus der Welt von anderen Agenten gesammelt werden können, werden z.B. in Büchern niedergeschrieben. Dadurch entsteht eine Möglichkeit gesammeltes

⁵⁰Vgl. WERSIG: Information-Kommunikation-Dokumentation. Ein Beitrag zur Orientierung der Informations- und Dokumentationswissenschaft, S.54-74.

⁵¹Vgl. WUCHTERL: Lehrbuch der Philosophie, S.72-78.

⁵²Vgl. HERMANS: Ontologiebasiertes Information Retrieval für das Wissensmanagement, S.4.

Wissen weiter an andere kognitive Agenten zu verteilen. Eine Verteilung von Wissen, das nicht immer wieder neu erzeugt werden muss, erweitert auch die Erkenntnisfähigkeit jedes Individuums und somit auch des Kollektivs. Vorausgesetzt es existieren Strukturen, die Träger von Erkenntnissen beinhalten (beispielsweise ein Buch) und es besteht die Möglichkeit (An)Fragen zu formulieren, um in dieser Struktur erfolgreich nach relevanten 'Objekten der Erkenntnis' - sprich Dokumenten - suchen zu können. Aus diesem Blickwinkel betrachtet, ist Suchen ein Teilbereich wissenschaftlichen Arbeitens und wissenschaftliche Arbeit stets ein Weg zu neuer Erkenntnisse, die dem Lösen von Problemen dienen.

Der Kognitionswissenschaftler NORMAN skizziert ein Modell, das beschreibt, wie ein Akteur in der Wirklichkeit agiert. Er unterscheidet Handlungen (*execution*) und Evaluierung der Ergebnisse der Handlungen (*evaluation*). Nach jeder getätigten Aktion muss ein Akteur mit seiner Wahrnehmung feststellen, in wie weit seine Handlung die Wirklichkeit verändert hat. Den Unterschied zwischen der vorhergehenden Annahme über ein Ergebnis und die tatsächliche Wahrnehmung davon, bezeichnet NORMAN als '*gulf of execution*'. Als '*gulf of evaluation*' bezeichnet er die Herausforderung zu entscheiden, ob das gesetzte Ziel erreicht wurde oder nicht.⁵³ Je weniger ein Akteur von der Wirklichkeit weiß, desto weniger ist er in der Lage das zu erreichen, was er möchte. Je geringer seine Informationskompetenz⁵⁴ ist, also die Effektivität eingesetzter Methoden und Technologien zum Finden von Informationsressourcen, desto schlechter werden die Suchergebnisse sein. Schlechter bedeutet in diesem Zusammenhang, dass das Informationsbedürfnis nicht gestillt wird, oder nur scheinbar gestillt wurde, es nicht der Wirklichkeit entspricht und die Befriedigung des Informationsbedürfnis manipuliert wurde. Es stellt sich die Frage mit welchen Kriterien man festmachen kann, ob eine Erkenntnis sicheres und wahres Wissen ist, oder ob es unwahr ist. Damit verbunden stellt sich die weitere Frage, ob wahres Wissen überhaupt existieren kann und inwieweit IR als Teil wissenschaftlichen Arbeit mit diesem Problem konfrontiert ist. Sind nur 'wahre' Suchtreffer von Relevanz?

Wahrheit

"Erkenntnis ist Wahrheitssuche - die Suche nach objektiv wahren, erklärenden Theorien."⁵⁵ Überschneiden sich Wirklichkeit und Erkenntnis, so geht man von einer **Korrespondenztheorie der Wahrheit** aus. Wahrheit ist ein variabler Begriff, es gibt Dinge in der subjektiven Wahrnehmung von Menschen, die für den Einzelnen wahr sind, für den anderen aber schlichtweg falsch. Für ARISTOTELES galt eine Annahme dann als wahr, wenn eine Menge an Personen dieselbe Meinung vertritt, dass etwas wahr ist. Dies bezeichnet man als **Konsenstheorie der Wahrheit**. Thomas von AQUIN erweiterte diese Annahme dahingehend, dass etwas als allgemein bewiesen gelten muss. Dies führt zur **semantischen Theorie der Wahrheit**, in der eine Metasprache und Kriterien gefunden werden müssen, um eine Wahrheit argumentieren zu kön-

⁵³Vgl. NORMAN, Donald A: The design of everyday things: Revised and expanded edition. New York, 2013.

⁵⁴Fähigkeit mit Information effektiv umgehen zu können, Vgl. BALCERIS, Michael: Medien-und Informationskompetenz: Modellierung und Messung von Informationskompetenz bei Schülern. 2011.

⁵⁵POPPER: Auf der Suche nach einer besseren Welt: Vorträge und Aufsätze aus dreißig Jahren, S.12.

nen.⁵⁶ Im 19. Jahrhundert arbeiteten KANT und HEGEL weiter am Begriff der Wahrheit. In ihren Augen muss ein Beweis erbracht werden, der von der Allgemeinheit nach bestimmten Kriterien nachvollziehbar betrachtet werden kann und danach akzeptiert wird. Aufbauend auf HEGEL wurde im dialektischen Materialismus ein Spannungsfeld aus absoluter Wahrheit und relativer Wahrheit aufgespannt. In diesem Verständnis ist absolute Wahrheit alles das, das als allgemeingültig akzeptiert ist und immer bewiesen werden kann. Ein Beispiel dafür ist die Existenz einer Kraft, die Körper nach unten, zum Erdinneren, zieht. Die Existenz dieser Kraft, der Erdanziehungskraft, kann zu jedem Zeitpunkt an jedem Ort einfach bewiesen werden. Relative Wahrheit ist eine Aktualisierung einer Annahme zu einer bestimmten Zeit, nachdem weitere Erkenntnisse gewonnen wurden.⁵⁷ So relativiert sich die Gravitationskraft im Makro- bzw. Mikrokosmos und verhält sich anders. Diese Erkenntnis lässt sich aber nicht ad hoc beweisen, sondern als fundiertes Ergebnis von Thesen und Experimenten.

Die zentrale These im kritischen Rationalismus geht davon aus, dass jede Erkenntnis einen hypothetischen Charakter aufweist und im Grunde fehlbar ist. Es gibt keine sichere Wahrheit und auch kein sicheres Wissen postuliert POPPER. Erst in der immer wiederkehrenden (selbst)kritischen Auseinandersetzung mit Thesen kann etwas, wie sichereres Wissen, entstehen. Erkenntnismonopole, nicht nur für die Wissenschaften, gehen laut POPPER immer auch mit autoritären Strukturen einher und sind deswegen abzulehnen.⁵⁸ Dennoch existieren einige Systeme, die eine absolute Wahrheit implizieren. Dabei handelt es sich beispielsweise um Ideologien, Religionen oder andere gesellschaftliche, wie ökonomische Strukturen, auf die in Kapitel 2.3 näher eingegangen wird.

"*Wissenschaft ist Wahrheitssuche*"⁵⁹ schreibt POPPER weiters. Es ist nicht von zentraler Bedeutung, dass eine Suchanfrage ein absolutes Ergebnis liefert, sondern, dass der Weg zu diesem Ergebnis (oder keinem Ergebnis) bereits ein zentrales Moment wissenschaftlicher Arbeit ist. IR-Systeme, in all ihren unterschiedlichen Formen, sind essentieller Teilbestand jeder Wissenschaft. Das Suchen nach Fragen und Antworten ist der Wissenschaft immanent und essentiell für effektive Wissenschaft, die sich den Problemen der Menschen stellt und dabei nicht außer Acht lässt ihre Ergebnisse nachvollziehbar zu halten. Dies gilt auf einer theoretischen, philosophischen Ebene, aber auch auf einer praktischen.

Als Wort des Jahres 2016 wurde das Adjektiv '*postfaktisch*' zum Sieger gekürt und im gleichen Jahr die Formulierung '*alternative Fakten*' strapaziert. Begriffe, die veranschaulichen, wie intensiv sich die gesellschaftliche Diskussion in dieser Zeit damit beschäftigt hat, wie wahr oder falsch Information im World Wide Web sein kann und welche Problemstellungen damit einhergehen. Es ist nicht wichtig wer eine auf Fakten basierte Diskussion führt, sondern wer medial

⁵⁶Vgl. HERMANS: Ontologiebasiertes Information Retrieval für das Wissensmanagement, S.4.

⁵⁷Vgl. POPPER: Auf der Suche nach einer besseren Welt: Vorträge und Aufsätze aus dreißig Jahren, S.43-50.

⁵⁸Vgl. SALAMUN: Was ist Philosophie?, S.203-217.

⁵⁹POPPER: Auf der Suche nach einer besseren Welt: Vorträge und Aufsätze aus dreißig Jahren, 50.

so auftreten kann, um seine politischen und ideologischen Ideen durchsetzen zu können.⁶⁰ Das Web ermöglicht es, alles, ob wahr oder falsch, einer breiteren Öffentlichkeit zugänglich zu machen.

Zwei zentrale Fragen bzw. Problemstellungen, die im Zuge dieser Arbeit nicht gelöst werden können, werfen sich auf:

- Welche Werkzeuge und Kriterien kann es geben, um Akteuren im Web eine Gewissheit darüber zu geben, welche gefundenen Informationsressourcen wahr und welche falsch sind? Im Semantic Web Stack in Abbildung 18, im letzten Kapitel der Arbeit, beschreiben die Bereiche *Proof* und *Trust* die Problemstellungen im Web, wie Inhalte vertrauenswürdig, richtig und 'wahr' zwischen Menschen und Maschinen, aber auch Menschen und anderen Menschen geteilt werden können.
- Inwieweit manipulieren IR-Systeme (z.B. Google) oder andere Akteure die gefilterte Wirklichkeit und liefern falsche, inkonsistente und verzerrende Antworten auf Suchanfragen und welche Lösungen kann es diesbezüglich geben? Diese Problemstellungen betreffen auch sämtliche Retrieval-Systeme und Prozesse im Kontext wissenschaftlicher Arbeit.

⁶⁰STEPHAN Felix, Zeitungsartikel, <http://www.zeit.de/kultur/2016-12/postfaktisch-wort-des-jahres-post-truth-demokratie-jill-lepore>, 22.12.2016.

2.3 Gesellschaftliche Perspektive

Paul MASON stellt in seinem Werk *Postkapitalismus* die kontroverse These auf, dass die Anpassungsfähigkeit des wirtschaftlichen Systems des Kapitalismus nicht mehr aufrecht zu halten ist. Dies begründet er mit der Tatsache, dass sich eine historisch einmalige Entwicklung in den letzten 25 Jahren etabliert hat: die Informationsgesellschaft. Die Verteilung und Vernetzung von Information über das World Wide Web und die starke Anbindung der Informations- und Kommunikationstechnologien an die Wirtschaft und die Lebenswelt der Menschen. Diese sogenannte 'Informationsrevolution', die Entwicklung von der Industrie- zur Informationsgesellschaft,⁶¹ hat drei folgenreiche Auswirkungen:

1. Informationstechnologien haben, und werden weiter, den Arbeitsaufwand in vielen Bereichen verringern und zu einem Aufbrechen der Beziehungen zwischen Arbeit und Freizeit, sowie Arbeit und Einkommen führen.
2. Märkte beruhen auf den Prinzipien Knappheit, Nachfrage und Angebot. Informationsgüter, die im Überfluss existieren und theoretisch beliebig vielfältig und verteilbar sind, zerstören die grundlegenden Prinzipien des Marktes, wie etwa die Preisbildung.
3. Es haben sich Strukturen gebildet, die nicht in das kapitalistische Verständnis passen, da sie nicht den Prinzipien des Marktes gehorchen. Diese Strukturen, beispielsweise *Wikipedia.org*, die *Open Access* oder *Open Source* Bewegungen, sind trotzdem erfolgreich⁶²

Dies soll keine ökonomische oder politische Arbeit sein oder gar ein politisches Statement. Die Beobachtung von MASON jedoch, dass Information in ihrer Form als Informationsgut,⁶³ einen essentiellen Stellenwert in unserer Ökonomie und Gesellschaft hat, ist nachvollziehbar. Der Begriff **Informationsökonomie** fasst dies zusammen und beschreibt einen volkswirtschaftlichen vierten Sektor, in dem die zentrale Ressource, die Information ist.⁶⁴

Auch der, in der aktuellen politischen Diskussion verwendete Begriff, 'Industrie 4.0', zielt in diese Richtung ab und beschreibt eine 4. industrielle Revolution, in der ganze Wertschöpfungsketten automatisiert ablaufen können. Produktionsprozesse werden durch die selbstständige Kommunikation zwischen Maschinen optimiert. BAUERNHANSL sieht in den Informations- und Kommunikationstechnologien, wie auch MASON, Schlüsseltechnologien, die es ermöglichen Produktionsfaktoren so zu wenden, damit ein Energie und Ressourcen nachhaltiges Wirtschaften auch in Zukunft möglich sein kann. Beispielsweise sind *Smart Grids*, intelligente Stromnetze, die selbständig Informationen über Stromverbrauch und Stromerzeugung austauschen und effizient verteilen, eine Technologie, um die Stromversorgung zu optimieren und

⁶¹Vgl. PILLER: Einführung: Informationsrevolution und industrielle Produktion, S.1-2.

⁶²Vgl. MASON: Postkapitalismus: Grundrisse einer kommenden Ökonomie, S.14-16.

⁶³Gut als Mittel um Bedürfnis zu befrieden, [https://de.wikipedia.org/wiki/Gut_\(Wirtschaftswissenschaft\)](https://de.wikipedia.org/wiki/Gut_(Wirtschaftswissenschaft)), 20.09.2016

⁶⁴Informationsökonomik, Gabler Wirtschaftslexikon, <http://wirtschaftslexikon.gabler.de/Archiv/54725/informationsoekonomik-v1.html>, 09.01.2017.

Energie zu sparen.⁶⁵ Die automatisierte Kommunikation zwischen Maschinen, die dafür sorgt, dass die gebrauchte Information zur richtigen Zeit am richtigen Ort ist, bedarf zur Umsetzung Methoden des IR.

Neben dem soeben genannten ökonomischen Aspekt, versucht KUHLEN den zentralen Stellenwert von Information und Wissen in seinem Modell der **Wissensökologie**, in Hinblick auf ganzheitlich gesellschaftliche Aspekte zusammenzufassen:

*Wissensökologie kann [...] definiert werden als der nachhaltige Umgang mit Wissen und Information. [...] Wissen verbraucht und erschöpft sich nicht im Gebrauch. Ganz im Gegenteil, je mehr es genutzt wird, desto mehr nutzt es vielen Leuten, und desto höher ist die Wahrscheinlichkeit, dass sich aus Wissen Innovationen, neue Produkte und Dienstleistungen entwickeln und dass neues Wissen aus existierendem entsteht.*⁶⁶

Damit versucht KUHLEN die Wechselwirkung zwischen Mensch und technisierter Gesellschaft zu beschreiben und führt sieben wesentliche Perspektiven an, die diese Wissensökologie begründen. Gleichzeitig geben diese Punkte ein modellhaftes Bild davon, wie stark Information und Wissen die Gesellschaft beeinflussen, weiterentwickeln und neu gestalten.

Umweltschutz für das Internet

Menschen stehen immer mit Information und Wissen, wie mit der Natur, in Interaktion. In der heutigen Zeit, mittels mobiler Applikationen, noch viel stärker als je zuvor. Wie bei natürlichen Ressourcen, wie etwa Luft und Wasser, gilt auch für Information und Wissen, dass sie bei falscher Nutzung oder Übernutzung 'verschmutzen' und unbrauchbar werden. Ein alltägliches Beispiel dafür ist Spam im Email-Ordner. Systeme die Information verarbeiten oder verteilen müssen gepflegt und gewartet werden. Besonders wenn sie dem Anspruch gerecht werden wollen Information zu verteilen.

Informationsökologie

Diese Perspektive manifestiert sich durch Räume wie etwa der (digitalen) Bibliothek, als '*[...] eine Einrichtung, die unter archivarischen, ökonomischen und synoptischen Gesichtspunkten publizierte Information für die Benutzer sammelt, ordnet und verfügbar macht.*'⁶⁷ Eine Bibliothek ist ein Informations- und Wissensspeicher, der der Gesellschaft zur Weiterentwicklung zur Verfügung steht und so weitere Prozesse aktiviert, die wiederum neue Informationsressourcen erzeugen. Effektive Retrieval-Möglichkeiten, zum Beispiel in einer digitalen Bibliothek, die ohne Retrieval-Funktionalitäten nicht denkbar wäre, fördern diese Entwicklung.

⁶⁵Vgl. BAUERNHANSL/TEN HOMPEL/VOGEL-HEUSER: Industrie 4.0 in Produktion, Automatisierung und Logistik: Anwendung· Technologien· Migration, S.5-34.

⁶⁶KUHLEN: Wissensökonomie und Wissensökologie zusammen denken, S.1.

⁶⁷EWERT, Gisela/UMSTÄTTER, Walther: Lehrbuch der Bibliotheksverwaltung. Stuttgart, 1997, S.10.

Funktionale Perspektive

Eine gesellschaftliche Hauptaufgabe und so auch eine der Wissensökologie ist die Frage nach einem nachhaltigen Nutzen der natürlichen Ressourcen. Information und Wissen dienen dazu, die bestmögliche Nutzung von Ressourcen zu gewährleisten, in dem sie neue Technologien, Diskurse und Bildung fördern. Die zuvor genannten *Smart Grids* sind ein Beispiel dafür. KUHLEN postuliert in diesem Zusammenhang, dass eine nachhaltige Entwicklung nur mit einem freien Zugriff auf Wissen erreicht werden kann. Ein uneingeschränkter Zugang steht aber der bestehenden Marktdominanz von Gütern gegenübersteht, ja schließt sie sogar aus.

Kommunikationsökologische Perspektive

Kommunikationsökologie beschäftigt sich mit der durchdringenden Wechselwirkung der Kommunikation zwischen Mensch(en), Kultur(en) und Gesellschaft(en). Das Kommunikationsverhalten selbst verändert sich stetig mit den verwendeten Kommunikationsmitteln. Die Kommunikation über eine Briefkorrespondenz verläuft anders, und zwar auf einer technischen wie auch auf einer kommunikativen Ebene, als ein synchroner Kommunikationsprozess mittels Instant-Messenger-Dienst über ein Smartphone. Aus dieser Perspektive heraus verändern sich auch Retrieval-Prozesse im Wandel technischer Entwicklungen. Die Suche nach einem Buch mittels Zettelkatalog unterscheidet sich von der Suche nach Büchern über eine Literatursuchmaschine wie beispielsweise *Google Scholar*,⁶⁸ das zur Literaturrecherche wissenschaftlicher Dokumente dient.

Zukunftsethische Perspektive

Eine weitere Frage, die das Konzept der Wissensökologie aufwirft, ist die Frage nach der Langzeitverfügbarkeit von Wissen für zukünftige Generationen. KUHLEN sieht es als Pflicht der lebenden Generationen, ihr Wissen den kommenden Generationen effektiv zur Verfügung zu stellen. Dafür bedarf es entsprechender Informations-, Dokumentations- und Langzeitarchivierungssysteme, die ein geeignetes Rüstzeug für IR in sich tragen. Kapitel 5.3.1 widmet sich der theoretischen und modellhaften Zugängen, wie digitale Langzeitarchivierung von Informationsressourcen umgesetzt werden kann.

Ökosoziale Perspektive

KUHLEN versteht Wissensökologie als Beitrag zur einer globalisierten, ökosozialen Marktwirtschaft, die nicht nur virtuelle Güter oder Dienstleistungen beinhaltet, sondern auch die realen Ressourcen. Ziel soll es sein Information und Wissen gerecht und fair, sowohl in ökologischer, wie auch in sozialer Sicht, zu nutzen, ohne an der Doktrin einer 'Ressourcen-Inanspruchnahme' und der Verknappung der Güter festzuhalten.

⁶⁸Google Scholar, <https://scholar.google.at>, 05.03.2017.

Wissensökologische Perspektive

Immaterielle Güter werden immer wichtiger. Eine Gesellschaft, die freies, offenes und nachhaltiges Wirtschaften mit Wissen ermöglicht hat einen Vorteil gegenüber stark regulierten Märkten. Eine zu starke Regulierung, beispielsweise wie das Urheberrecht, geht mit einer Begrenzung innovativer Arbeit einher.⁶⁹

Politische Dimension

Information und Wissen werden von IR-Systemen in der Gesellschaft verteilt. Ihr Stellenwert ist genau so wichtig wie die Informationsgüter selbst. Um nachhaltig, kollaborativ, ökologisch und ökonomisch mit diesen Gütern umzugehen, bedarf es Systemen, die es den Akteuren erlauben Information oder formalisiertes Wissen zu suchen und zu entdecken. Das Ziel dabei ist die Annäherung an ein Idealbild der Gesellschaft, das mit Hilfe nachhaltiger und offener Wissensstrukturen - wie etwa der Wissensökologie - die ökonomischen, politischen, sozialen und kulturelle Entwicklungen vorantreibt. Die These von KUHLEN kann auf folgenden gemeinsamen Nenner gebracht werden: Je freizügiger der Umgang mit Wissen und Information ist, desto innovativer ist die politische Struktur, die Wirtschaft und die Wissenschaft und desto 'besser' kann sich eine Gesellschaft weiterentwickeln.

Ein Paradebeispiel dieser sogenannten Informationsgesellschaft ist der **Suchmaschinenanbieter Google**. Suchmaschinen sind Informationssysteme, die Daten unbekannter Eigenschaften verwalten, die einem Suchenden zumindest prinzipiell bekannt sind.⁷⁰ Gleichzeitig ist der kommerzielle Konzern Google ein Beispiel das einem Ideal entgegensteht. LEWANDOWSKI führt die dominierende Position von Suchmaschinen im Nutzerverhalten und in technischer Dimension an. Sie sind nicht nur Retrieval-Systeme, sondern auch der Einstiegspunkt zum World Wide Web, schlechthin. Aus diesem Grund ist es notwendig, die wechselseitige Beziehung von technischen, wie auch gesellschaftlichen Aspekten zusammen zu denken, um Suchmaschinen verstehen zu können. Die massenhafte Nutzung von Suchmaschinen, die Monopolstellung von Google und das 'Vertrauen', dass besser gereichte Suchtreffer, die 'wahren' Ergebnisse sind, positionieren Suchmaschinen an der Spitze der Informations- und Wissensverteilung.⁷¹

Für KUHLEN ist Google ein erstes, aber nicht zufriedenstellendes Ergebnis, das erfolgreiche unternehmerische Tätigkeit mit Aspekten der Wissensökologie verbunden hat. Auf der einen Seite ist der Zugang zu unzählbaren Informationsressourcen kostenlos, auf der anderen Seite ist Google einer der größten Konzerne der Welt, der nicht gesellschaftliche, sondern unternehmerische Ziele verfolgt.⁷² Unternehmerische Ziele müssen nicht mit gesellschaftlichen Zielen

⁶⁹Vgl. KUHLEN, Rainer: Erfolgreiches Scheitern — eine Götterdämmerung des Urheberrechts? In Schriften zur Informationswissenschaft Band 48 Herausgegeben vom Hochschulverband für Informationswissenschaft (HI) eV Konstanz., S.433-449.

⁷⁰Vgl. STOCK: Information retrieval: Informationen suchen und finden, S.6.

⁷¹Vgl. LEWANDOWSKI, Dirk: Suchmaschinen verstehen. Berlin and Heidelberg, 2015, S.1-5.

⁷²Vgl. KUHLEN: Erfolgreiches Scheitern — eine Götterdämmerung des Urheberrechts?, S.449-451.

einhergehen, oder können diesen sogar gegenüberstehen.

Google ist in der Lage politischen und ökonomischen Nutzen aus dem Wirtschaftsgut Information zu ziehen. Die Ergebnisse einer jeden Suchmaschine, die einer Nutzerin angezeigt werden, sind nicht neutral. GOLDMAN führt das Phänomen der *'search engine bias'* an. Darunter versteht man die Kontrolle darüber, welche Suchtreffer einer Nutzerin angeboten werden, ihre Reihung und welche Suchtreffer weggelassen werden und welche weiteren IR-Schritte vorgenommen werden können.⁷³ Zum einen hat das positive Effekte, da versucht wird Relevantes und nicht Relevantes zu trennen und Suchprozesse zu verbessern, zum anderen hat Google Möglichkeiten Inhalte zu bestimmen. Google verfügt über ein Erkenntnismonopol und entscheidet was gefunden wird und was nicht. Daraus lässt sich ein gewisses Machtpotenzial erkennen. Die Berechnung der Relevanz und somit dem Ranking einer Informationsressource erfolgt bei Google mittels mehreren Algorithmen. Einer davon ist der **PageRank**-Algorithmus. Auf Basis der Verlinkung einer Webseite zu anderen Webseiten und vice versa, wird ein Wahrscheinlichkeitswert berechnet, der beschreibt, wie relevant eine Webseite ist. Dieser Wert hilft dabei Informationsressourcen im Web vor zu sortieren und beliebten Seiten eine höhere Relevanz zu geben. Dadurch werden sie in der Trefferliste weiter oben anzuführen, als unbekanntere Seiten.⁷⁴ Zwar ist hier ein kollektiver Prozess im Gange, da mit der Beteiligung mehrere Individuen die beliebtesten Seiten besser erfasst werden können, jedoch wird nach dem Prinzip *'the rich get richer and the poor get poorer'* agiert und beliebte Seite stets höher gewertet als andere. Aus dieser Tatsache heraus führt LOBET-MARIS an, dass dieser gefilterter Zugang zu Information und zu Wissen, drei wechselseitige Beziehungen bedingen:

*'[...] für die Bedeutung, da sie eine bestimmte Ordnung der Welt implizieren, für die Macht, da sie eine bestimmte Machtverteilung unter den Informations-Akteuren nach sich ziehen, und für die Normen, da sie Wohlverhalten bei der Nutzung und eine entsprechende Einstellung mit einem guten Index und einem guten Ranking belohnen.'*⁷⁵

Der Erfolg von Google ist auf Grund seiner einfachen und effektiven Bedienung zu erklären, das nahezu ein Design-Paradigma für Retrieval-Systeme festgelegt hat. Auf Grund der hohen Nutzerzahlen bestimmt Google, welche Informationsressourcen gefunden, oder eben nicht gefunden werden. Dieses Machtpotenzial erlaubt es eine bestimmte Ordnung der Welt zu implizieren.

⁷³Vgl. SPINK/ZIMMER: Web search: Multidisciplinary perspectives, S.121.

⁷⁴Vgl. STOCK: Information retrieval: Informationen suchen und finden, S.555.

⁷⁵Vgl. LOBET-MARIS Claire, Vom Vertrauen zur Spurenauswertung. Eine neue Sicht der Technikfolgenabschätzung, S.85-97 In: BECKER/STALDER: DEEP SEARCH: Politik des Suchens jenseits von Google, S.90-94.

2.4 Fazit

MASON's Kapitalismuskritik auf Basis des Gutes Information, die von KUHLEN definierte Wissensökologie und Google als Paradebeispiel eines kommerziellen Unternehmens sind Beispiele dafür, wie stark Information und Wissen und somit auch IR die Lebenswelt der Menschen berührt. Wissenschaft, Wirtschaft und Gesellschaft bedienen sich tagtäglich dem Werkzeug IR um Daten, Information und Wissen zu verteilen, zu strukturieren und zugänglich zu machen. Wer eine führende Rolle an der Verteilung von Information und Wissen hat, der verfügt über eine Machtposition, sowohl im ökonomischen, wie auch im politischen Sinn.

POPPER's Erkenntnistheorie spiegelt die Vernetzung von Philosophie und Gesellschaft wider. Mit seiner Grundidee einer kritischen Auseinandersetzung mit Information und Ideen und das Wahrheit nur dann existieren kann, wenn sie auch angezweifelt werden kann, liefert er ein wichtiges Instrumentarium für die Wissenschaft, aber auch für die Gesellschaft. Die Formulierung einer These, die gar nicht widerlegt werden kann, ist eine unbrauchbare These. Anstelle von alleiniger, absoluter und statischer 'Pseudo-Wahrheit', die mit autoritären Strukturen einhergeht, sieht POPPER im 'ständigen Ringen' um Wahrheit die Möglichkeit des Erkenntnisgewinns. Dieses 'Ringen um Wahrheit' ist mit einem kritischen Suchen nach Dokumenten, Information und Wahrheit gleichzusetzen. Somit ist die Auseinandersetzung damit, wie solches Suchen ermöglicht oder verbessert werden kann, wichtiger Bestandteil der Gesellschaft und wird in den folgenden Kapitel theoretisch abgehandelt und im letzten Kapitel beispielhaft gezeigt.

3 Suchen: Information Retrieval

3.1 Einleitung

Information Retrieval beschreibt das effektive Suchen, Speichern, Organisieren, Wiederfinden und zugänglich Machen aller relevanten Informationsressourcen, die das Informationsbedürfnis eines kognitiven Agenten, hinsichtlich eines spezifischen Informationsbedarfs stillen.⁷⁶

Dieses Kapitel verfolgt zwei Ziele: Zum einen die Grundbegriffe der Informationswissenschaft, zum anderen die des IR zu definieren. Die Begriffe *Signal*, *Daten*, *Information*, sowie *Wissen* und ihre Beziehung zueinander und zum Retrieval-Prozess werden im Kapitel 3.2 weiter erörtert. Ausgehend von diesen Definitionen und vom Modell des Retrieval-Prozesses in Abbildung 1 werden die Fachbegriffe des IR in Kapitel 3.3, am Beispiel der digital edierten Karteikarten der digitalen Sammlung des Hans Gross Kriminalmuseum,⁷⁷ geklärt. In diesem theoretischen Beispiel stellt ein Kognitiver Agent eine Suchanfrage auf Grund eines Informationsbedürfnisses an einen Informationskorpus und erhält darauf eine Antwort, die vom Agenten weiter verarbeitet werden kann. Jedem dieser Bereiche wird ein Unterkapitel gewidmet, in dem die dazugehörigen Begriffe erörtert werden. Insgesamt ergibt sich dadurch eine Übersicht der Grundlagen, sowohl für die Ausführungen zum RD in Kapitel 4, als auch zur Umsetzung solcher Prinzipien in Kapitel 5.

3.2 Grundbegriffe der Informationswissenschaft

Wer sich mit der Thematik des Suchens beschäftigt, bewegt sich sogleich in der Disziplin der Informationswissenschaft, die STOCK wie folgt definiert:

*"Informationswissenschaft untersucht das Auswerten, Bereitstellen, Suchen und Finden von relevantem [...] Wissen durch Informations- und Kommunikationsprozesse."*⁷⁸

Für KUHLEN ist Informationswissenschaft mehr als die angeführte Definition. Die Entwicklung der Definition des Faches Informationswissenschaft beruht auf dem Verständnis des Begriffes Information, der unterschiedlich aufgefasst werden kann. Informationswissenschaft geht über das Wiederfinden von Informationsressourcen hinaus und versucht Ursachen und Wirkungen im Kontext von 'Information' zu beschreiben und zu erklären. Sogar von Informationswissenschaft *'als eine Wissenschaft vom Verstehen, als 'Geisteswissenschaft' im besten Sinne'*⁷⁹ ist die Rede. Gesellschaftliche, rechtliche, ökonomische, sowie Fragen zur Telemediatisierung - dem Zusammenspiel von Informatik, Telekommunikation und Hypermedia - und ihre Auswir-

⁷⁶Vgl. POETZSCH: Information Retrieval: Einführung in Grundlagen und Methoden, S.13-20.

⁷⁷Siehe Kapitel 5 zur weiteren Erklärung der Karteikarten, die die zentrale Datenquelle in der Projektumsetzung darstellen.

⁷⁸STOCK: Information retrieval: Informationen suchen und finden, S.4.

⁷⁹KUHLEN: Information - Informationswissenschaft, S.14.

kungen, sind Themenbereiche des Faches.⁸⁰ Zum Teil wurden diese Aspekte im vorangegangenen Kapitel skizziert.

In diesem Zusammenhang ist es notwendig die Begriffe *Daten*, *Information* und *Wissen*, als Grundbegriffe der Informationswissenschaft, festzumachen. Für diese Begrifflichkeiten gibt es eine Vielzahl unterschiedlicher Definitionen und Zugänge, über welche dieses Kapitel einen Überblick geben soll.

3.2.1 Der Ausgangspunkt: Signal und Daten

Die vier Begriffe - *Signal*, *Daten*, *Information* und *Wissen* - die stark ineinander greifen, unterscheiden sich durch ihren Grad an Strukturierung und ihre Gebundenheit an einen kognitiven Agenten, wie Abbildung 2 veranschaulichen soll. Dabei erhöht sich mit jedem Schritt die Wertigkeit des Begriffes. Desto komplexer die Strukturen werden, desto stärker ist die Bindung an den kognitiven Agenten, der diese Strukturen (mit)konstruiert. Signale sind simpel und können losgelöst ohne einen kognitiven Agenten existieren, wohingegen der Inhalt (die Information) in einem Buch erst bei der Leserin zu Wissen wird. Der Informationsbegriff ist aber gegenüber dem Wissensbegriff nicht abwertend zu verstehen. Trotzdem ist es notwendig die essentiellen Unterscheidungsmerkmale hervorzuheben.⁸¹

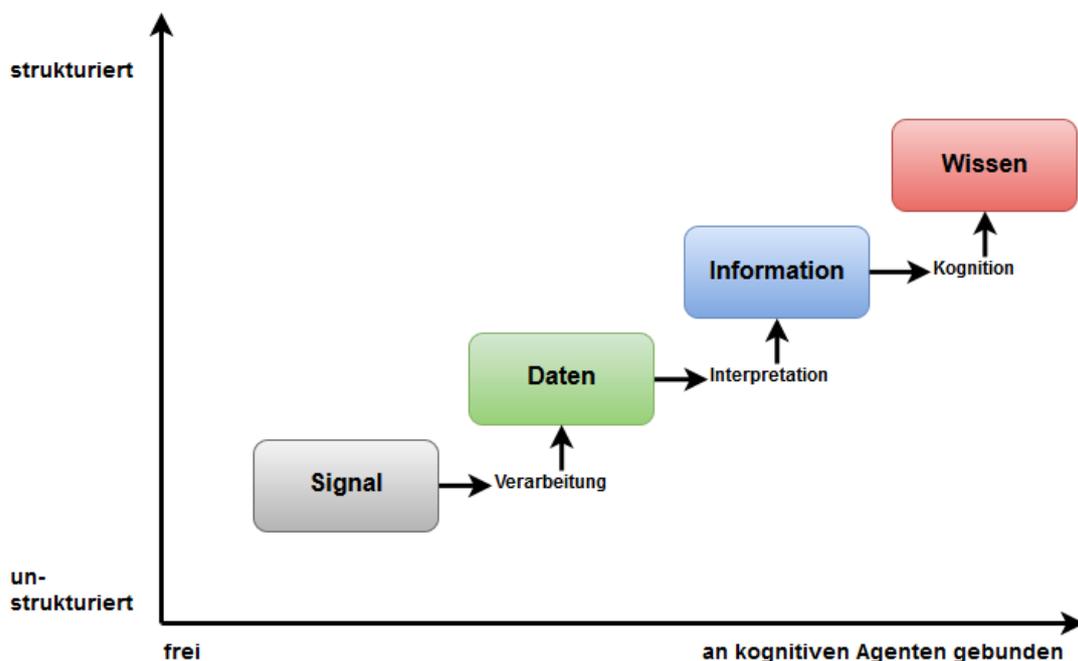


Abbildung 2: Signal, Daten, Information und Wissen

Signale sind Zeichen, die unabhängig von physikalische Entitäten, von einem Sender an einen Empfänger übermittelt werden. WERSIG definiert Signal als

⁸⁰Vgl. KUHLEN: Information - Informationswissenschaft, S.5-8.

⁸¹Vgl. BALLOD, Matthias: Informationen und Wissen im Griff: effektiv informieren und effizient kommunizieren. W. Bertelsmann Verlag, 2011, S.8-9.

*'ein zeitabhängiges, physikalisches Ereignis, das von einer Signalquelle ausgesandt oder von einem Signalempfänger empfangen wird'*⁸².

KARRENBERG definiert Signale als (physikalische) Schwingungen oder Wellen, die Information transportieren. 'Natürliche' Signale liegen in analoger Form vor und werden in der Messung durch Sensoren in elektrische Signale umgewandelt.⁸³ Eine Definition des Begriffes **Zeichen**, als Stellvertreter eines Objektes, beinhaltet die Existenz eines Bezeichners und eines Bezeichneten.⁸⁴ Bezeichner einigen sich darauf, dass die Anordnung von Druckerschwärze nach einem bestimmten Muster auf einem Papier Zeichen darstellt und etwas aus der Wirklichkeit, das Bezeichnete, stellvertretend repräsentiert. Sie können hinsichtlich ihrer Beziehung zu anderen Zeichen (Syntaktik), ihrer Beziehung zu bezeichneten Dingen (Semantik) und ihrer Beziehung zu ihren Benutzern (Pragmatik) betrachtet werden. Hier ist nicht automatisch ein Informationsgehalt vorhanden, sondern es gibt Stellen am Papier die schwarz sind und welche die weiß sind. Dieser Unterschied kann von einem Empfänger verarbeitet werden. Das Schema der Signalübertragung nach SHANNON⁸⁵ beschreibt genau diesen Vorgang, dass eine Quelle bzw. ein Sender über einen Kanal Zeichen an einen Empfänger überträgt. Dabei ist nicht relevant, ob die empfangenen Zeichen sinnvoll sind, sondern nur, dass diese richtig bei einem Empfänger ankommen. FAVRE-BULLE bezeichnet dies als **nachrichtentechnisches Informationsparadigma**.

Auf diesem Kanal können Signale gestört werden. Dies wird als **Rauschen** bezeichnet und verzerrt oder zerstört den Informationsgehalt übertragender Zeichen bzw. Signale und sollte im Sinne der Nachrichtentechnik vermieden werden.⁸⁶ Um Signale richtig aufnehmen zu können, müssen Sender und Empfänger über einen gemeinsamen Satz an Zeichen verfügen und in der Lage sein, die Signale beim Senden zu kodieren, und beim Empfangen zu dekodieren. 'Natürliche Daten' oder 'natürliche Information', wie sie in der analogen Wirklichkeit vorkommen, müssen, um sie im Computer verarbeiten zu können, digitalisiert werden. Musik beispielsweise, charakterisiert durch Tonhöhe, Klang oder Ähnlichem, muss dafür als Bitstrom⁸⁷ repräsentiert werden. **Kodierung** beschreibt den Vorgang der Umwandlung der 'natürlichen Information' in einen Bitstrom, der von einer Maschine interpretiert werden kann. Die **Dekodierung** ist die

⁸²WERSIG: Information-Kommunikation-Dokumentation. Ein Beitrag zur Orientierung der Informations- und Dokumentationswissenschaft, S.78.

⁸³Vgl. KARRENBERG, Ulrich: Signale-Prozesse-Systeme: eine multimediale und interaktive Einführung in die Signalverarbeitung. Springer-Verlag, 2016, S.24.

⁸⁴Vgl. FAVRE-BULLE: Information und Zusammenhang: Informationsfluß in Prozessen der Wahrnehmung, des Denkens und der Kommunikation, S.41-43.

⁸⁵Siehe auch SHANNON, C. E.: A mathematical theory of communication. The Bell System Technical Journal, Vol. 27, 1948.

⁸⁶Vgl. KARRENBERG: Signale-Prozesse-Systeme: eine multimediale und interaktive Einführung in die Signalverarbeitung, S.58-60.

⁸⁷Ein Bitstrom ist eine Folge von Nullen und Einsen, die von einer Maschine interpretiert werden kann, <https://en.wikipedia.org/wiki/Bitstream>, 02.03.2017.

Rückwandlung der Musik von einem Bitstrom in Schallwellen, die dann wieder von einem kognitiven Agenten wahrgenommen werden können. Bei der Kodierung von Information kann auch eine verlustbehaftete und eine verlustfreie Kompression erfolgen, um den notwendigen Speicherplatz zu verringern.⁸⁸ Darstellung 3 veranschaulicht SHANNON's Modell, in der ein Sender 's' aus seinem Kontext heraus einen Inhalt codiert und über einen Kanal an einen Empfänger 'e' übermittelt, der wiederum das Signal decodiert und auf Grundlage seines Kontextes (alle Erfahrungen und das gesamte Wissen) in Information umwandelt.

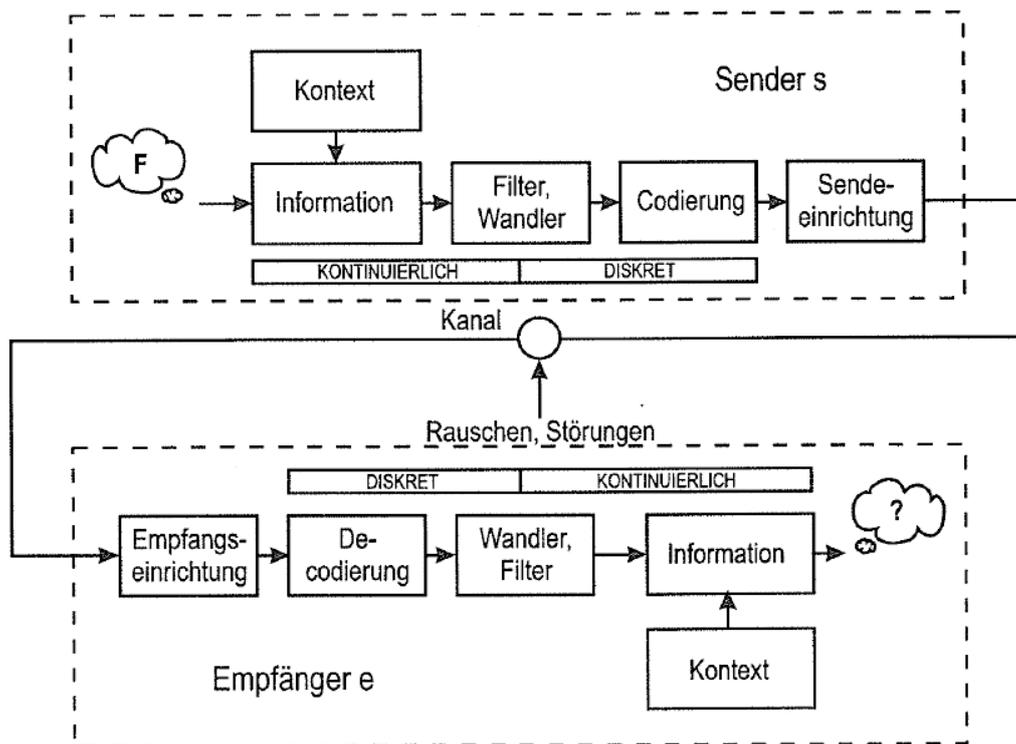


Abbildung 3: Modell der Informationskette für codierte Signalübertragung

Im Verständnis historischer Texte als Nachricht eines historischen Senders an einen gegenwärtigen Rezipienten führt THALLER zwei Strömungen im Diskurs der hier erörterten Begriffe an.

*'That data – as transmitted in signal streams – and information, as handled by humans, are not identical is a truism. They have long been seen as separate strata in information theory. A main difference between Shannon and the 'data – information – knowledge – wisdom' hierarchy has always been, that the former leads directly to an intuitive understanding of systems which can be realized by software engineering, while the later cannot.'*⁸⁹

⁸⁸Vgl. KARRENBERG: Signale-Prozesse-Systeme: eine multimediale und interaktive Einführung in die Signalverarbeitung, S.345-348.

⁸⁹THALLER, Manfred: What is a text within the Digital Humanities, or some of them, at least? dh2012 - Book of Abstracts, 2012 (URL: <http://www.dh2012.uni-hamburg.de/conference/programme/abstracts/beyond-embedded-markup>), 29.01-2017.

THALLER versteht die unterschiedliche Zugänge Daten bzw. Information zu definieren, führt aber auch an, dass die Hierarchie Daten-Information-Wissen noch nicht direkt zur Umsetzung von Informationsarchitekturen geführt hat.

Von **Daten** könnte man sprechen, wenn Signale nach einem bestimmten und geregelten Muster verarbeitet werden. Werden die Signale - Druckerschwärze auf Papier - in ihrer spezifischen Form und Anordnung erkannt, so kann der Empfänger darin Buchstaben und Wörter erkennen, wenn diese im lateinischen Alphabet kodiert wurden und Sender, wie auch Empfänger, diese Kodierung entschlüsseln können. Bei Daten befindet man sich auf der syntaktischen Ebene.

Ist der Brief zwar im lateinischen Alphabet, aber in einer Sprache verfasst, die der Empfänger nicht versteht, wurden die Signale in Daten verarbeitet und erkannt, aber kein Informationsgewinn erzielt. Hier kommt die semantische Ebene ins Spiel. Bei FAVRE-BULLE findet sich der Begriff **semantisches Informationsparadigma**, indem die Übertragung des Inhalts einer Nachricht von Bedeutung ist.⁹⁰ Erst wenn die Semantik, die Zuordnung von Bedeutung zu den syntaktischen Gerüsten, existiert, kann **Information** übertragen werden. Diese Zuordnung zu einer Bedeutung ist eine Interpretation des Empfängers, basierend auf den erhaltenen und entschlüsselten Daten.⁹¹

3.2.2 Der Informationsbegriff

GAUS führt sieben Definitionsvorschläge für den Begriff Information an. Eine klare und eindeutig Definition zu diesem vielschichtigen und im Alltag unterschiedlich verwendeten Begriff existiert nicht und erscheint auf Grund der Breite des Begriffes auch nicht sinnvoll zu sein. Diese Definitionsvorschläge berühren unterschiedliche Disziplinen, die jeweils mit ihrem eigenständigen Informationsbegriff arbeiten. GAUS sieht folgende Möglichkeiten eine Definition zu formulieren:⁹²

Information als Struktur

Erst dadurch, dass die Dinge in der Welt in einer gewissen Verbindung zueinander existieren, also eine Struktur aufweisen, erhalten sie einen Informationsgehalt. WERSIG beispielsweise verfolgt einen solchen Zugang mit seinem kybernetischen Informationsbegriff. Dabei strukturiert ein Organismus bzw. kognitiver Agent seine Wahrnehmungen der Außenwelt, die er mit seinen Rezeptoren aufnehmen kann. Erhält der Agent dadurch Information spricht WERSIG von Information im engeren Sinne als '*Reduktion von Ungewissheit aufgrund von Kommunikationsprozessen*'.⁹³

⁹⁰Vgl. FAVRE-BULLE: Information und Zusammenhang: Informationsfluß in Prozessen der Wahrnehmung, des Denkens und der Kommunikation, S.18-32.

⁹¹Vgl. STOCK: Information retrieval: Informationen suchen und finden, S.17-19.

⁹²Vgl. GAUS: Dokumentations-und Ordnungslehre: Theorie und Praxis des Information Retrieval, S.29-30.

⁹³WERSIG: Information-Kommunikation-Dokumentation. Ein Beitrag zur Orientierung der Informations- und Dokumentationswissenschaft, S.74.

Information als Erkenntnis

Es werden nur solche Strukturen als Information verstanden, die auch von einem kognitiven Agenten als solche wahrgenommen werden und zu Erkenntnis führen können, wie es im Kapitel 2.2 auch angeführt wird. Wahrnehmung ist im Sinne des Konstruktivismus von WATZLA-WICK vom jeweiligen Betrachter aus konstruiert.⁹⁴ Alles was unerkannt bleibt, also nicht Teil der wahrnehmbaren Wirklichkeit ist, ist für den jeweiligen kognitiven Agenten keine Information. Von Information spricht man aus dieser Perspektive nur, wenn ein kognitiver Agent die Strukturen auch als Information erkennt.

Information als Signal

Information wird als Signal von einem Sender an einen Empfänger übermittelt, wobei die angekommene Information gemessen werden kann. Dieser nachrichtentechnische Zugang zum Informationsbegriff fußt auf SHANNON's Modell der Signalübertragung, das in Kapitel 3.2.1 bereits angesprochen wurde.

Information als Nachricht

Als Nachricht versteht die Informationstheorie eine Information, die sich als Signal und Folge von Zeichen mit semantischem Wert definieren lässt. Das Signal 'rote Lampe' kann unterschiedliche und kontextabhängige semantische Werte besitzen. Beispielsweise um vor einer Gefahr zu warnen oder als Stoppsignal, wie bei einer roten Ampel.⁹⁵

Information als verstandene Nachricht

Der nachrichtentechnische Zugang ist reduziert. Die Kommunikation zwischen Menschen ist komplexer, da eine nicht verstandene Nachricht zu einer gescheiterten Informationsübertragung führt. Es müssen weitere Grundvoraussetzungen, wie etwa eine gemeinsame Sprache oder eine ähnliche Wissensbasis, aller an der Kommunikation beteiligten Akteure, erfüllt sein. Einen Vortrag in einer fremden Sprache kann man hören, aber nicht verstehen was er behandelt.⁹⁶ Die analogen und 'natürlichen' Daten in Form von Schallwellen wurden wahrgenommen, aber der kognitive Agent war nicht in der Lage, das Gehörte als eine Erkenntnis erzeugende Nachricht zu verstehen.

Information als Wissensvermehrung

Bei diesem Ansatz spricht man erst dann von Information, wenn die angekommene und verarbeitete Nachricht dazu führt, dass etwas, das vorher noch nicht gewusst wurde, nun gewusst

⁹⁴Vgl. VON GLASERSFELD, Ernst: Konstruktion der Wirklichkeit und der Begriff der Objektivität. In VON FOERSTER, Heinz (Hrsg.): Einführung in den Konstruktivismus. München, 1992, und Vgl. Radikaler Konstruktivismus, https://de.wikipedia.org/wiki/Radikaler_Konstruktivismus, 04.11.2016.

⁹⁵Vgl. FAVRE-BULLE: Information und Zusammenhang: Informationsfluß in Prozessen der Wahrnehmung, des Denkens und der Kommunikation, S.34-35.

⁹⁶Vgl. GAUS: Dokumentations- und Ordnungslehre: Theorie und Praxis des Information Retrieval, S.30.

wird. Es ist zu einem Informationszuwachs gekommen. KENDAL et CREEN sehen Information aus der Perspektive des Wissensmanagements im Dreischritt: Daten ('*The temperature is 5*'), die in einem kognitiven Prozess eines Agenten zu Information werden ('*It is cold outside*') und in einem pragmatischen Schritt zu Wissen werden ('*It is cold - put on a warm coat*').⁹⁷

Information als Vorgang

Hier steht die Informationsübertragung, -gewinnung, und -verarbeitung als Prozess im Mittelpunkt, wobei die Informatik die Disziplin darstellt, die sich mit der maschinellen Verarbeitung von Information beschäftigt. GUMM et SOMMER definieren den Begriff Information im Wechselspiel von Daten, die sie als Abstraktion von Information in der maschinenlesbaren Form von Nullen und Einsen festmachen. Eine Folge von Bits kann eine abstrahierte Darstellung für eine Maschine sein, die beispielsweise Text repräsentiert, wobei jede Verarbeitung an sich wiederum eine in Nullen und Einsen kodierte Information darstellt⁹⁸

Jede Definition ist aus ihrer Perspektive und Argumentationsweise heraus schlüssig und legitim. Zusammenfassend lässt sich sagen, dass Information das Ergebnis eines Interpretationsprozesses einer Nachricht durch einen kognitiven Agenten ist. Dabei ist ein Dualismus zwischen Behälter (der Syntax) und dem Inhalt (der Semantik) erkennbar.⁹⁹ Diesen Dualismus von Signal und Inhalt findet man auch bei STOCK. Er sieht nach der Definition von KUHLEN Information als *Wissen in Aktion*¹⁰⁰. Dabei ist das Wissen statisch und die Information etwas, das gebunden an ein Medium, in Bewegung ist. Bei BALLOD wiederum liest man:

*'Wissen ist immer an Prozesse eingebunden. Es wird generiert, beschafft, erfasst, vernetzt, verteilt, transferiert, gespeichert, aktualisiert oder vergessen. Wissen ist mehr flüchtiges Phänomen, denn statisches Objekt'*¹⁰¹

Man sieht, in der Fachliteratur gibt es unterschiedliche Gewichtungen, ob Information Wissen bedingt, oder umgekehrt und ob das eine oder das andere an Prozesse oder Akteure gekoppelt ist. Im nächsten Kapitel wird der Wissensbegriff näher untersucht.

3.2.3 Der Wissensbegriff

Wissen ist die systeminterne Repräsentation vorliegender Erfahrung eines kognitiven Agenten zu einem bestimmten Zeitpunkt, die einem zu überprüfenden Anspruch auf Gültigkeit ausge-

⁹⁷Vgl. KENDAL, Simon/CREEN, Malcolm: An Introduction to Knowledge Engineering. London, 2007, S.2-7.

⁹⁸Vgl. GUMM, Heinz Peter/SOMMER, Manfred: Einführung in die Informatik. 9. Auflage. Oldenbourg Verlag, 2011, S.1-11.

⁹⁹Vgl. FAVRE-BULLE: Information und Zusammenhang: Informationsfluß in Prozessen der Wahrnehmung, des Denkens und der Kommunikation, S.15-38.

¹⁰⁰STOCK: Information retrieval: Informationen suchen und finden, S.21.

¹⁰¹BALLOD: Informationen und Wissen im Griff: effektiv informieren und effizient kommunizieren, S.9.

setzt sein muss. Als solches prägt Wissen das Handeln und Denken des kognitiven Agenten auf den unterschiedlichsten Ebenen und dient zur Lösung von Problemen. Das jeweils aktuelle Wissen bildet einen kontextuellen Rahmen, in dem ankommende und bestehende Information interpretiert und zu neuen Erfahrungen verarbeitet werden.¹⁰²

Es gibt unterschiedlichste Zugänge und Definitionen für den Wissensbegriff. Dieser Absatz ist das Ergebnis der Definition von Wissen von FAVRE-BULLE, der versucht sich diesen Begrifflichkeiten auf interdisziplinärer Ebene anzunähern.¹⁰³ In dieser Definition sind Aspekte angeführt, die in vielen Definitionen auftauchen und mit der Entwicklung eines Wissensbegriffes einhergehen. An dieser Stelle soll versucht werden die facettenreichen Aspekte des Wissensbegriffs weiter aufzuschlüsseln. FAVRE-BULLE führt dazu 5 Punkte an, die in diesem Zusammenhang von zentraler Bedeutung sind:¹⁰⁴

Bindung an einen kognitiven Agenten.

Heutzutage wird auch von der sogenannten 'Wissensgesellschaft' gesprochen. BAUER merkt an, dass die Verteilung und Zugänglichkeit von Fakten und Wissen alleine noch nicht auf eine Wissensgesellschaft schließen lassen kann.¹⁰⁵ Erst durch das kritische Hinterfragen, das Reflektieren und Strukturieren entsteht Wissen. Genau diese Möglichkeiten zu Handeln und mit der Welt zu interagieren, beschreiben einen kognitiven Agenten. Ohne Kognition, wie in Abbildung 2 dargestellt, wird aus Information kein Wissen.

Momentaufnahme der bis dahin verarbeiteten Ereignisse eines kognitiven Agenten.

Ein kognitiver Agent kann etwas nicht wissen, das er während seiner Existenz noch nicht bewusst wahrgenommen und verarbeitet hat und er kann nur neues Wissen mit dem Wissen abgleichen, das er bis zu diesem Zeitpunkt gesammelt bzw. gelernt hat. Wissen unterliegt aus diesem Blickwinkel auch einer zeitlichen Dimension.

Einfluss auf Verhalten und die Zustände eines kognitiven Agenten.

Darunter kann das Phänomen des Lernens verstanden werden, das auch von KUHLEN hervorgehoben wird. Lernen ist dann gegeben, wenn Informationen dauerhaft in Wissensstrukturen eines kognitiven Agenten abgelegt wurden.¹⁰⁶ Gelernt werden kann **deklaratives Wissen** und **prozedurales Wissen**. Lernt man die Hauptstädte Europas auswendig, so ist die Kenntnis über diese Fakten deklaratives Wissen. Es ist explizites, dem kognitiven Agenten bewusstes, und sensorisches Wissen. Werden Fertigkeiten wie das Radfahren erlernt, spricht man von prozeduralem Wissen, das durch ausprobieren und üben gekennzeichnet ist. Es ist implizites, dem

¹⁰²Vgl. FAVRE-BULLE: Information und Zusammenhang: Informationsfluß in Prozessen der Wahrnehmung, des Denkens und der Kommunikation, S.93-97.

¹⁰³Vgl. ebd., S.20-25.

¹⁰⁴Vgl. ebd., S.98-100.

¹⁰⁵BAUER, Reinhard: Die digitale Bibliothek von Babel: über den Umgang mit Wissensressourcen im Web 2.0. Hülsbusch, 2010.

¹⁰⁶Vgl. ebd., S.45.

kognitiven Agenten unbewusstes und motorisches Wissen.¹⁰⁷ Die interne Information (Erfahrung), die externen Informationen (Wahrnehmung), sowie das Wissen darüber, beeinflussen das Verhalten des kognitiven Agenten.¹⁰⁸

Wissen bedingt einen Gültigkeitsanspruch.

LIESSMANN führt an, dass Wissen sich auf Erkenntnis bezieht. Somit ist die Frage nach der Wahrheit eine Grundvoraussetzung für Wissen. Dieser Gültigkeitsanspruch, darf nach KUHLEN aber nicht absolut sein, viel eher gibt es für ihn Wahrheitsansprüche aus unterschiedlichen Perspektiven heraus.¹⁰⁹ CHRISHOLM sieht Wissen darin, dass die Voraussetzungen Akzeptanz, Wahrheit und Evidenz zur Gänze vorliegen und damit eine wahre Aussage existiert. Akzeptiert ein Subjekt eine Aussage als wahr, und diese Aussage ist nach objektiven Kriterien wahr und kein Zufall, dann ist darin eine Definition von Wissen gefunden.¹¹⁰ Hinsichtlich des Wahrheitsbegriffes wird auf Kapitel 2.2 verwiesen. In diesem Zusammenhang sollen auch die einführenden Überlegungen in dem soeben genannten Kapitel im Zusammenhang mit dem Wissensverständnis POPPER's gebracht werden. Welt 2, die an das menschliche Denken gebunden ist, wie etwa ein Gefühl, könnte man als subjektives Wissen verstehen. Objektives Wissen hingegen würde der Welt 3, dem vom Menschen unabhängigen Denken, entspringen. Subjektives und objektives Wissen haben einen unterschiedlichen Gültigkeitsanspruch. Ein Gefühl kann nur von einem Menschen wahrgenommen bzw. als wahr aufgefasst werden, im Unterschied zum objektiven Wissen.

Wissen bezieht sich auf etwas und erzeugt Kontexte.

Für LIESSMANN ist Wissen die Interpretation von Daten in kausalen Zusammenhängen und ihrer inneren Konsistenz. Erst wenn Information über eine eigene Struktur verfügt, spricht man von Wissen.¹¹¹ Diese Definition kann auch bei WERSIG gefunden werden, der Wissen als *'die Struktur des internen Außenweltmodells'*¹¹² des Organismus bzw. kognitiven Agenten betrachtet. Die Struktur der Innenwelt entsteht dadurch, dass die Außenwelt wahrgenommen und verarbeitet wird. In diesem Prozess wird Ungewissheit reduziert und Information entsteht.¹¹³ Als Wissen versteht WERSIG alles, das ein kognitiver Agent als wichtig und nachvollziehbar hält um die Interaktion mit seiner Umwelt organisieren zu können und um die Ungewissheit in der Außenwelt zu reduzieren.¹¹⁴

¹⁰⁷Vgl. FAVRE-BULLE: Information und Zusammenhang: Informationsfluß in Prozessen der Wahrnehmung, des Denkens und der Kommunikation, S.100-105.

¹⁰⁸Vgl. ebd., S.96.

¹⁰⁹Vgl. BAUER: Die digitale Bibliothek von Babel: über den Umgang mit Wissensressourcen im Web 2.0, S.45-46.

¹¹⁰Vgl. STOCK/STOCK: Wissensrepräsentation: Informationen auswerten und bereitstellen, S.20.

¹¹¹Vgl. BAUER: Die digitale Bibliothek von Babel: über den Umgang mit Wissensressourcen im Web 2.0, S.45.

¹¹²WERSIG: Information-Kommunikation-Dokumentation. Ein Beitrag zur Orientierung der Informations- und Dokumentationswissenschaft, S.64.

¹¹³Vgl. ebd., S.42-43.

¹¹⁴Vgl. WERSIG, Gernot: Die Komplexität der Informationsgesellschaft. Konstanz, 1996, S.146.

3.3 Grundbegriffe des Information Retrieval

Eine Historikerin interessiert sich für Kriminalfälle des frühen 20. Jahrhunderts. Aus diesem Grund besucht sie die Webpräsenz der digitalen Sammlung des Hans Gross Kriminalmuseums. Für ihre Forschungsfrage benötigt sie weiteres Quellenmaterial, um ihre Thesen adäquat untermauern zu können.

Dies könnte ein realistisches Beispielszenario sein, mit dem die digitale Sammlung konfrontiert werden könnte. Der kognitive Agent - in diesem Szenario eine Wissenschaftlerin - hat ein konkretes Informationsbedürfnis in Form ihrer Forschungsfrage und versucht relevantes Quellenmaterial zu finden. Dieses Quellenmaterial sind die Karteikarten im Kriminalmuseums, die Wilderei-Delikte behandeln.¹¹⁵ An Hand dieses Beispiels eines Retrieval-Prozesses, wie er in der Einleitung in Abbildung 1 skizziert wurde, sollen die wichtigsten Grundbegriffe und Denkmuster des IR ausführlicher beschrieben werden. Das Kapitel wird mit einer Übersicht gängiger Ordnungsstrukturen und Wissensrepräsentationen abgeschlossen.

3.3.1 Informationsbedarf

Eine Wissenschaftlerin hat ein Informationsbedürfnis, in Form ihrer Forschungsfrage und benötigt Daten- bzw. Quellenmaterial. Als **Informationsbedürfnis** versteht man die konkrete, subjektive Situation mit der eine Person konfrontiert ist, wenn sie den Bedarf nach einer Information verspürt. STOCK unterscheidet Informationsbedürfnis vom **Informationsbedarf**, der das Fehlen von Information in einem objektiven Sinn meint. Informationsbedarf beschreibt Qualität, Art und Menge an Information, die ein kognitiver Agent benötigt, um eine bestimmte Aufgabenstellung bewältigen zu können.¹¹⁶ Ein Informationsbedürfnis oder auch ein Informationsbedarf werden nur gestillt, wenn die, durch einen Retrieval-Prozess angebotenen Dokumente, auch von **Relevanz** bzw. **Pertinenz** sind. Ein Dokument ist relevant, wenn es objektiv einen Informationsbedarf deckt und damit eine Wissenslücke schließt ohne den Faktor Mensch miteinzubeziehen. Ein objektiv relevantes Dokument, das aber in einer nicht verständlichen Sprache oder von einem nicht geschätzten Autor verfasst wurde, ist zwar relevant, der Nutzer verwendet es aber nicht. Der Begriff Pertinenz bezieht sich auf die subjektive Wahrnehmung eines Nutzers und dessen subjektives Informationsbedürfnis und darauf, ob der Nutzer ein Dokument für geeignet oder ungeeignet befindet.¹¹⁷ Die Entscheidung ob etwas relevant ist oder nicht spiegelt eine Prozess wider, aus dem sich vier Handlungen, angelehnt an SUTCLIFF et al., ableiten lassen können:

- Die Identifikation der Forschungsfrage.
- Die Artikulation des Informationsbedürfnisses.

¹¹⁵Siehe Kapitel 5 für weiter Information

¹¹⁶Vgl. HEINRICH, Lutz J/RIEDL, René/STELZER, Dirk: Informationsmanagement: Grundlagen, Aufgaben, Methoden. Oldenbourg, 2014, S.63-64.

¹¹⁷STOCK: Information retrieval: Informationen suchen und finden, S.68-79.

- Die Formulierung der Frage, um das aus der Forschungsfrage abgeleitete Informationsbedürfnis zu stillen
- Die Evaluierung der Antworten bzw. Ergebnisse.¹¹⁸

Ist dieser Zyklus, der nicht statisch sondern dynamisch zu betrachten ist, abgeschlossen, so gilt ein Informationsbedürfnis als gestillt.

Zwei Werte die dabei helfen die Relevanz eines Dokuments zu evaluieren, sind die Begriffe **Precision**, als Messwert für die Genauigkeit, und **Recall**, als Messwert für die Vollständigkeit, aller relevanter Dokumente. Dabei ist es wichtig, dass Recall und Precision in einem ausgewogenen Verhältnis zueinander stehen. Geht man von einem sehr großen Recall-Wert aus, so hat man sehr viele Dinge gefunden, aber viele davon sind nicht brauchbar. Bei einem zu hohen Precision-Wert, sind die Treffer zwar gut verwertbar, aber viel Information bleibt verborgen.¹¹⁹ STOCK¹²⁰ und BAEZA-YATES et. al.¹²¹ unterscheiden zwischen zwei Arten von Informationsbedarf und ihrer Befriedigung. Der statische **Information Lookup** bezeichnet ein konkretes Informationsbedürfnis, bei dem das Bedürfnis nach einem Faktum gestillt werden will (Wer war Hans Gross?). Unter dem Begriff **Exploratory Search** versteht man einen problemorientiertes und dynamisches Informationsbedürfnis, das durch eine Kette von Abfragen und Verknüpfungen gestillt werden kann (Suche nach Quellenmaterial zu einer wissenschaftlichen Forschungsfrage).¹²² In Kapitel 4 wird genauer auf diese Form des Informationsbedarf eingegangen.

3.3.2 Dokumente

Dem Informationsbedarf eines kognitiven Agenten steht ein Korpus an Informationsressourcen gegenüber, der die Antwort auf eine Query enthalten könnte. Dieser Informationskorpus beinhaltet Dokumente. Der Bestand aller Karteikarten im Kriminalmuseum kann als ein solcher Informationskorpus verstanden werden. Jede einzelne Karteikarte darin stellt ein eigenständiges **Dokumente** dar, nachdem gesucht oder gestöbert bzw. das gefunden werden kann. Für GAUS ist ein Dokument alles, das auf irgendeine Art und Weise gesammelt, geordnet und erschlossen werden kann, um es für Nutzerinnen wieder verwertbar bzw. auffindbar zu machen. So zum Beispiel eine Karteikarte oder die digitale Repräsentation davon in einer digitalen Sammlung, die mittels Metadaten und Index in einer Ordnungsstruktur gesucht und gefunden werden kann. Vier Kriterien führt STOCK an, die gegeben sein müssen, um etwas als ein Dokument verstehen zu könnten:

- Materialität: Ein Objekt ist digital oder real existent.

¹¹⁸Vgl. SUTCLIFFE, Alistair/ENNIS, Mark: Towards a cognitive theory of information retrieval. Interacting with computers, 10 1998, Nr. 3.

¹¹⁹Vgl. POETZSCH: Information Retrieval: Einführung in Grundlagen und Methoden, S.20-21.

¹²⁰Vgl. STOCK: Information retrieval: Informationen suchen und finden, S.51-67.

¹²¹Vgl. BAEZA-YATES et al.: Modern information retrieval. The concept and technology behind search, S.22-23.

¹²²Vgl. STELZER, Dirk: Informationsbedarf. In MERTENS, P. (Hrsg.): Lexikon der Wirtschaftsinformatik. 4. Auflage. Berlin, 2004, S.238.

- Intentionalität: Ein Objekt ist mit einer Bedeutung verknüpft.
- Erarbeitung: Ein Objekt wurde geschaffen bzw. erzeugt.
- Wahrnehmung: Ein Objekt wird als Dokument bezeichnet und auch als solches verstanden.

Der Dokumentbegriff kann auch anders verstanden werden. Ein kognitiver Agent versucht mittels der zur Verfügung stehenden Syntax einen bestimmten Inhalt, die Semantik, zu formulieren, um einen Kommunikationsprozess anzustoßen. Die Struktur, der Informationsgehalt und das Medium - in dem Struktur und Informationsgehalt zusammengefasst sind - kann in seiner Gesamtheit als Dokument verstanden werden. Unterschiedliche Dokumenttypen verfolgen unterschiedliche Ziele. Die Darstellung eines Dokumentes und sein Ziel sind weitere Bestandteile des Dokumentbegriffes.¹²³ So verfolgt ein handgeschriebener Brief eine andere Intention also ein Kriminalakt, woraus auch ihre Struktur und ihre Repräsentation erklären lassen.

Diese Kriterien gelten auch für nicht-textuelle Dokumente. Multimediale Dokumente umfassen Bilder, Videos, Audio oder 'Rich Media'.¹²⁴ Multimediale Dokumente sind mit weiteren Herausforderungen im Bereich des IR konfrontiert.¹²⁵ Gibt es keine digitalen Vollformen eines Dokumentes, in der jeder Bereich indexiert werden kann, so bedarf man zusätzlicher Ressourcen um IR umsetzen zu können.¹²⁶ Dazu werden sogenannte **dokumentarische Bezugseinheiten** als Hilfsmittel verwendet, die dabei helfen, Dokumente in einem Bestand bzw. Informationskorpus wiederzufinden. Ein Findbuch in einem Archiv ist beispielsweise eine dokumentarische Bezugseinheit und stellt deskriptive und administrative Metadaten¹²⁷ zur Verfügung, um Dokumente auffindbar zu machen.¹²⁸

Eine **Signatur**, die bei der Erschließung vergeben wird, dient der Erkennung von Dokumenten in einem Bestand. Eine dokumentarische Bezugseinheit verweist auf Signaturen, über die Dokumente gefunden werden können. Sie müssen eindeutig, kurz, sortierfähig, maschineneignet, erweiterungsfähig und mnemotechnisch (merkbar) sein.¹²⁹

Im **Information Indexing** bilden dokumentarische Bezugseinheiten kleinste Formen der Repräsentation einer Dokumentationseinheit. Autor, Veröffentlichungsdatum oder ähnliche Information wird in Verbindung zu einem Dokument gesetzt. Diese **Deskriptoren**, also Bezeichnungen zur inhaltlichen Beschreibung eines Objektes, ermöglichen es, diese Begriffe (und nicht

¹²³Vgl. BAEZA-YATES et al.: Modern information retrieval. The concept and technology behind search, 203-204.

¹²⁴Im Sinne interaktiver Dokumente, wie beispielsweise Social Media Anwendungen oder Games, die als Zusammenspiel von Text, Bild, Audio und Video verstanden werden können, https://en.wikipedia.org/wiki/Interactive_media, 21.02.2017.

¹²⁵Vgl. WITTEN, Ian H./BAINBRIDGE, David/NICHOLS, David M.: How to Build a Digital Library, Second Edition. San Francisco, 2010, S.215.216.

¹²⁶Vgl. RÜGER: Multimedia Resource Discovery, S.39-55.

¹²⁷Siehe Kapitel 5.3.2

¹²⁸STOCK: Information retrieval: Informationen suchen und finden, S.82-93.

¹²⁹Vgl GAUS: Dokumentations-und Ordnungslehre: Theorie und Praxis des Information Retrieval, S.11-16.

Worte) für das IR zu nutzen. Ein **Begriff** unterscheidet sich hinsichtlich eines Wortes darin, dass die Bedeutung eindeutig und an ein semantisches Umfeld gekoppelt ist.¹³⁰ Das semiotische Dreieck in Abbildung 4 zeigt, was darunter verstanden wird.

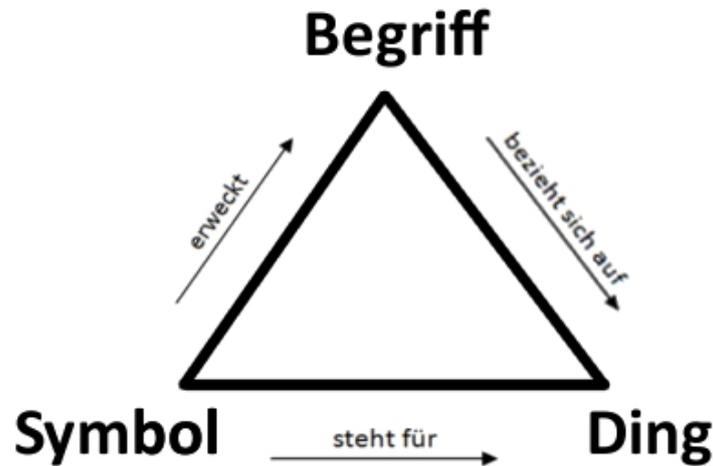


Abbildung 4: Semiotisches Dreieck

Die Unterscheidung zwischen Wort und Begriff ist deswegen notwendig, um den Problemen, die beispielsweise mit Synonymen¹³¹ oder Homonymen¹³² verbunden sind, entgegenzutreten. Karteikarten sind mit einer Signatur bzw. Inventarnummer versehen. Im semiotischen Dreieck ist das reale Dokument ein *Ding*. Der Begriff 'Karteikarte' bezieht sich auf alle realen Objekte, die die gleichen Eigenschaften besitzen. Dem gegenüber steht die reine syntaktische Zeichenfolge 'Karteikarte'. Die Zeichenfolge (Symbol) erweckt beim kognitiven Agenten einen Begriff, der in Abhängigkeit zum jeweiligen Kontext steht. Unter 'Karteikarte' kann auch ein Zettel zum Lernen von Vokabeln verstanden werden. In der digitalen Sammlung stellen 'Karteikarten' einen eigenständigen Sammlungsbereich dar.¹³³ Für das IR ist diese Trennung wichtig, da es den Suchprozess in einer Dokumentensammlung beeinflussen kann. Eine Volltextsuche nach 'Karteikarte' unterscheidet sich von der Suche nach dem Begriff 'Karteikarte'. Eine Suche nach unstrukturierten Text, im Gegensatz zum semantischen Konzept, erlaubt diesbezüglich keine Differenzierung.

3.3.3 Suchanfrage und ihre Formulierung

Suchanfragen werden im IR auch als *Query* bezeichnet. Nutzerinnen haben unterschiedliche Suchstrategien und Werkzeuge zur Verfügung um ihre individuellen Bedürfnisse zu stillen. Sie

¹³⁰Vgl. STOCK: Information retrieval: Informationen suchen und finden, S.270-291.

¹³¹Wort mit unterschiedlicher Syntax, aber gleicher Semantik, Tierarzt und Veterinär, Vgl. STOCK: Wissensrepräsentation: Information auswerten und bereitstellen, S.54.

¹³²Wort mit gleicher Syntax, aber anderer Semantik. Die 'Bank' als Sitzgelegenheit und die 'Bank' als Geldinstitut, Vgl. STOCK: Wissensrepräsentation: Information auswerten und bereitstellen, S.54-55.

¹³³Vgl. STECHOW, Arnim von: Schritte zur Satzsemantik. 2004, S.9-12.

müssen ihr Informationsbedürfnis in eine geeignete Suchanfrage 'übersetzen'. LEWANDOWSKI fasst alle Fertigkeiten im Prozess des Formulierens einer Suchanfrage als Informationskompetenz zusammen. Weiters beschreibt er Teilschritte eines Suchprozesses:

- Auswahl eines geeigneten Retrieval-Systems.
- Formulierung und Eingabe der Suchanfrage.
- Betrachtung der Suchergebnisse und Auswahl geeignet erscheinender Treffer.
- Evaluierung dieser Treffer.
- Suche im gefundenen Dokumente oder Neuformulierung der Anfrage.

Ein solcher Prozess ist weniger eine Informationsübertragung von einem Sender an einen Empfänger, sondern viel mehr ein Dialog. Der Empfänger der Suchanfrage, beispielsweise eine Suchmaschine, wird selber zum Sender und SHANNON's Modell um einen zweiten Informationskanal erweitert. Weiters kann von jedem Schritt aus zu einem der davor liegenden gewechselt werden.¹³⁴ Liefert das System nicht zufriedenstellende Suchtreffer zurück, so kann die Nutzerin wieder zu einem vorherigen Schritt zurückkehren, um Suchanfragen umformulieren, ergänzen oder neu beginnen zu können.¹³⁵

IR-Systeme müssen folglich unterschiedliche Wege anbieten, die zu den gewünschten Zielen führen. Dabei unterscheidet BAEZA-YATES et al. zwischen **Search**, der 'klassischen' Suche mittels Suchwörtern, und **Browsing**. Das englische Wort für 'stöbern' bezeichnet den Vorgang eines Nutzers dargestellt Informationsressourcen zu überfliegen und geeignet erscheinende Dokumente auszuwählen. User bevorzugen das Browsing gegenüber der Eingabe von Suchwörtern besonders dann, wenn sie in der Formulierung der Suchwörter unsicher sind.¹³⁶ Diese Unterscheidung ist essentiell für die Darstellung der Such- und Browse-Funktionalitäten eines IR-Systems. Generell ist festzuhalten, dass eine User im seltensten Fall mit nur einer Anfrage sofort ans Ziel kommt. Vielmehr muss er auf Basis der gefundenen Dokumente seine Anfrage neu formulieren und neu entscheiden, ob die angezeigten Ergebnisse für ihn relevant sind. Diese schrittweise Annäherung an das Ergebnis wird als '**Berrypicking**' bezeichnet.

Die **Volltextsuchanfrage in der digitalen Sammlung** soll als Beispiel dienen, um Suchanfragen und ihre Formulierung weiter auszuführen. Dort gibt es unterschiedliche Möglichkeiten Suchanfragen zu modifizieren und einzugrenzen. Die **Trunkierung** erlaubt es mit sogenannten Trunkierungssymbolen (Wildcards) Platzhalter für Wortvariationen zu verwenden. Das '*'-Symbol dient als Platzhalter für eine beliebige Anzahl von Zeichen, das '?'-Symbol für genau

¹³⁴Vgl. REGINALD: Information Retrieval-Suchmodelle und Data-Mining-Verfahren für Textsammlungen und das Web, S.24-26.

¹³⁵Vgl. LEWANDOWSKI, Dirk: Informationskompetenz und das Potenzial der Internetsuchmaschinen. In SÜHL-STROHMENGER, W. (Hrsg.): Handbuch Informationskompetenz. Berlin, 2012, S.1-9.

¹³⁶Vgl. BAEZA-YATES et al.: Modern information retrieval. The concept and technology behind search, S.24-25.

ein Zeichen. **Boolsche Operatoren** erlauben es mit 'AND', 'OR' und 'NOT' logische Operatoren zu verwenden, um die Abfrage weiter zu verknüpfen und zu konkretisieren.¹³⁷ In der **Fuzzy-Logik** kann mit dem Tilde-Symbol ein Grad der Ähnlichkeit und die Suche nach benachbarten Wörtern in der Query angegeben werden. In Kapitel 3.4 werden bool'sche Operatoren und die Fuzzy Logik im Zuge von IR-Modellen weiter erläutert. Diese beispielhaften Möglichkeiten werden als **Query-Reformulation** bzw. **Query-Expansion** zusammengefasst.¹³⁸

Die Eingabe des Suchbegriffes im Suchschlitz der digitalen Sammlung '1936' liefert sieben Treffer.¹³⁹ Fünf Objekte gehören zum Sammlungsbereich 'Karteikarten' und zwei Treffer zum Sammlungsbereich 'Objekte'. Die gefundenen Objekte beinhalten die Zeichenkette '1936', die ein kognitiver Agent sofort als Datum interpretieren kann. Wird die Suchanfrage auf '1936 AND GAMS'¹⁴⁰ erweitert, so findet sich nur noch eine Karteikarte. Dieses Dokument enthält die Zeichenkette '1936' und 'Gamsen', als ähnliches Wort zu 'Gams', das mittels des Tilde-Symbols gefunden wurde.

3.3.4 Index und Ranking

Ein **Index** ist eine Datenstruktur, in der alle Dokumente eines Informationskorpus mittels eines Sets repräsentativer Deskriptoren erschlossen sind. Anstatt alle Dokumente eines Bestandes nach einer Suchanfrage zu prüfen, ob sie der Suchanfrage entsprechen, wird der Index abgearbeitet. Dadurch erhöht sich die Performanz des Retrieval-Prozesses deutlich, besonders dann, wenn man von einer großen Datenmenge ausgeht. Lange Ladezeiten sind für jedes Informationssystem ein zu vermeidendes Problem.¹⁴¹ Folgender Text findet sich in einer Karteikarte des Hans Gross Kriminalmuseums:

*"Täter 1) hat im Juli 1936 im Revier der Bundesforstverwaltung Hinterberg unter der Rötelssteinwand gewildert und einen Gamsbock i. W. v. 35 S - erlegt u. gestohlen. Bei dieser Gelegenheit hat er ohne behördliche Bewilligung zum Tragen verbotener Waffen diesen abnehmbaren Kugelstutzen mit auswechselbaren Läufen getragen"*¹⁴²

Zu Erstellung eines Index für alle Karteikarten könnte man nun beispielsweise alle Nomen und Verben aus diesem Text herausziehen und in normalisierter Form in den Index aufnehmen. Normalisiert bedeutet in diesem Zusammenhang, dass Verben in ihrer Infinitivform, Nomen in

¹³⁷Vgl. BAEZA-YATES et al.: Modern information retrieval. The concept and technology behind search, S.258.

¹³⁸Vgl. ebd., S.32-33.

¹³⁹Eingabe der Query: <http://gams.uni-graz.at/search/gsearch?query=1936&hitPageSize=10&hitPageStart=1&pid=km&x2=http://gams.uni-graz.at/km/km-search.xsl>, 28.02.2017.

¹⁴⁰Eingabe abgeänderter Query: <http://gams.uni-graz.at/search/gsearch?query=1936ANDGAMS~&hitPageSize=10&hitPageStart=1&pid=km&x2=http://gams.uni-graz.at/km/km-search.xsl>, 28.02.2017.

¹⁴¹Vgl. REGINALD: Information Retrieval-Suchmodelle und Data-Mining-Verfahren für Textsammlungen und das Web, S.84-85.

¹⁴²Ausschnitt der digitalen Repräsentation einer Karteikarte im Hans Gross Kriminalmuseum, <http://gams.uni-graz.at/o:km.87>, 16.12.2016.

ihrer Singularform übernommen werden und andere Wörter oder Satzzeichen ignoriert werden. Für dieses Dokument würde folgender Eintrag im Index existieren. Die Literatur führt dieses Set aller **Indexterme** (Deskriptoren) als Menge *K* an.¹⁴³

K = { Täter, haben, Juli, 1936, Revier, Bundesforstverwaltung, Hinterberg, Rötelsteinwand, wildern, Gamsbock, 35, erlegen, stehlen, Gelegenheit, Bewilligung, Tragen, Waffe, Kugelstutze, Lauf, tragen}

Das Ergebnis ist eine Menge an Termen, die die wichtigsten Informationen der Karteikarte repräsentieren. Aus dieser Sichtweise kann man ein Dokument als eine oder mehrere Mengen an Termen verstehen. In dieser Form sind alle Terme gleich relevant. Es handelt sich um eine **ungewichtete Indexierung**. Ein kognitiver Agent könnte bestimmte Terme als relevanter erachten als andere. So sind zum Beispiel die Jahresangabe '1936' oder die Ortsangabe 'Hinterberg' womöglich wichtigere Terme, als beispielsweise der Term 'tragen'. Nun ist es möglich bestimmte Terme manuell oder automatisiert einer **gewichteten Indexierung** zu unterziehen. Diese Gewichtung ermöglicht es gewissen Kategorien, wie etwa 'Jahr' oder 'Ort' den Wert 1 zu geben und dem Verb 'tragen' die Gewichtung 0. Durch unterschiedliche Wertigkeiten in der Gewichtung werden unterschiedliche Grad der Relevanz ausgedrückt, mit denen das **Ranking** umgesetzt werden kann. Die Terme 'Jahr' und 'Ort' werden vor dem Term 'tragen' gereiht. Dokumente in denen mehr relevante Terme vorkommen, werden als ganzes höher gewertet.¹⁴⁴ Die Vorgehensweise der Indexierung kann unterschiedlich erfolgen, je nach dem welches Ziel damit erreicht werden soll und welche Kapazitäten zur Umsetzung vorhanden sind. Das soeben angeführte Beispiel stellt eine Möglichkeit einer automatisierten und simplen Indexierung dar. Die Reduktion auf gemeinsame Wortstämme, die Auswahl aller Nomen und Verben, das Weglassen von Stoppwörtern und die Gewichtung der extrahierten Terme kann durch einen einfachen Algorithmus erfolgen. Weitere Verfahren, um aus Dokumenten einen Index zu generieren, könnten beispielsweise eine lexikalische Analyse, das Eliminieren von Stoppwörtern, das 'Steering', die Auswahl von Stichwörtern umfassen oder andere Verfahren, um textuelle Dokumente zu erschließen.¹⁴⁵ Indexieren, die praktische Zuordnung von Begriffen zu dokumentarischen Bezugseinheiten, kann aber auch manuell bzw. halbautomatisch erfolgen, gerade dann, wenn keine Volltextsuche umgesetzt werden soll, wie im soeben genannten Beispiel, sondern eine Suche nach Begriffen.¹⁴⁶

¹⁴³Vgl. LINCKELS/MEINEL: E-librarian service: user-friendly semantic search in digital libraries, 82-83.

¹⁴⁴Vgl. FUHR, Norbert: Modelle im Information Retrieval. In KUHLEN, Rainer/SEMAR, Wolfgang/STRAUCH, Dietmar (Hrsg.): Grundlagen der praktischen Information und Dokumentation: Handbuch zur Einführung in die Informationswissenschaft und -praxis. 6. Auflage. Berlin and Boston, 2013, S.321.

¹⁴⁵Vgl. BAEZA-YATES et al.: Modern information retrieval. The concept and technology behind search, S.223-228.

¹⁴⁶Vgl. NOHR, Holger: Grundlagen der automatischen Indexierung: ein Lehrbuch. Berlin, 2003, S.24-30.

3.3.5 Ordnungsstrukturen und Wissensrepräsentation

Eine weitere Form der Erschließung und Formalisierung eines Informationskorpus kann mittels Ordnungsstrukturen und ihrer Wissensrepräsentation umgesetzt werden. In der digitalen Sammlung des Kriminalmuseums gibt es unterschiedliche Typen von Objekten, die wiederum aus unterschiedlichen Materialien bestehen, die wiederum in unterschiedlichen Kontexten stehen können. Um diese Information bzw. dieses Wissen darstellen zu können, bedarf es Möglichkeiten der Strukturierung und Ordnung. Begriffe in Form von Deskriptoren werden verwendet, um als Stellvertreter in einem Index angezeigt werden zu können. In diesem Zusammenhang kann man **kontrolliertes** und **freies Indexieren** unterscheiden, um geeignete Deskriptoren zu finden und festzulegen. Beim freien Indexieren werden einem Dokument Deskriptoren ohne weitere Regeln zugeordnet. Beim gebundenen Indexieren werden festgelegte Begriffe und Regeln verwendet - beispielsweise auch von Fachexperten einer Wissensdomäne - mit dem Ziel, deren Wissen formalisieren zu können. Erst durch die Formalisierung kann es auch maschinell verarbeitet und genutzt werden.¹⁴⁷

STOCK fasst die wichtigsten Ordnungsstrukturen, aus denen sich Wissensrepräsentationen ableiten lassen, in Abbildung 5 zusammen. Er beschreibt den Zusammenhang zwischen Umfang der jeweiligen Wissensdomäne und der Ausdrucksstärke der Methode. Wissensrepräsentation versucht Vertreter von Dokumenten zu erzeugen, die die Dokumente nicht nur auf einer Metaebene abbilden, sondern auch Retrieval-Prozesse ermöglichen. Der Bereich des Indexieren fasst diese Methoden, die sich in ihren Möglichkeiten hinsichtlich Größe der Wissensdomäne und gewünschten Erschießungsgrad unterscheiden, zusammen. In vielen Fällen ist es notwendig, dass mehrere dieser Methoden zur Wissensrepräsentation in Zusammenarbeit verwendet werden.¹⁴⁸ Je nach vorhandener Kapazität und welches Ziel durch Retrieval-Prozesse erreicht werden soll, bieten sich unterschiedliche Ordnungs- und Repräsentationsstrukturen an. Beispielsweise ist der Bestand des Kriminalmuseums als geisteswissenschaftliche Quellensammlung zu verstehen, die direkt an Wissenschaftlerinnen gerichtet ist. Der Bestand des Kriminalmuseums wurde von qualifizierten Mitarbeitern und Mitarbeiterinnen nach bestimmten Kriterien erschlossen. Aus diesem Grund ist es sinnvoll die Expertise und das Wissen dieser Personen zu formalisieren und für Retrieval-Prozesse zu nutzen. Je stärker Information oder Wissen formalisiert werden soll, desto eher bedarf man ausdrucksstarker Methoden, um dieses Wissen darzustellen und nutzen zu können. Je stärker Information strukturiert und an einen kognitiven Agenten gebunden ist, desto ausdrucksstärker muss folglich auch die Methode der Wissensrepräsentation sein. Zur Strukturierung bieten sich für diesen Fall, bis auf die Folksonomie, alle Formen der Wissensrepräsentation an. Konkret wurden Klassifikationen und Thesauri in der Projektumsetzung verwendet.

¹⁴⁷Vgl. STOCK: Information retrieval: Informationen suchen und finden, S.451-459.

¹⁴⁸Vgl. STOCK/STOCK: Wissensrepräsentation: Informationen auswerten und bereitstellen, S.37-45.

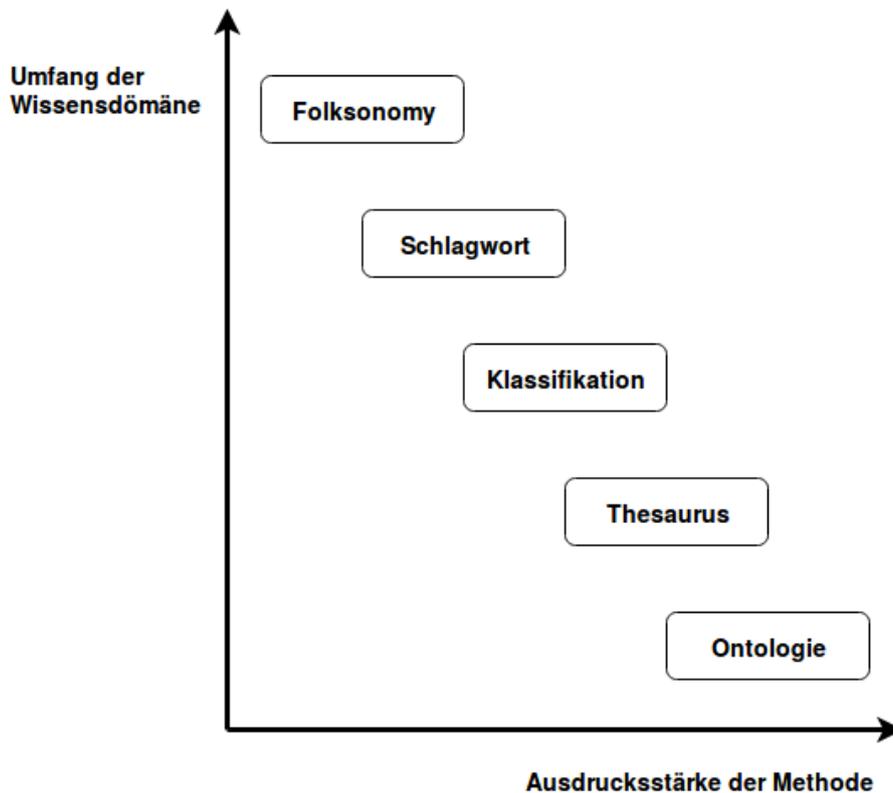


Abbildung 5: Ausdrucksstärke von Methoden der Wissensrepräsentation und Größe der Wissensdomäne

Folksonomie

Webpräsenzen, wie etwa die Videoplattform *Youtube*, ermöglichen ihren Nutzern Videos zu verschlagworten, also mit Deskriptoren frei zu indexieren. Dies wird als *Tagging* bezeichnet. Schlagwörter dienen als Deskriptoren und werden mittels Algorithmen verarbeitet und gewichtet.¹⁴⁹ So effizient ein Kollektiv an Nutzern auch eine große Zahl an Dokumenten sichten und erschließen kann, so kann nie eine tief gehende, fachliche und qualitativ überprüfte Auszeichnung erfolgen.¹⁵⁰

Schlagwörter

Bei der Verschlagwortung vergibt eine Person nach einem zuvor festgelegten Schema Schlagwörter, die ein Dokument beschreiben. Passende Schlagwörter für eine Karteikarte sind relevante Terme, die den Inhalt beschreiben, wie etwas 'Wilderei', 'Steiermark' oder '1936'. Der Unterschied zur Folksonomie besteht darin, dass dieser Prozess qualitativ geleitet und auf eine Personengruppe (z.B. Fachwissenschaftler) eingeschränkt ist, wohingegen bei der Folksonomie eine Vielzahl von Akteuren am gemeinschaftlichen Tagging beteiligt sind.¹⁵¹

¹⁴⁹Vgl. CARLIN, Sascha: Schlagwortvergabe durch Nutzende (Tagging) als Hilfsmittel zur Suche im Web. Darmstadt, 2006.

¹⁵⁰Vgl. BAEZA-YATES et al.: Modern information retrieval. The concept and technology behind search, S.232.

¹⁵¹Vgl. HORT, I: Holger Nohr: Grundlagen der automatischen Indexierung: ein Lehrbuch. ABI TECHNIK, 24 2004, Nr. 1, S.74-75.

Klassifikation

Beschreibt die Gruppierung und Einteilung von Wissen nach methodischen Prinzipien in Klassen. Nach dem Prinzip 'Gleiches zu Gleichem' werden Dokumente, wenn sie über mindestens ein gemeinsames Kriterium verfügen, in einer Klasse zusammengefasst. Die Kriterien für eine Klassifikation sind:

- Eine systematische Ordnung von Begriffen.
- Relationen werden durch nicht-natürlichsprachige Notation formalisiert.
- Möglichkeit von Wortkombinationen, jedoch mit geringer Ausdrucksstärke.¹⁵²

Thesaurus

Ein Thesaurus bezeichnet eine geordnete Sammlung von Begriffen und ihren Beziehungen. Er dient dazu Inhalte einer Wissensdomäne zu indexieren, speichern und für Retrieval-Prozesse zu erschließen. Die Merkmale des Thesaurus umfassen:

- Eine terminologische Kontrolle, in der Begriffe und Bezeichnungen aufeinander bezogen sind.
- Darstellung von Beziehungen zwischen Begriffen.
- Einen natürlichsprachlichen Zugang
- Flexibilität, Ausdrucksstärke und Möglichkeiten der Kombination von Begriffen.¹⁵³

Die Objekte des Bestandes im Kriminalmuseum werden mittels zweier Thesauri zu Objekt- und Materialtypen zugeordnet. Die technische Umsetzung diesbezüglich findet sich in Kapitel 5.4.1.

Die **Ontologie** wird in Kapitel 5.4.1 weiter beschrieben, da die Erläuterung mit der technischen Umsetzung zusammen sinnvoll ist. Mit einer Ontologie könnte man beispielsweise die Kontexte der Objekte abbilden. So wäre man in der Lage Personen, Objekte, Orte oder die Geschichte eines Objektes miteinander in Verbindung zu setzen.¹⁵⁴

¹⁵²Vgl. MANECKE, Hans-Jürgen: Klassifikation, Klassieren. In KUHLEN, Rainer/SEEGER, Thomas/STRAUCH, Dietmar (Hrsg.): Grundlagen der praktischen Information und Dokumentation: Handbuch zur Einführung in die Informationswissenschaft und -praxis. 5. Auflage. München, 2004, S.127–131.

¹⁵³Vgl. BURKART, Margarete: Thesaurus. In KUHLEN, Rainer/SEEGER, Thomas/STRAUCH, Dietmar (Hrsg.): Grundlagen der praktischen Information und Dokumentation: Handbuch zur Einführung in die Informationswissenschaft und -praxis. 5. Auflage. München, 2004, S.141–154.

¹⁵⁴Vgl. STOCK/STOCK: Wissensrepräsentation: Informationen auswerten und bereitstellen, S.176-272.

3.4 Information Retrieval-Modelle

'Information Retrieval-Modelle spezifizieren, wie zu einer gegebenen Anfrage die Antwortdokumente aus einer Dokumentenkollektion bestimmt werden'.¹⁵⁵

IR-Modelle befinden sich zwischen der Theorie und der Praxis. Ihre Entwicklung entspringt tendenziell dem Prinzip 'trial and error'. Weder die eine, noch die andere Seite alleine ist in der Lage effektive IR-Systeme umzusetzen. Die theoretischen Grundlagen ergeben sich aus den formalen Modellen, die unterschiedliche Herangehensweisen für unterschiedliche Arten von Problemen skizzieren. Kein Modell ist besser oder schlechter als ein anderes, aber manche Modelle eignen sich für bestimmte Problemstellungen besser als andere. Auch können die Ansätze der Modelle miteinander kombiniert werden, um die Effizienz steigern zu können.

Ein IR-Modell besteht aus vier Teilbereichen: einer Menge von Dokumenten, einer Menge von Queries, einem Framework für Dokumentenrepräsentation - dem Index - und einem Verfahren, das aus Queries und Dokumenten ein Ranking erstellt.¹⁵⁶ Als Framework versteht man die nach STOCK¹⁵⁷ und HIEMSTRA¹⁵⁸ angeführten Prinzipien, die in IR-Modellen formalisiert sind. Auch bei BAEZA-YATES et al. findet sich ein Kapitel zu den in der Taxonomie in Abbildung 6 angeführten Modellen, die einen ersten Überblick geben soll.¹⁵⁹ Je nach Dokumentart, ob textuelles oder multimediales Dokument bzw. ob das Dokument einen Webinhalt darstellt, bedarf es unterschiedlichen Herangehensweisen, nach welchem Regelwerk die Dokumente strukturiert, geordnet und gewichtet werden.

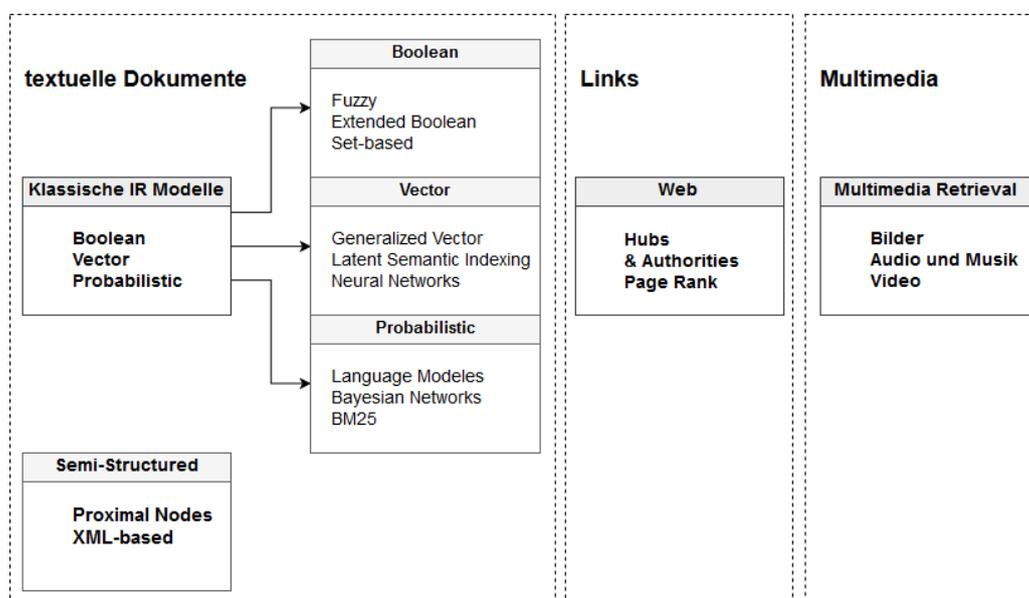


Abbildung 6: Taxonomie der IR-Modelle

¹⁵⁵FUHR: Modelle im Information Retrieval, S.321.

¹⁵⁶Vgl. BAEZA-YATES et al.: Modern information retrieval. The concept and technology behind search, S.57-58.

¹⁵⁷Vgl. STOCK: Information retrieval: Informationen suchen und finden, S.102-103.

¹⁵⁸Vgl. GOKER/DAVIES: Information retrieval: searching in the 21st century, S.1-15.

¹⁵⁹Siehe BAEZA-YATES et al.: Modern information retrieval. The concept and technology behind search, S.58-131.

Neben den klassischen Modellen zur Aufbereitung von textuellen Dokumenten gibt es für Multimedia- und Webdokumente eigene Verfahren. Beispielsweise zählt der **PageRank Algorithmus**, der Rankings auf Basis der Verlinkung von Dokumente im Web gewichtet, zu solchen Retrieval-Modellen. Das nun folgenden Kapitel widmet sich jedoch klassischen IR-Modellen für textuelle Dokumente, da sich mit diesen die grundlegenden Prinzipien von Retrieval-Modellen beschreiben lassen.

3.4.1 Boolesches Modell

Mit dem Modell des Booleschem Retrieval werden Dokumente beim Prozess des Indexierens auf einfache Indexterme abgebildet, die weiter nicht gewichtet werden. Diese Terme können nun mit einer Suchanfrage mittels Booleschen Operatoren (AND, OR, NOT) verknüpft werden. Die erzielten Ergebnisse werden nicht weiter in einem Ranking gereiht. Die Suche nach 'Gams' AND ('Graz' OR 'Wien') könnte theoretisch, wenn Dokumente mit diesen Textfolgen indexiert wurden, alle Karteikarten finden, die 'Gams' und 'Graz' oder 'Wien' beinhalten. Folgende Abbildung visualisiert die möglichen Ergebnisse dieser logischen Operatoren. 'AND'-Verknüpfungen geben nur Dokumente zurück, in denen sich Indextermen überschneiden. Die 'OR'-Verknüpfung ist als eine Vereinigung von Dokumenten zu betrachten, wohingegen ein 'NOT' alle Dokumente von einem Suchprozess ausschließt, die diesen Term beinhalten.

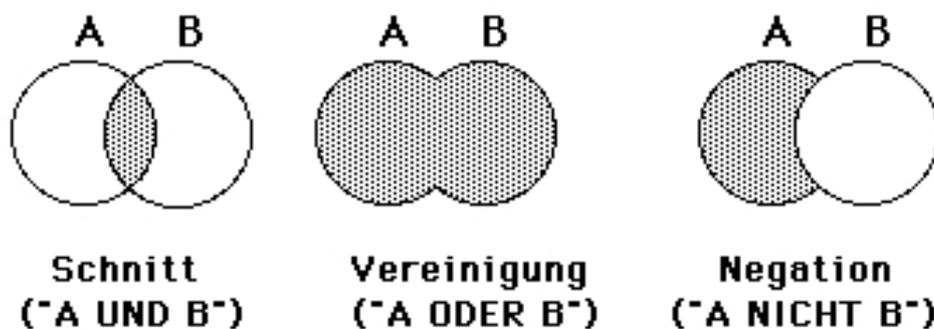


Abbildung 7: Darstellung logische Operatoren für das Boolesche Modell

Diese Art der Suche setzt bei der Nutzerin Vorkenntnisse über die Suchbegriffe voraus. Der Verweis auf ein Dokument alleine zeigt einer Nutzerin nicht an, wie relevant das Ergebnis zur Befriedigung ihres Informationsbedürfnisses ist, sondern nur, dass es damit indexiert wurde. Auch die Handhabung der logischen Operatoren muss richtig eingesetzt werden, da beispielsweise AND-Verknüpfungen nur Ergebnisse liefern, wenn alle Bedingungen erfüllt sind. Ein 'Laie' könnte eventuell unwissend vorhandene Dokumente in seiner Query ausschließen, obwohl sie wichtig wären. Ein weiteres Problem ist mit Synonymen verbunden. Das lateinische Wort '*Rupicapra rupicapra*' ist der wissenschaftliche Begriff für 'Gams'. Eine Suche in diesem

Modell nach dem Term 'Gams' findet keine Dokumente in denen 'Rupicapra rupicapra' synonym für 'Gams' verwendet wird, obwohl sie zur Befriedigung des Informationsbedürfnisses relevant gewesen wären.

Die Vorteile des Booleschen Modells sind besonders dann wahrnehmbar, wenn der Informationskorpus überschaubar ist, ein kontrolliertes Vokabular existiert und der Nutzer über ausreichend Erfahrung im Umgang mit logischen Operatoren verfügt, sowie eine tendenziell einfache technische Umsetzung dieses Modells. Auf der anderen Seite ist es tendenziell, als ein Retrieval von Daten zu sehen, als ein Retrieval von Information, da die wiedergewonnenen Dokumente als unkommentiertes Ergebnis, ohne Bewertung der Relevanz und ohne das Auflösen semantischer Unschärfe (z.B. Auflösung von Synonymen) vorliegen.

Das **Erweiterte Boolesches Retrieval** versucht diesen Problemen zu begegnen. Die binäre Natur der Booleschen Logik wird aufgehoben und stattdessen werden Werte eingesetzt, die sich zwischen wahr und falsch bewegen können. Damit verbunden ist das **Fuzzy-Retrieval**, mit dem man in der Lage ist auch unscharfe Mengen zu beschreiben. In diesem Ansatz werden die gleichen Strukturen, wie beim Booleschen Retrieval, verwendet. Erweitert wird es um einen gewichteter Index auf dem Intervall $[0,1]$, wodurch ein echtes Ranking entsteht. Die oben angeführten Beispiele könnten folgende Gewichtungen (aus einem automatisierten oder manuellen Verfahren heraus) in einem Dokument erhalten: 0.9 Gams, 0.5 Graz, 0.3 Wien. Für die Abfrage 'Gams' AND ('Graz' OR 'Wien') ergibt sich dann aus $\min(0.9, \max(0.5, 0.3))$ ¹⁶⁰ der Wert 0.5 für die Gewichtung des Terms 'Gams'. Aus diesen Werten kann sich ein einfaches Ranking abbilden, wobei FUHR anführt, dass die Qualität des Ergebnisses dieses Retrieval-Modells relativ schlecht ist da der Nutzer vollkommen in diesem Prozess ausgeblendet wird und Relevanz nur auf Basis von Dokumenten bestimmt wird.¹⁶¹

3.4.2 Vektorraummodell

Viele Dokumente im Internet sind nicht mit einem kontrollierten Vokabular verschlagwortet, im Gegensatz zum Bestand in einer Bibliothek. Das Vektorraummodell bietet eine Lösung an, Dokumente nur auf Basis der darin vorkommenden Terme zu finden und nutzt zu diesem Zweck geometrische Interpretationen. Das Vorkommen eines einzelnen Wortes wird in allen Dokumenten gezählt. Daneben existiert ein Index aller vorhandener Wörter und den Verweis in welchem Dokument sie vorkommen. Folgendes Beispiel soll dieses Modell veranschaulichen: Es gibt drei Dokumente, die als Vektor dargestellt die Anzahl der Wörter im Dokument beschreiben. Die Häufigkeit der Wörter 'Gams' und 'Graz' wird in den Vektoren D1, D2 und D3 und angezeigt:

D1(2,3) 'Gams' kommt zweimal vor und 'Graz' dreimal.

¹⁶⁰ $\min()$ bildet das 'AND', $\max()$ das 'OR'.

¹⁶¹ Vgl. FUHR: Modelle im Information Retrieval, S.321-322.

D2(0,2) 'Gams' kommt nicht vor und 'Graz' zweimal.

D3(3,2) 'Gams' kommt dreimal vor und 'Graz' zweimal.

Nun wird die Query auch als Vektor dargestellt, wobei nur einmal 'Gams' im Vektor vorkommt, also **Q(1,0)**. Dies entspricht einer einfachen Suche nach dem Term 'Gams'. In Abbildung 8 ist das Dokument von größter Relevanz, dessen Winkel zu Q am Kleinsten ist. D3 und D1 werden als Treffer angezeigt, wobei D3 im Ranking den ersten Platz einnimmt. D2 würde als Suchergebnis wegfallen. Würde man weitere Worte abfragen wollen würden schlichtweg weitere Dimensionen im Vektorraum hinzukommen.

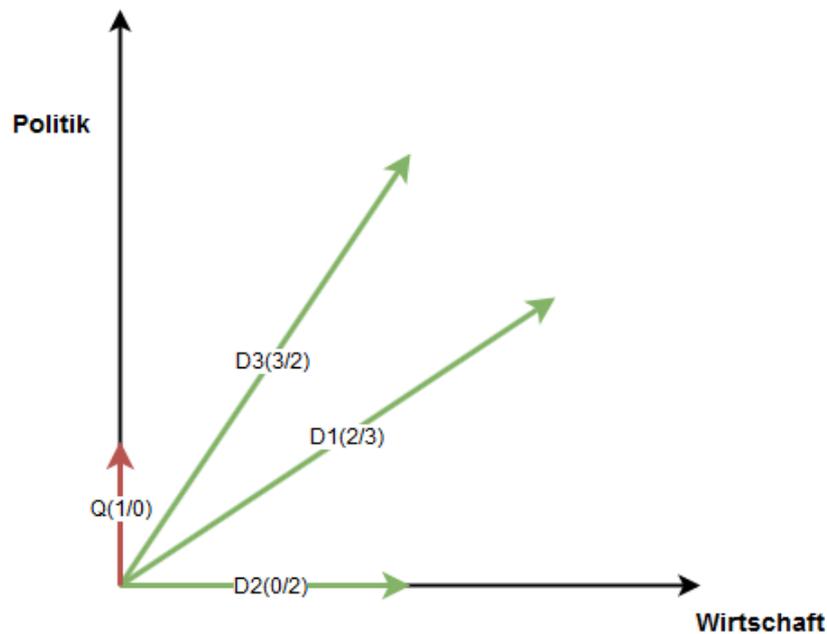


Abbildung 8: Beispiel für das Vektormodell

Die Vorteile des Vektormodells sind eine nicht binäre Gewichtung und Reihung der Suchtreffer nach möglicher Relevanz. Probleme treten bei sehr großen Dokumenten auf, bei Synonymen oder Homonymen und Dokumenten mit ähnlichem Inhalt, aber anderem Wortschatz, auf. Die Qualität dieses Retrieval-Modells ist hoch und viele Web-Suchmaschinen basieren auf diesem Modell.

Das **Latent Semantic Indexing**¹⁶² ist ein auf diesem Modell beruhendes Variation, in dem die Abhängigkeit zwischen Termen zusätzlich berücksichtigt wird. Eine andere Variation wird mittels **neuronalen Netzen**¹⁶³ umgesetzt, die die Vektoren und ihre Abhängigkeiten maschinell erzeugen können.¹⁶⁴

¹⁶²Vgl. DEERWESTER, Scott et al.: Indexing by latent semantic analysis. Journal of the American society for information science, 41 1990, Nr. 6, 1-27.

¹⁶³Vgl. KENDAL/CREEN: An Introduction to Knowledge Engineering, S.36-54.

¹⁶⁴Vgl. FUHR: Modelle im Information Retrieval, S.322-325.

3.4.3 Probabilistische Modell

Die bis jetzt angeführten IR-Modelle berechnen wie ähnlich Suchanfrage und Dokumentenbeschreibungen sind, ohne Aussagen darüber zu tätigen, inwieweit eine hohe Qualität des Retrievals erreicht werden kann. Aufgrund der natürlichen Sprache mit der ein Query formuliert werden muss, kann nicht sicher festgestellt werden, ob ein Dokument relevant ist oder nicht. Ausgehend von der Frage wie groß die Wahrscheinlichkeit ist, dass ein Dokument für eine Anfrage als relevant eingeschätzt wird, wird mittels Vereinfachungen, Unabhängigkeitsannahmen und der Häufigkeit von Termen in einem Dokument, ein Wert berechnet. An Hand dieses Wertes wird ermittelt wie ähnlich ein Dokument der Suchanfrage ist. Je ähnlicher Anfrage und ein Dokument sind, desto höher ist dieser Wert und desto höher wird das Dokument in der List der Suchtreffer gereiht, wobei es erst einen gewissen Schwellenwert zu überschreiten gilt, um angezeigt zu werden. Während diesem Vorgang kann es Möglichkeiten für die Nutzerin geben mit einem Feedback zu steuern, wie relevant ein Dokument war. Diese gewonnenen Informationen können genutzt werden, um den Retrieval-Prozesse für Nutzerinnen zu verbessern. Es gibt eine Vielzahl von probabilistischen Verfahren. Eines der bekanntesten ist das **Retrievalmodell mit binärer Unabhängigkeit**.¹⁶⁵ Als Vorteile in diesem Modell kann die Loslösung des Users von Gewichtung der Suchbegriffe, Möglichkeit des Rankings, eine natürlichsprachliche Eingabe der Suchanfrage ohne Operatoren, angeführt werden. Dennoch führt die Literatur an, dass das probabilistische Modell einen experimentellen Charakter aufweist und in der Praxis nicht oder nur bedingt umgesetzt werden kann.¹⁶⁶

3.5 Fazit

Wie Kapitel 3.2 zeigt, werden die Begriffe Daten, Information und Wissen unterschiedlich definiert. Zusammenfassend kann gesagt werden, dass Wissen alle Kompetenzen und Konzepte eines kognitiven Agenten über die Wirklichkeit zu einem bestimmten Zeitpunkt umfasst. Die Interpretation der wahrgenommen Wirklichkeit - sprich Information - führt zu neuer Erkenntnis und neuem Wissen, das über einen gewissen (subjektiven) Wahrheitsgehalts verfügt. Suchprozesse sind stark an den Erkenntnisprozess gekoppelt. Um Suchprozesse und somit Erkenntnisgewinn zu fördern werden Retrieval Modelle und Methoden entwickelt, um das Informationsbedürfnis eines kognitiven Agenten zu befrieden, in dem alle relevanten Dokumente in einem Suchszenario geliefert werden können. Dabei wird der Prozess des Retrievals vom Informationsbedürfnis, der Formulierung der Suchanfrage, der Ordnung und Strukturierung des Informationskorpus, dem Abgleich zwischen Suchanfrage und Dokumentspeicher, die Ausgabe und das Feedback des Ergebnisses durchlaufen. Alle so eben genannten Teilaspekte des Retrieval-Prozesses wurden in diesem Kapitel vertiefend erörtert. Das Boolesche Modell, das Vektorraummodell und das probabilistische Modell beschreiben theoretische Lösungsansätze, um die Herausforderung des

¹⁶⁵BIRKNER, Michael et al.: Guideline zur Langzeitarchivierung. 2016.

¹⁶⁶Vgl. FUHR: Modelle im Information Retrieval, S.325-327.

IR umsetzen zu können. Sie formalisieren Strukturen, um den Abgleich zwischen Suchanfrage und Informationskorpus zu gewährleisten und ermitteln die Ordnung, sowie Relevanz bzw. Reihung der Dokumente.

So sind grundlegende Aspekte des IR und der durch die Modelle gegebenen Überlegungen zur Umsetzung auch notwendig für die praktische Implementierung von Discovery-Funktionalitäten. Auch dort muss die Relevanz von Objekten, ihre Ordnung und Struktur formalisiert und ermittelt werden. Das folgende Kapitel widmet sich weiteren RD-Prinzipien, die über diese Grundlagen hinausgehen, sich aber dennoch auf diese berufen.

4 Stöbern: Resource Discovery

4.1 Einleitung

*"Resource Discovery is more than just search: it is browsing, searching, selecting, assessing and evaluating."*¹⁶⁷

Resource Discovery ist das Kennenlernen, Stöbern und Entdecken von Informationsressourcen in einem Informationskorpus. Ziel ist es IR-Prozesse zu verbessern und eine Möglichkeit zu schaffen einen Bestand an Objekten kennen zu lernen und womöglich zusätzliche Informationsressourcen zu erschließen, die zuvor noch nicht bekannt waren, aber relevant sind. Vor allem große Suchräume oder Bestände mit untypische Ressourcen, wie beispielsweise die Objekte des Kriminalmuseums, konfrontieren den Nutzer damit, nicht zu wissen was er überhaupt suchen kann. Das Informationsbedürfnis ist vage und dem Nutzer fällt es schwer eine Suchanfrage zu formulieren oder mit den erhaltenen Suchergebnissen effizient umzugehen. Auch die Handhabung von Retrieval-Systemen ist unterschiedlich. Retrieval-Systeme von Bibliothekskatalogen beispielsweise, unterscheiden sich von kommerziellen Websuchmaschinen, da sie ihre Informationsressourcen stärker auf inhaltlicher Ebene strukturieren und die Suche mit einzelnen miteinander kombinierbaren Begriffen ermöglichen. Werden Begriffskombinationen falsch verknüpft, so führt die Suche zu keinen Treffern. Laut WIESENMÜLLER und PFEFFER können Resource Discovery Systeme Verbesserungen für Bibliothekskataloge mit sich bringen.¹⁶⁸ Ein weitere Gedanke hinter Discovery-Prozessen besteht auch darin, dass ein Informationsbedürfnis nicht auf einzelne Dokumente abzielt, sondern auf interne und externe Kontexte, in die die Dokumente eingebunden sind. Semantische Technologie bieten ein Werkzeug, um solche Verbindungen sichtbar, verwertbar und explorativ zu machen, indem nach Bedeutung und nach den Kontexten, aus denen heraus Dokumente stammen, gesucht werden kann. Aus diesem Grund widmet sich Kapitel 5 der Ausführlichen Dokumentation der Umsetzung des Studienprojektes auf Basis semantischer Technologien.¹⁶⁹ Folgende, sich überschneidenden Zugänge zur Umsetzung des RD, werden in diesem Kapitel behandelt:¹⁷⁰

- **User-Interaktion** mit einem Informationskorpus.
- **(Selbst)Organisation von Informationsressourcen.**
- An die Ressourcen angepasste **Informationsvisualisierung.**
- **Semantische Suche**

¹⁶⁷RÜGER: Multimedia Resource Discovery, S.39.

¹⁶⁸Vgl. WIESENMÜLLER, Heidrun/PFEFFER, Magnus: Resource Discovery Systeme. In STRAUB, Martina/SÜHL-STROHMENGER, Wilfried (Hrsg.): Handbuch Informationskompetenz. 2. Auflage. Berlin, 2016, S.105-107.

¹⁶⁹Vgl JORDANOUS: Enhancing information retrieval and resource discovery from data using the Semantic Web, S.105.

¹⁷⁰Vgl. GESER: Resource Discovery - Position Paper: Putting the Users First, S.9-12.

4.2 Ein erweitertes Informationsbedürfnis: Exploratory Search

Als **exploratives Verhalten** versteht man die Bereitschaft eines kognitiven Agenten die Umwelt wahrzunehmen und Schlüsse daraus zu ziehen - sie zu erkunden. DÖRNER sieht eine enge Verbindung zwischen der Fähigkeit des Menschen Probleme zu lösen und seinem Verhalten die Welt bzw. die Wirklichkeit kennen lernen zu wollen. Die kognitive Landkarte, so DÖRNER, wird breiter, je mehr Informationen verarbeitet werden bzw. je mehr Erkenntnis aus der wahrnehmbaren Wirklichkeit gewonnen werden können. Ein Mehr an Information führt dazu, dass bekannte Areale besser in Erinnerung bleiben, das wiederum die Motivation neues zu Entdecken erhöht.¹⁷¹

Information Lookup und **Exploratory Search** - sprich entdeckendes Suchen - beschreiben zwei unterschiedliche Formen eines Informationsbedürfnisses und seiner Befriedigung. Folgendes Beispiel soll das komplexere Informationsbedürfnis, sowie den Prozess der *Exploratory Search* beschreiben:

Eine Forscherin hat eine soziohistorische Forschungsfrage zu Wilderei-Delikten und gibt geeignete Begriffe in eine Suchmaschine. Dabei stößt sie auf die Person Hans Gross, eine zuvor unbekannte Person. Die Kenntnis über diese Person hat die kognitive Landkarte der Forscherin erweitert. Die Information wird aufgenommen, verarbeitet und in eine neue Suchanfrage umformuliert. Das Ergebnis dieses kognitiven Prozesses ist ein URL zur digitalen Sammlung des Hans Gross Kriminalmuseums. Nach Verwendung der dort umgesetzten Discovery-Möglichkeiten¹⁷² entdeckt sie neue Begriffe und neue Dokumente mit neuen Informationen, die ihr helfen, ihre Suchanfragen anzupassen, um für sie relevante Karteikarten zu finden.

WHITE und ROTH sehen im von Kognitionsprozessen geprägten entdeckenden Suchen Überschneidungen, aber auch erweiternde Aspekte des klassischen Informationsbedürfnisses, die sie in Darstellung 9 veranschaulichen. Den statischen Prozessen des Navigierens, Verifizierens und Faktensammelns, stehen die investigativen Prozesse des Analysierens, Evaluierens und Interpretierens gegenüber. Gemeinsam haben beide Formen, dass Lernprozesse angestoßen werden, in denen Informationen mit dem 'internen Außenweltmodell' des kognitiven Agenten abgeglichen werden.¹⁷³ Entwickelt sich neues Wissen beim kognitiven Agenten auf Grund positiver Effekte, die mit neu gefundener und genutzter Information einhergehen, wird, nach WERSIG, das 'interne Außenweltmodell' um neue Regeln und 'Programme' erweitert. Diesen Prozess kann man als 'Lernen' bezeichnen, der jeder Kognition im Sinne eines explorativen Suchprozesses immanent ist.

¹⁷¹Vgl. DÖRNER, Dietrich: Vom Umgang mit Unbestimmtheit und Komplexität. Bern, 1983.

¹⁷²RD im Bestand der Wildererkarteikarten, <http://gams.uni-graz.at/archive/objects/query:km.wildererkk/methods/sdef:Query/get?mode=query>, 03.03.2017.

¹⁷³Vgl. WHITE, Ryan W/ROTH, Resa A: Exploratory search: beyond the query-response paradigm (Synthesis lectures on information concepts, retrieval & services). Morgan and Claypool Publishers 3 2009, S.9-23.

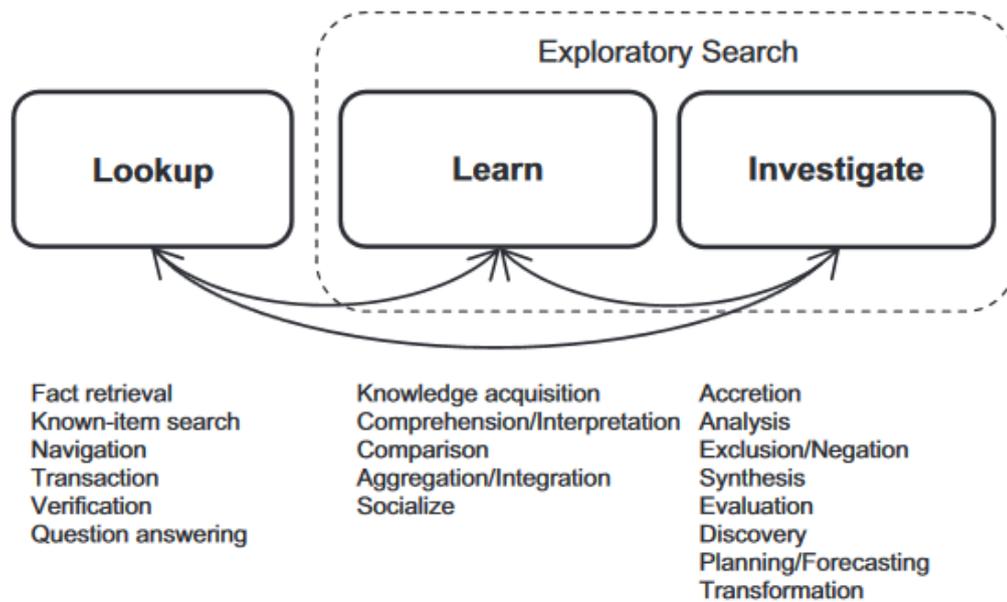


Abbildung 9: Darstellung Information Lookup vs. Exploratory Search

Folgende Zugänge zur Umsetzung von RD-Funktionalitäten können einem explorativen Informationsbedürfnis entsprechen.

4.3 User-Interaktivität: Als Hilfestellung zur Query-Formulierung

Interaktivität prägt das Web. Nutzer von Suchmaschinen sind in ihrer Mehrzahl 'Laien'. Ihre Informationskompetenz ist gering, ohne dies wertend zu verstehen. Aus diesem Grund sind komplexere Informationsbedürfnisse für diese Personen eine große Herausforderung, da sich auch die Formulierung einer zielführenden Suchanfrage als komplex erweist. JOEMON fasst diese Problemstellung für einen User treffend mit dem Satz *'I don't know what I'm looking for, I'll know it when I find it.'*¹⁷⁴ zusammen. Als User-Interaktivität werden in dieser Arbeit alle Handlungen verstanden, die ein User durchführen kann, um Informationsressourcen zu entdecken und einen Informationskorpus besser kennenlernen zu können.

Eine Studie von JANSEN et al. hat in einem kurzen Zeitraum 3 Millionen Einträge im Log-Protokoll der Suchmaschine *AltaVista* empirisch ausgewertet und dabei festgestellt, dass 52% aller Suchanfragen von Nutzern mindestens einmal neu formuliert wurden.¹⁷⁵ Der Zugang die Query immer wieder neu zu adaptieren, um die Ergebnisse zu verbessern, ist eine gängige Art im Sinne des explorativen Suchens und wird als **Query Reformulation** bezeichnet.

HEARST führt aus, dass es unterschiedliche Zugänge seitens Retrieval-Systemen geben soll, um diese Vorgehensweise von Nutzern zu unterstützen und gleichzeitig Werkzeuge des Ent-

¹⁷⁴JOEMON, Jose: Personalisation Techniques in Information Retrieval. In GESER, Guntram/PEREIRA, John (Hrsg.): DigiCULT Thematic Issue 6 – Resource Discovery Technologies for the Heritage Sector. Salzburg and Glasgow, 2004, S.22.

¹⁷⁵Vgl. JANSEN, B.J./SPINK, A./PEDERSEN, J.O.: A Temporal Comparison Of Altavista Web Searching. Journal Of The American Society For Information Science And Technology, 56(6), 2005, S.559–570.

deckens von Informationsressourcen anzubieten. Abgesehen von Rechtschreib- oder Tippfehlern, die von Retrieval-Systemen ausgebessert werden müssen (**Spelling Suggestion**), sieht man in Abbildung 10, wie Google auf Suchanfragen dynamisch reagiert.



Abbildung 10: Screenshot einer Term Suggestion bei Google

Die Eingabe des Suchbegriffs 'information' bei Google lieferte neben der Eingabe die Wörter 'retrieval', 'synonym', 'technology' und 'and computer engineering'. Dahinter stecken Begriffe, die Google nach bestimmten Kriterien mit der Eingabe 'information' in Verbindung gesetzt hat. Diese **Term Suggestions** erweitern und verfeinern Suchergebnisse während des Suchprozesses und geben auch eine Übersicht darüber, was im Bestand vorhanden ist oder was andere Nutzer bereits gesucht haben. Weiters könnten vernetzte Konzepte als Sinneinheiten für semantische Technologien damit verbunden sein. Während der Interaktivität des Eingebens der Suchanfrage in den Suchschlitz, erhält der Nutzer ein dynamisches Feedback, das ihm beim Formulieren seiner Suchanfragen unterstützt. HEARST führt mehrere Studien an, die zeigen, dass automatisierte Empfehlungen für Rechtschreibung und für Begriffe in ungefähr 35% aller Fälle von Nutzerinnen verwendet werden. Dabei ist vor allem die interaktive Komponente für die Nutzerin hervorzuheben.¹⁷⁶

Die Einbindung des Nutzers im Suchprozess ist wichtig, und, wie in Kapitel 3.4 gezeigt, für klassische Retrieval-Modelle oft nicht möglich. Als **Relevance Feedback** versteht man ein Verfahren zur schrittweisen Verbesserung von Suchergebnissen eines Suchprozesses. Oft bestehen Suchanfragen nur aus 2 Wörtern. Weiters können Wortkombinationen mehrdeutige Korrelationen ergeben, die zu unscharfen und nicht relevanten Suchergebnissen führen. Über die Ergebnisse vorangegangener Suchen und dem Feedback eines Nutzers, ob Dokumente relevant oder nicht relevant waren, wird die Suchanfrage erneut durchlaufen. Das Feedback wird im neuen Suchprozess berücksichtigt und führt zu einem anderen Ergebnis. Nach mehreren Iterationen können so immer weitere relevante Dokumente gefunden werden. Ein Nachteil ist der damit verbundene Aufwand seitens der Nutzerin das Feedback, als user-seitige Interaktion, zu geben.¹⁷⁷ Für **Explicit Feedback Systems** wird das Verhalten eines Nutzer beobachtet.

¹⁷⁶Vgl. HEARST: Search User Interfaces, S.142-156.

¹⁷⁷Vgl. BÜTTCHER, Stefan/CLARKE, Charles LA/CORMACK, Gordon V: Information retrieval: Implementing

Die Augenbewegung, die Verweildauer auf einer Seite, Ort und Anzahl der Mausklicks werden statistisch erhoben und ausgewertet. Auf Basis solcher Auswertungen - Suchtreffer, die öfters angeklickt werden sind relevanter - wird die Relevanz eines Dokumenten berechnet.¹⁷⁸ **Implicit Feedback Systems** wiederum verwerten Information, die ein Nutzer bewusst eingibt oder als relevante oder nicht relevante Information kennzeichnet. Ein Beispiel dafür sind die Nutzerbewertungen mittels 'Sternen' auf Amazon. Dokumente, die viele und bessere Bewertungen haben, erhalten auch eine höhere Relevanz.¹⁷⁹



Abbildung 11: Screenshot einer Recommender-Empfehlung bei Amazon

Eine weitere Möglichkeit ist die Empfehlung von anderen Informationsressourcen, wie es bei Amazon auf Basis eines **Recommender Systems** vorzufinden ist. Diese Empfehlungsdienste berechnen an Hand der Suchprofile der Einzelpersonen und dem Kollektiv, welche Dokumente im Zusammenhang mit anderen Dokumente relevant erscheinen. Amazon nutzt diese Dienste, um das Interesse ihrer Nutzer für angebotene Waren zu wecken.¹⁸⁰ Recommender Systeme haben einen starken Einfluss auf ökonomische Entwicklungen, in deren Zentrum die gesteuerte Auswertung und Verteilung von Information steht. Diese Überlegungen wurden in Kapitel 2.3 zur gesellschaftlichen Perspektive diskutiert.

Eine weitere User-Interaktivität, die RD fördert ist das **Browsing**, das BAEZA-YATES et. al. als ein '*discovery paradigm*' bezeichnen. Die Darstellung von Thumbnails und von basalen Metadaten zu einem Dokument auf einer zweidimensionalen Ebene - dem **Flat Browsing** - erlaubt es einem Nutzer einen Überblick über eine Menge von Dokumente zu erhalten. Auch sind stärker strukturierte Formen des Browsings denkbar, in der z.B. die hierarchische Strukturen herangezogen werden. Der 'Laie' hat so die Möglichkeit einen Bestand kennen zu lernen, ohne eine Abfrage formulieren zu müssen, der Experte wiederum kann eventuell mit seiner pro-

and evaluating search engines. Cambridge and London, 2016, S.273-275.

¹⁷⁸Vgl. BAEZA-YATES et al.: Modern information retrieval. The concept and technology behind search, S.185-188.

¹⁷⁹Vgl. ebd., S.195-200.

¹⁸⁰Vgl. JANNACH, Dietmar et al.: Recommender systems: an introduction. Cambridge, 2010, S.1-6.

funden Kenntnis des Bestandes und der Inhalte bereits am Thumbnail oder an den Metadaten entscheiden, ob die Informationsressource für ihn relevant ist oder nicht.¹⁸¹

4.4 Organisation von Suchergebnissen: Als strukturgebende Methode

Die Möglichkeit Suchergebnisse zu ordnen und zu strukturieren bildet ein weiteres Themenfeld für das RD. Eine Studie von KÄKI zeigt, dass Nutzer ihr exploratives Suchverhalten ändern, wenn sie ihre Suchergebnisse selbst organisieren können.¹⁸² Die Organisationsformen der Facettenkategorien und des Clusterings gelten als zwei etablierte Vorgehensweisen in diesem Zusammenhang. Ein **Kategorisierungssystem** ist ein Ordnungssystem, in dem jede Kategorie über einen Begriff definiert ist. Jedes Objekt kann einer solchen Kategorie zugeordnet werden. Die am öftesten verwendeten Strukturierungen für Kategorisierungen sind flache, hierarchische und facettierte Formen. Abbildung 12 zeigt ein Beispiel für eine hierarchische Struktur in Form einer einfachen Liste von Begriffen, die weitere Unterbegriffe aufweist.

- Schusswaffe
 - Bogen
 - Feuerwaffe
 - Kurzwaffe
 - Pistole
 - Revolver
 - Langwaffe
 - Flinte
 - Gewehr
 - Selbstschussgerät

Abbildung 12: Ausschnitt aus dem SKOS-Objektthesaurus des Kriminalmuseums

In dieser Darstellung wird die hierarchische Struktur von Objektkategorien, die zur Erschließung des Bestandes im Kriminalmuseum verwendet wurden, dargestellt. Die Kategorie *Schusswaffen* fasst zwei Unterkategorien, *Bogen* und *Feuerwaffen*, zusammen. So können Suchtreffer Schritt für Schritt eingeschränkt werden. Jedoch ist es nicht mehr möglich von Unterkategorien, z.B. *Gewehr*, auf andere Unterkategorien eines anderen Teilbaums zuzugreifen, z.B. *Pistole*, und diese in einer Suchanfrage zusammen zu führen. Auf der anderen Seite kann dies auch als gern gesehenes Ergebnis betrachtet werden, da so Suchergebnisse eingeschränkt und dadurch besser für den kognitiven Agenten fassbar werden.

Die **facettierte Begriffsordnung**, als polyhierarchisches Klassifikationssystem, erlaubt es mehrere unabhängige Begriffszuordnungen vorzunehmen.¹⁸³ *Holz* und *Gewehr* sind unterschiedli-

¹⁸¹Vgl. BAEZA-YATES et al.: Modern information retrieval. The concept and technology behind search, S.498-500.

¹⁸²Vgl. KÄKI, Mika: Findex: search result categories help users when document ranking fails. In Proceedings of the SIGCHI conference on Human factors in computing systems. 2005.

¹⁸³Vgl. STOCK/STOCK: Wissensrepräsentation: Informationen auswerten und bereitstellen, S.273-281.

che Kategorien, aber sie könnten als geisteswissenschaftliches Quellenmaterial im Kontext einer musealen Repräsentation, hinsichtlich ihrer Gemeinsamkeiten bezüglich Material, Zeit oder Ort von Interesse sein. Definiert man für Materialien, Entstehungszeiten oder Entstehungsorte - also Objekt beschreibende Metadaten - jeweils Facetten, so ermöglicht dies eine facettierte Darstellung. Innerhalb einer Facette werden Selektionen mit 'OR', über Facetten hinweg mit 'AND', verknüpft. So kann ein Suchergebnis sowohl *Holz*, wie auch *Gewehr* enthalten, sofern sie diesen Facette zugeordnet ist.¹⁸⁴ Das Projekt 'Repositorium Steirisches Wissenschaftserbe' bietet eine facettierte Suche an.¹⁸⁵ Werden dort die Facetten Material mit *Holz*, Kategorie mit *Gewehre* und Sammlung mit *Hans Gross Kriminalmuseum* selektiert, so werden 214 Treffer¹⁸⁶ gefunden. Alle diese Treffer sind Objekte aus dem Kriminalmuseum, eine Feuerwaffe und alle bestehen auch aus dem Material Holz. Über weitere Facetten kann so der Suchraum spezifischer eingegrenzt oder erweitert werden. Die technische Umsetzung der facettierten Suche in diesem Beispiel erfolgt mit den Open Source Produkten **Apache Lucene**¹⁸⁷ und **Solr**, die Softwareprodukte zur Indexierung von Dokumenten darstellen.¹⁸⁸

Eine weitere denkbare Umsetzung einer Selbstorganisation von Suchergebnissen ist das Speichern von Informationsressourcen in einem für die Nutzerin verfügbaren **Datenkorb**, wie es etwas ein 'Warenkorb' bei Amazon ist. Gefundene relevante Treffer können so gesammelt und in einer zweiten Auswahl wieder entfernt werden und anschließend für lokale Nutzung exportiert werden.¹⁸⁹

4.5 Informationsvisualisierung: Als Discovery-Strategie

Die Visualisierungsmöglichkeiten von Daten oder Information sind so unterschiedlich wie die Ressourcen selbst. Ressourcen können als Bilder, Symbole, Videos, Objekte in einem Museum oder als Digitalisat auftreten. Jede Darstellung in Retrieval-Prozessen ist an die Visualisierung der Inhalte in einer zweidimensionalen Welt gebunden: auf Papier und Bildschirm. Dennoch, so führt TUFTE an, gibt es eine große Bandbreite an Möglichkeiten Information in dieser zweidimensionalen Welt effektiv darzustellen.¹⁹⁰ Die Möglichkeiten kann man als Informationsvisualisierung zusammenfassen.

'Information visualization is the use of computer-supported, interactive, visual representati-

¹⁸⁴Vgl. BAEZA-YATES et al.: Modern information retrieval. The concept and technology behind search, S.35-48.

¹⁸⁵Beispiel einer facettierten Suche, <http://gams.uni-graz.at/context:pardus>, 05.02.2017.

¹⁸⁶Facettierte Suche nach Gewehren im Kriminalmuseum, http://gams.uni-graz.at/archive/objects/context:pardus/methods/sdef:Context/get?mode=search&locale=de#/?q=*&dc_type=WaffeundMunition&dc_medium=Holz&dc_isPartOf=HansGrossKriminalmuseum&x=0,04. 03.2017.

¹⁸⁷Lucene, <http://lucene.apache.org>, 05.02.2017.

¹⁸⁸Vgl. WIESENMÜLLER/PFEFFER: Resource Discovery Systeme, S.107-109.

¹⁸⁹Beispiel eines Datenkorbs in einer digitalen Edition, <http://gams.uni-graz.at/context:srbas>, 07.04.2017

¹⁹⁰Vgl. TUFTE: Envisioning Information, S.9-13.

*ons of abstract data in order to amplify cognition.*¹⁹¹

Informationsvisualisierung hilft dabei, die menschliche Kognition zu verbessern, indem Muster, Zusammenhänge und Strukturen, mit Hilfe von computergestützten Methoden, visuell und interaktiv aufbereitet werden, damit ein kognitiver Agent in die Lage versetzt wird, Information besser verarbeiten zu können.¹⁹²

Eine entsprechende Darstellung von Information ist sehr gut dafür geeignet, um Discovery-Prozesse zu unterstützen und umzusetzen. Es kann den Einstieg in eine Suche erleichtern, kann Veränderungen, Filterungen, Verdichtungen oder Erweiterungen von Suchergebnissen auf eine Art und Weise abbilden, das eine Nutzerin sich besser im Informationskorpus zu recht finden kann. Dargestellt werden kann eine räumliche, zeitliche, thematische oder strukturelle Dimension.¹⁹³

Diese Dimensionen stecken beispielsweise in den Karteikarten des Kriminalmuseums. Die praktische Umsetzung in Kapitel 5.4.4 widmet sich genau dieser Fragestellung. Die **räumliche** Dimension kann mit Karten dargestellt werden, die **zeitliche** mittels Abbildung von Prozessen realisiert werden. Ein Beispiel dafür ist die Darstellungen von Dokumenten nach einer durch eine Zeitstrahl gegebenen Ordnung, wie es der Dariah GeoBrowser oder der Zeitstrahl im praktischen Teil zeigt. Eine **thematische** Visualisierung ist mit der virtuellen Ausstellung des Wilderer-Bestandes im Kriminalmuseum, einem narrativen Discovery-Szenario, gegeben. Von dort aus existieren Sprungpunkte aus dem thematischen Feld der 'Wilderer' in die digitale Sammlung.¹⁹⁴ Ein Beispiel für eine **strukturelle** Visualisierung kann beispielsweise die Darstellung von Dokumenten und ihrer Relationen in einem Netzwerk umfassen. Die Ausgangslage eines solches Netzwerkes kann mittels semantischen Technologien modelliert wurde.

4.6 Semantische Suche: Als Suche nach Bedeutung

*'We use the term Semantic Search when semantics are used during any part of the phases in the search process'*¹⁹⁵

Das Wort Semantik ist ein Synonym für das Wort Bedeutung. Eine semantische Suche ist demzufolge eine Suche nach einer bestimmten Bedeutung. 'Klassisches' IR indexiert Text als 'Bag-of-Words' und gewichtet diesen Text mit Zahlen. In einer semantischen Suche wird etwas in

¹⁹¹CARD/MACKINLAY/SHNEIDERMAN: Readings in Information Visualization: Using Vision to Think, S.7.

¹⁹²Vgl. FORTMÜLLER: Informationsvisualisierung: Analyse und Evaluierung ausgewählter Applikationen zur interaktiven visuellen Exploration von Datenbeständen, S.5-9.

¹⁹³Vgl. JÄGER-DENGLER-HARLES, Ingeborg: Informationsvisualisierung und Retrieval im Fokus der Informationspraxis. 2013, S.149-152.

¹⁹⁴Virtuelle Ausstellung als thematischer Discovery-Einsprungpunkt, <http://gams.uni-graz.at/o:km.story>, 12.04.2017.

¹⁹⁵HILDEBRAND/VAN OSSENBRUGGEN/HARDMAN: An analysis of search-based user interaction on the semantic web, S.2.

einem bestimmten Kontext gesucht. Die Notwendigkeit dafür entsteht dann, wenn das gesuchte Dokument, einen höheren Grad an Komplexität aufweist.¹⁹⁶ Die Fähigkeit eines Retrieval-Systems zum '**Konzeptbewusstsein**', wie GRAUPMANN anführt, ermöglicht es einer Suchmaschine unterscheiden zu können in welchem semantischen Konzept sich ein Suchbegriff befindet. Mit einem 'Konzeptbewusstsein', kann eine Suchmaschine beim Indexieren erkennen, ob 'Rice' dem Kontext Nahrungsmittels zugeordnet werden soll, oder dem Kontext eines Namens und einer Person. Ein '**Abstraktionsbewusstsein**' einer Suchmaschine würde es weiters erlauben im Umfeld eines Kontextes suchen zu können. Die Suchanfrage '*Wilderei-Delikte des 19. Jahrhunderts*' sollte Karteikarten und Objekte aus den Jahren 1800-1899, die zum Kontext von Wilderei-Delikten gehören ausgeben. Semantische Suche imitiert auf gewisse Art und Weise die menschliche Erkenntnisfähigkeiten - darf aber nicht mit künstlicher Intelligenz gleichgesetzt werden - in dem Assoziationen und bereits vorhandenes Wissen in den Suchprozess durch eine Maschine miteinbezogen werden. Die Suchmaschine ließt nicht nur die Zeichenfolge '*Wilderei-Delikte des 19. Jahrhunderts*', sondern versteht das Konzept Wilderei, sowie das Konzept Delikt als einen juristischen Ausdruck und 19. Jahrhunderts als die Zeitangabe, die sie ist.

Das Konzeptbewusstsein bzw. das Hintergrundwissen, das ein Maschine braucht, um den Kontext miteinzubeziehen, kann aus einer Wissensrepräsentation gewonnen werden, wie sie in Abbildung 5 in Kapitel 3.3.4 angeführt werden. Das notwendige Wissen wird in einer für eine Maschine interpretierbaren Struktur formalisiert, wie etwa in einem Thesaurus oder in einer Ontologie, die wiederum in maschinenlesbaren Sprachen, wie etwa XSL serialisiert werden. Um eine Maschinen interpretierbare Form zu erreichen bedarf es standardisierter Technologien, wie RDF, SKOS oder OWL, wie sie in Kapitel 5.4.1 beschrieben und angewandt werden. Die Modellierung der Wissensdomäne ist dabei ein zentrales Problem, da eine manuelle Modellierung zeitintensiv, eine automatisierte Modellierung ein aktuelles Forschungsgebiet darstellt. GRAUPMANN beschreibt das Vorgehen der Sphere-Search-Suchmaschinen, ausgehend von einer semantischen Struktur in XML (hierarchische Baum Struktur). Auf Basis eines flexiblen Graphenmodells werden Suchanfragen berechnet und ausgewertet.¹⁹⁷ Dazu müssen Informationen extrahiert, in ein Modell gebracht und diese wiederum abgerufen und dargestellt werden.¹⁹⁸ Folgende Darstellung beschreibt die Architektur einer semantischen Suchmaschine, wie etwas Sphere-Search.

¹⁹⁶Vgl. YU, Liyang: Introduction to the semantic web and semantic web services. New York, 2007, S.24-34.

¹⁹⁷Vgl. GRAUPMANN: Semantische Suche: intelligente Suchmaschinen durch innovative und zukunftsweisende Konzepte und Technologien, S.9-16.

¹⁹⁸Vgl. DAVIS, John/DUKE, Alistair/KIRYAKOV, Atanas: Semantic Search. In GOKER, Ayse/DAVIES, John (Hrsg.): Information retrieval: searching in the 21st century. Chichester, 2009, S.179-181.

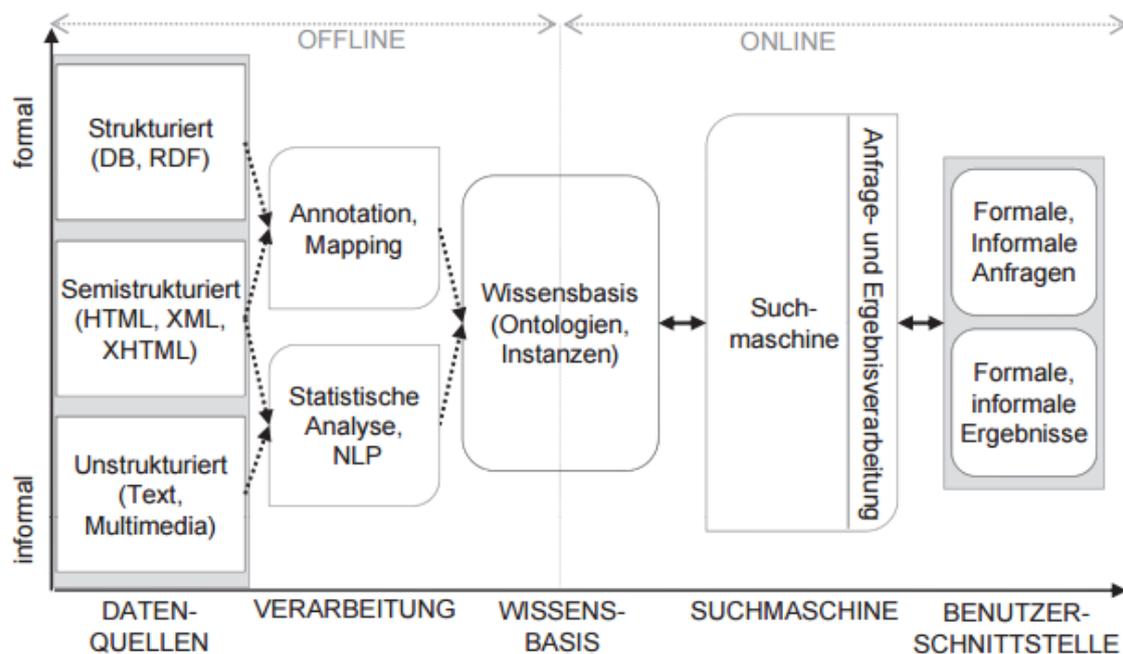


Abbildung 13: Architektur einer semantischen Suchmaschine

Eine Datenquelle, in der sich unstrukturierte (reiner Text), semistrukturierte (XML) oder strukturierte Dokumenten (RDF) befinden, wird erschlossen und verarbeitet. Dies kann mittels statistische Verfahren bzw. Annotation geschehen, um aus den Daten eine formale Struktur in Form einer Wissensbasis (z.B. Thesaurus oder Ontologie) zu generieren. Die Suchmaschine wird von einer Nutzerin mit einer Anfrage konfrontiert, die zu einer Query umformuliert wird, die in der Lage ist die Anfrage mit der Wissensbasis zu vergleichen und ein Ergebnis zurückliefert.¹⁹⁹

Dieser Zugang kann auch für RD-Funktionalitäten umgesetzt werden. Im **Semantic (Resource) Discovery** werden Dokumente, Webinhalte oder Web Services so aufbereitet, dass Such- und Discovery-Möglichkeiten nach Bedeutungen bzw. nach formalisierten Wissen existieren. Die Struktur der Wissensbasis an sich kann zur Visualisierung und Darstellung von Inhalten, wie etwa die Formalisierung eines Datensatzes als Graphen herangezogen werden. Aus dem Graphenmodell wiederum lassen sich andere Darstellungs- bzw. Discovery-Formen erzeugen. Die praktische Umsetzung des RD im der digitalen Sammlung des Kriminalmuseums im folgenden Kapitel soll genau dies zeigen.

4.7 Fazit

Die Wirklichkeit ist mehr als wir wahrnehmen und erkennen können. Bei komplexeren Formen des Informationsbedürfnisses bzw. wenn ein kognitiver Agent sein Informationsbedürfnis noch nicht kennt oder besser gesagt es noch nicht formulieren kann, stößt 'klassisches' Retrieval schnell an seine Grenzen. Synonyme, falsche logische Verbindungen von Termen, zu große

¹⁹⁹Vgl. DENGEL: Semantische Technologien: Grundlagen - Konzepte - Anwendungen, S.230-251.

Datenbestände oder multimediale Inhalte gehen mit erweiterten Erfordernissen des Retrievals einher. Gerade im Web sind explorative Elemente von Suchstrategien eine Herausforderung. In diesem Kapitel wurden vier Zugänge ausgehend vom 'klassischen' Retrieval aus Kapitel 3 beschrieben, die exploratives Verhalten fördern sollen. User-Interaktivität ermöglicht es während der Formulierung einer Query zu helfen, in dem mögliche Querys angeboten werden. Weiters kann die direkte und indirekte Partizipation von einer Vielzahl von Nutzerinnen gemessen werden und als Hilfestellung für einzelne Nutzerinnen verwendet werden, wie Bewertungen oder Recommender Systeme. Die Organisation von Suchergebnissen kann einerseits einen Informationskorporus selbst umfassen, wenn Beispielsweise Inhalte mittels Facettenklassifikation strukturiert sind und so eine facettierte Suche umgesetzt werden kann, oder eine nutzer-seitige Möglichkeit darstellen, Suchtreffer zu verwalten, speichern und zu exportieren. Informationsvisualisierung kann alle geeigneten Darstellungen von Strukturen, Ordnungen und Umfang eines Informationskorporus umfassen, die für Nutzerinnen relevant sein könnten. Die Suche nach Bedeutung und nicht nach Text kann mit einer semantischen Suche umgesetzt werden. Im folgenden Kapitel werden die semantischen Technologien beschrieben und im Detail auf die Umsetzung von RD-Funktionalitäten auf Basis der Prinzipien semantischer Suche eingegangen.

5 Prinzipien semantischer Suche am Beispiel der virtuellen Repräsentation des Hans Gross Kriminalmuseums

5.1 Einleitung

Das abschließende Kapitel befasst sich mit der praktischen Umsetzung des Studienprojektes und dokumentiert meine Aufgabenbereiche im Projekt. Diese Aufgabenbereiche umfassten das Projektmanagement und die Implementierung der Webpräsentation, sowie der Information Retrieval bzw. Resource Discovery Funktionalitäten der digitalen Sammlung des Hans Gross Kriminalmuseums.

Gegenstand des Kapitels ist zu Beginn ein Einschub, der das Studienprojekt, das Museum und die projektspezifische Umsetzung definiert, um die Ausgangslage des Projektes zu skizzieren. Daran anschließend wird die Infrastruktur - Das Geisteswissenschaftliche Asset Management System GAMS - beschrieben, auf deren Basis das Projekt umgesetzt wurde. Für das bessere Verständnis werden zuvor noch Grundbegriffe, wie *Langzeitarchivierung*, *Metadatenstandards* und *Semantische-*, wie auch *X-Technologien*, die zentral für die Infrastruktur und für die Arbeit im Ganzen sind, definiert. Ein besonderer Fokus wird auf die Beschreibung der Implementierung der RD-Funktionalitäten der Sammlung gelegt, die auf Prinzipien semantischer Suche beruhen.

5.2 Projektbeschreibung des Studienprojektes

Der Masterstudiengang EuroMACHS der Karl-Franzens Universität Graz führte im Zuge des viersemestrigen Studiums im Wintersemesters 2015/2016 und Sommersemesters 2016 ein Studienprojekt durch, in dem der Bestand des Hans Gross Kriminalmuseums digitalisiert und digital repräsentiert wurde. In meiner Funktion als Projektmanager wurde ein Projektplan erstellt, dessen Inhalte im Zuge dieser Arbeit dokumentiert werden. Diese Dokumentation umfasst die Definition des Projektes, die Ziele, die Rollen und ihre Zuordnung zu Work-Packages, die Meilenstones, sowie einer knappen Beschreibung des Hans Gross Kriminalmuseums selbst, um das Projekt im größeren Kontext fassbar machen zu können.

5.2.1 Projektdefinition und Projektziele

In diesem Studienprojekt galt es eine digitale Sammlung und in einem weiteren Schritt eine virtuelle Ausstellung zu planen und umzusetzen. Unter **digitaler Sammlung** ist eine mit Metadaten angereicherte Darstellung des Bestandes im Sinne der Langzeitarchivierung digitaler Objekte und deren Metadaten zu verstehen. Die digitale Sammlung beinhaltet zusätzliche Informationsressourcen und Funktionalitäten, die es ermöglichen, die vorhandenen Ordnungsstrukturen besser erschließen zu können, sprich im Bestand gezielt suchen, aber auch darin 'stöbern' zu können. Die virtuelle Ausstellung hingegen selektiert einen Teil des Bestandes - den Bestand

der Wilderer Objekte - und versucht das vielschichtige Phänomen für den User aufzubereiten und erfahrbar zu machen. Die virtuelle Ausstellung kann als Sprungmarke bzw. als eine Form des geleiteten RD des Bestandes gesehen werden. Sie eröffnet dem Nutzer die Möglichkeit etwas über den Bestand der digitalen Sammlung zu erfahren, ohne adäquate Hintergrundinformationen über die Sammlung des Museums zu besitzen. Hier liegt neben der Darstellung der Objekte auch das kontextualisierte Erfahren und Erleben sogenannter *Musealien*²⁰⁰ im Vordergrund. Dazu wird eine 'Geschichte' erzählt, mit der ein Angebot geschaffen wird, sich den Objekten und ihrem Kontext anzunähern. Diese Annäherung ist ein möglicher Ausdruck der Subjekt-Objekt-Beziehung und trägt zum musealen Wert bei.

Aus dieser Projektdefinition lassen sich folgende zwei **Projektziele** ableiten, die es im Zuge des Projektes zu erreichen galt:

- Die **digitale Sammlung** erlaubt es, die Objekte des Kriminalmuseums in Form eines Webportals öffentlich, zugänglich zu machen. Dabei sind die Objekte und ihre Metadaten erschlossen und auf standardisierte Form als digitale Objekte im Sinne der Langzeitarchivierung digitaler Ressourcen in einem Repository abgelegt. Die Erschließung beinhaltet ihre bildliche Repräsentation und eine digitale Repräsentation der Metadaten in dementsprechenden Standards, sowie die daraus erzeugten Repräsentationen.
- Ein zweites Ziel stellte die Erarbeitung und Umsetzung einer **virtuellen Ausstellung** zur Präsentation der Wilderer-Objekte dar, um ein ort-, und zeitungebundenes 'Erfahren' der ausgewählten Objekte ermöglichen zu können. Dies wird mit StoryMap.js²⁰¹ verwirklicht, einem Storytelling-Tool das einen starken Fokus auf geographische Karten bzw. hochauflösende Bilder legt, die etappenweise durchgeklickt werden können. Eine virtuelle Ausstellung ist kein Ersatz oder gar Gegenspieler eines 'echten' Museums, sondern soll viel mehr als eigenständige Ergänzung und Modifikation des Realen verstanden werden.

5.2.2 Das Hans Gross Kriminalmuseum und vorhandenes Quellenmaterial

Ein Abriss der Geschichte des Hans Gross Kriminalmuseums²⁰² und eine Beschreibung des Bestandes sollen zu einem besseren Verständnis der Umsetzungen des Projektes und seiner Ziele führen.

1896 gründet der Grazer Jurist Hans Gross das Kriminalmuseum. Primär umfasste das Museum am Anfang eine Lehrmittelsammlung zur Ausbildung von Studenten, Untersuchungsrichtern und Kriminalbeamten. Es handelte sich um eine von ihm zusammengetragene und dokumentierte Sammlung von sachlichen Beweismitteln, die dem Gross'schen Ideal einer '*systematisch*

²⁰⁰Objekte als Träger der Beziehung zwischen Mensch und Realität, die einen erhaltenden, vermittelnden Charakter aufweist: Vgl. WAIDACHER, Friedrich: *Museologie-knapp* gefasst. Wien and Köln and Weimar, 2005, S.320.

²⁰¹StoryMap, <https://storymap.knightlab.com>, 05.02.2016.

²⁰²Webpräsenz des Hans Gross Kriminalmuseums, <https://kriminalmuseum.uni-graz.at>, 02.07.2016.

umfassenden, enzyklopädischen Kriminologie’ entsprachen. Sein Bestreben war es die Kriminologie als moderne und eigenständige Wissenschaft, beruhend auf real-wissenschaftlichen Methoden und Richtlinien, gepaart mit einem stark pragmatischen Zugang zur Arbeit eines Juristen, zu entwickeln und zu etablieren. Sein "Handbuch für Untersuchungsrichter" galt bis in die 1970iger Jahre als Standardwerk der Kriminologie und beschreibt die zielführende Arbeit von Untersuchungsrichtern und deren methodisches Vorgehen. 1912 kam es zur Eröffnung des 'K.k. Kriminalistischen Instituts an der Universität Graz'. Als erstes dieser Art führte es zur Etablierung der Kriminologie als wissenschaftliche Disziplin. Basis für das heutige Kriminalmuseum ist dessen 'Kriminalistische Sammlung', die aus ungefähr 2500 Objekten besteht und über einen Zeitraum von ca. 50 Jahren zusammengetragen wurde. Das Kriminalmuseum wurde wiederum von Hans Gross in 32 Objektkategorien gegliedert, die den Bestand der Sammlung weiter definieren.²⁰³

Als Forschungsschwerpunkt der 'Grazer Schule der Kriminologie' des Instituts lassen sich wiederum folgende Punkte extrahieren:

- Kriminalbiologie
- Kriminelle Typologie
- Resozialisierung von Straftätern
- Verfeinerung kriminalistisch-technischer Verfahren
- Rückkoppelung der kriminalistischen Forschung an kriminalistische Praxis

Wenn auch die Grazer Schule international anerkannt und wissenschaftlich erfolgreich agierte, so war sie auch geprägt von biologistisch-sozialdarwinistischen, eugenischen und empirisch fragwürdigen Forschungsansätzen. In der Kriminalbiologie erforschte beispielsweise Adolf Lenz, der Gross als Leiter nachfolgte, ob und wie sich Persönlichkeitsmerkmale von Tätern typologisieren lassen: In der NS-Zeit dehnten sich diese Überlegungen auf rassenbiologische Aspekte aus.²⁰⁴

Täter, Tat und Opfer werden nicht nur rekonstruiert, sondern können auch als ein Spiegelbild der damaligen Lebensumstände im Sinne des 'Zeitgeists' betrachtet werden. Dies gilt nicht nur für die Quellen und Musealien, sondern auch für eine wirtschaftsgeschichtliche Perspektive auf Personen, wie z.B. Hans Gross selbst. Diese Aspekte müssen aus ihrer Zeit heraus betrachtet werden. Die Sammlung, die in gewissen Gesichtspunkten als unwissenschaftlich, brutal und abstoßend aufgefasst werden kann, ist ein Überbleibsel der Geschichte selbst. Somit verfügt die Sammlung an sich über ein Maß an Musealität. WAIDACHER führt Musealien als Bedeutungsträger und somit als Zeugnisse vom Menschen selbst an, die sowohl die Freuden, wie auch die Leiden, Hoffnungen und Ängste auf ihre Art widerspiegeln.²⁰⁵ Genau diesen Umgang mit

²⁰³Vgl. BACHHIESL: Räuber, Mörder, Sittenstrolche. 37 Fälle aus dem Kriminalmuseum der Karl-Franzens-Universität Graz, S.7-16.

²⁰⁴Vgl. KOCHER: Das K.k kriminalistische Universitätsinstitut in Graz, S.21-34.

²⁰⁵Vgl. WAIDACHER: Museologie-knapp gefasst, S.257-264.

diesen Objekten und den damit verbundenen Geschichten versucht das Kriminalmuseum seinen Besuchern zu vermitteln und 'erlebbar' zu machen. Im Falle des Kriminalmuseums sind es nun mal die 'Schattenseiten' des Lebens.

Zusammenfassend ist der Bestand aus heutiger Sicht eine dokumentierte Quellensammlung, aus der sich unterschiedlichste Fragestellungen in den Bereichen Wissenschaftsgeschichte des späten 19. und frühen 20. Jahrhunderts, der Technikgeschichte, Sozialkunde und Soziologie, gesellschaftliche Aspekte dieser Zeit, sowie vielerlei multiperspektivische Fragestellungen der Kriminologie formulieren lassen. Durch die Digitalisierung des Kriminalmuseums und seine Repräsentation im Web ist der Bestand zeit-, und ortsungebunden für interessierte Forscherinnen und Besucherinnen zugänglich. Die Standardisierungen eröffnen die Anwendung digitaler Methoden, die den klassischen Methoden-Kanon der Geisteswissenschaften adäquat erweitern. Der Bestand des Museums kann in unterschiedliche historisch gewachsene Sammlungsbereiche gegliedert werden. So besteht der Gesamtbestand aus Artefakten, Karteikarten, einer pornographischen Sammlung, Schautafeln, Giften und einem Bestand an juristischen Akten, die mit anderen Objekten in Verbindung stehen.

5.2.3 Rollen, Work-Packages und Milestones

Für die Projektplanung wurden Arbeitsphasen bzw. -pakete definiert, die im Folgenden als Work-Packages bezeichnet werden. Jedes Work-Package besteht aus mehreren Einzelaufgaben, sogenannten Tasks. Am Ende jedes Work-Package stehen ein oder zwei Milestones. Jeder Milestone definiert ein Ereignis, in dem die Arbeit an Tasks zur Erreichen des Zieles evaluiert und diskutiert wurde, um feststellen zu können, welche Entscheidungen für den weiteren Projektverlauf getätigt werden müssen. Jede Person im Projektteam nimmt zwei Rollen im Projekt ein, die jeweils einem Work-Package zugeschrieben werden.²⁰⁶ So erfolgte beispielsweise die Zuordnung einer Person zum Work-Package *Projektmanagement* und zum Work-Package *Implementierung*. Die Zuordnung basierte auf den Interessen und Fertigkeiten der Personen und stand in Zusammenhang mit dem Themengebiet der geplanten Abschlussarbeit. Zur Realisierung des Projektes wurden folgende Work-Packages festgemacht:

- **Projektmanagement**

Das Projektmanagement ist für alle administrativen und organisatorischen Abläufe, die das Projekt im Vorhinein planen, steuern und zum Abschluss führen, zuständig. Dies umfasste die Formulierung eines Projektplans und Pflichtenheftes, einer zeitlichen Einteilung, sowie das Aufbauen und Warten von Kommunikationskanälen im Projektteam, sowie zu beteiligten Personen außerhalb, den sogenannten Stakeholdern. Ein Pflichtenheft umfasst eine schriftliche, allgemeine, aber dennoch konkrete Zusammenführung aller Umsetzungen, Funktionalitäten und Anforderungen in einem (Software)Projekt: auf die-

²⁰⁶Vgl. KUSTER, Jürg et al.: Handbuch Projektmanagement. Berlin and Heidelberg, 2011, S.8-24.

ser Basis wurden die Ziele ausführlich definiert.²⁰⁷

Zur Kommunikation wurden ein Trello-Board²⁰⁸ angelegt, das es ermöglicht eine kollaborative Plattform über das Web zu nutzen, in der die Work-Packages, ihre Tasks und Milestones einer Person und einem Zeitraum zugeordnet wurden. Der Kommunikationsprozess wurde, neben wöchentlichen, realen Treffen, in einer geschlossenen Facebook Gruppe durchgeführt.

Milestone: Erstellung eines Projektplans und eines Pflichtenhefts, das die Funktionalitäten der digitalen Sammlung und der virtuellen Ausstellung festlegt; Evaluierung des Projekts.

- **Dokumentation und Analyse**

Dokumentation und Analyse umfassen den Aufgabenbereich zur inhaltlichen Erschließung des Kriminalmuseums und dessen Bestandes. Die Ergebnisse dieses Work-Package dienen als Ausgangspunkt für die folgenden Aufgabenbereiche: Die Sichtung, Erschließung, Dokumentation und Ordnung des Bestandes. Beispielsweise wurden die Karteikarten zu Kriminalfällen der Wildereidelikte ausgewählt, die vorhandene Inventarliste der Objekte normalisiert und ein Material- und Objektthesaurus definiert, um Objekte nach diesem Thesaurus katalogisieren zu können. Weiter wurden in diesem Zusammenhang Recherchearbeiten hinsichtlich der wissenschaftlichen Ausarbeitung der virtuellen Ausstellung, in Form eines Ausstellungsexpooes getätigt.

Milestone: Dokument mit allen Rechercheergebnissen in Form eines Ausstellungsexpooes; Ein ausgearbeitetes Konzept für die Modellierung von Ordnungsstrukturen (Thesauri) und Normalisierung der vorhandenen Daten.

- **Präsentation und Design**

Dieses Work-Package umfasst zwei zentrale Ziele. Zum einen wird das Design des Webportals der digitalen Sammlung festgelegt. Dies umfasst statische Designentwürfe, sowie Fragen der Accessibility²⁰⁹, Responsiveness²¹⁰ und der Usability²¹¹. Zum anderen wird ein Konzept für die Umsetzung der virtuellen Ausstellung, ausgehend von den vorhandenen Daten, bestehend aus Ausstellungsexpooes und dem Daten- und Quellenmaterial des Museums, definiert.

Milestone: Ausarbeitung eines Konzepts bzw. Storyboards für die virtuelle Ausstellung; Ausarbeitung eines Designentwurfs und Konkretisierung der Funktionalitäten der digitale Sammlung.

²⁰⁷Vgl. KUSTER et al.: Handbuch Projektmanagement, S.112-113.

²⁰⁸Web-basierte Projektmanagementsoftware, <https://trello.com/>, 09.12.2016.

²⁰⁹Barrierefreie Implementierung von Informations- und Kommunikationsangeboten, <https://en.wikipedia.org/wiki/Accessibility>, 12.02.2017.

²¹⁰Paradigma der Entwicklung von Applikationen für unterschiedliche Devices, https://de.wikipedia.org/wiki/Responsive_Webdesign, 12.02.2017.

²¹¹Gebrauchstauglichkeit einer Webressource, <https://en.wikipedia.org/wiki/Usability>, 12.02.2017.

- **Informationsmodellierung und Metadatenmanagement**

In diesem Work-Package werden die erschlossenen Daten in Dateiformate, Datenmodelle und Standards überführt, um die Daten in die digitale Sammlung einspielen zu können. Folgende Überführungen werden in diesem Zusammenhang durchgeführt:²¹²

- Die Karteikarten zu Wildereidelikten werden transkribiert und in TEI/XML ausgezeichnet. Beim Ingest in die digitale Sammlung wird für jede Karteikarte ein eigenständiges digitales Objekt angelegt.
- Die normalisierte Excel-Liste wird mittels XSL-Transformation zu einzelnen LIDO-Objekten verarbeitet. Jedes LIDO-Objekt erzeugt ein eigenes digitales Objekt in der GAMS-Infrastruktur. Dieser Arbeitsschritt wurde vom ZIM durchgeführt.
- Der Material- und der Objektthesaurus wird in SKOS modelliert.

Milestone: 'Digitalisierung' von Objekten und Überführung in Ordnungsstrukturen und Metadatenstandards für die Digitale Sammlung; Umsetzung dementsprechender Workflows.

- **Implementierung**

Das letzte hier angeführte Work-Package beschreibt die Umsetzung aller Ergebnisse der vorhergehenden Work-Packages. Somit umfasst es die Implementierung der digitalen Sammlung und der virtuellen Ausstellung in der GAMS-Infrastruktur, auf Basis der Ergebnisse der vorangegangenen Arbeitspakete. Die Implementierung der digitalen Sammlung erfolgt mittels XSL-T. Aus den standardisierten, digitalen Objekten werden HTML-Repräsentation erzeugt, die zusammen betrachtet, das Webportal der digitalen Sammlung erzeugen. Die virtuelle Ausstellung wird mit dem JavaScript-Template²¹³ StoryBoards.js umgesetzt, das mit Hilfe eines XML-Templates erstellt wurde.

Milestone: Implementierung eines Demonstrators der digitalen Sammlung und der virtuellen Ausstellung.

Abbildung 14 visualisiert den Zeitplan und den Projektverlauf des Studienprojektes. Darin enthalten sind die, für jede Phase definierten, Milestones, an deren Abschluss eine kurze Review-Phasen eingeschoben wird, um gemeinsam in der Gruppe, die vorhandenen Ergebnisse zu reflektieren, analysieren und das weitere vorgehen gemeinsam zu entscheiden. Die Bereiche Präsentation und Design und Informationsmodellierung und Metadatenmanagement werden parallel durchgeführt, da diese Work-Packages nicht direkt voneinander abhängig sind. Jedes Work-Package besteht aus einer Anzahl von Einzelaufgaben, die mit *Trello* dynamisch von der Projektgruppe und dem Projektmanager verwaltet werden, wie Abbildung 15²¹⁴

²¹²Die genannten Technologien werden im Kapitel 5.3 beschrieben.

²¹³JavaScript, Programmiersprache, <https://de.wikipedia.org/wiki/JavaScript>, 04.03.2017.

²¹⁴Trelloboard des Studienprojektes, <https://trello.com/b/bgW70TtY/kriminalmuseum>, 12.02.2017.

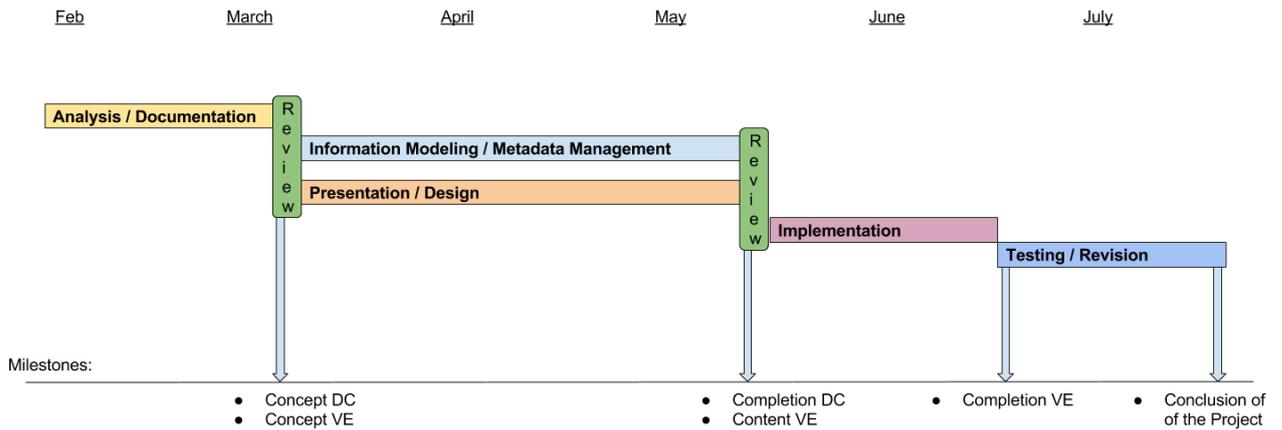


Abbildung 14: Zeitplan und Projektablauf

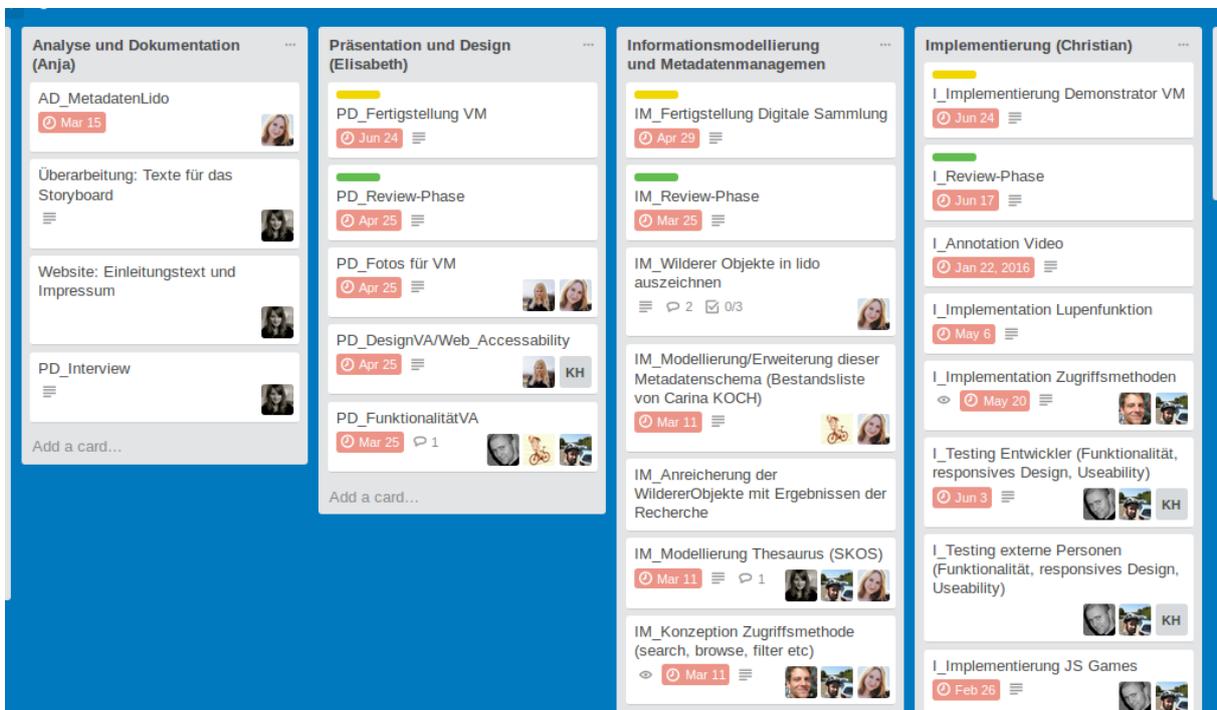


Abbildung 15: Ausschnitt des Trello-Boards zur Darstellung einzelner, Work-Packes zugeordneten, Tasks

5.3 Technische Grundlagen zur Umsetzung des Studienprojektes

Das Geisteswissenschaftliche Asset Management System, kurz GAMS, wird definiert als 'OAIIS-konformes Asset Management System zur Verwaltung, Publikation und Langzeitarchivierung digitaler Ressourcen aus allen geisteswissenschaftlichen Fächern'²¹⁵. Die Umsetzung der digitalen Sammlung und der virtuellen Ausstellung des Hans Gross Kriminalmuseums ist in das GAMS-Repositorium inkludiert. Eine ausführliche Beschreibung dieses Repositorium und der ihr zu Grunde liegenden Modelle und Begrifflichkeiten werden im Folgenden weiter ausgeführt.

5.3.1 Digitale Langzeitarchivierung und digitales Objekt

Vor der Digitalisierung der Gesellschaft waren Gedächtnisorganisationen wie Archive, Museen und Bibliotheken zuständig für ein verlässliches und vertrauensvolles Bewahren und Zugänglichmachen von Dokumenten und Objekten. Das digitale Zeitalter konfrontiert wirtschaftliche Unternehmen, politische Organisationen, sowie Gedächtnisorganisationen mit neuen Herausforderungen, die großen Datenmengen zu bewahren und zu verwalten. KEITEL und SCHÖGER nennen folgende Herausforderungen:²¹⁶

- Der physische Erhalt digitaler Daten auf kurzlebigen elektronischen Datenträgern.
- Die Abhängigkeit digitaler Information von komplexen technischen Systemen, die einem rasanten technologischen Wandel unterworfen sind.
- Das Problem der Authentizität und Vertrauenswürdigkeit von digitaler Information, die Möglichkeiten der Manipulation bieten.
- Die Möglichkeit der Vervielfältigung und Verteilung von rechtlich geschützten Informationsressourcen über das World Wide Web.

Das Kompetenznetzwerk zur digitalen Langzeitarchivierung im deutschsprachigen Raum *NESTOR* ist ein Kooperationsverbund von Museen, Archiven, Bibliotheken und Experten, der Lösungen für das Problem der Langzeitarchivierung und Langzeitverfügbarkeit digitaler Quellen behandelt. Dieses Problem beschreiben LIEGMANN und NEUROTH als einen unbestimmten Zeitraum, der von unbekanntem Faktoren aus dem Bereich der technologischen, wie auch der soziokulturellen Entwicklung beeinflusst ist. Es ist nur teilweise abschätzbar, welche digitalen Ressourcen, Speichermedien oder institutionellen Bedingungen in der Zukunft existieren und wie weit es aus heutiger Sicht möglich war, diese Ressourcen ohne Informationsverlust zu erhalten. Dabei bedeutet *Langzeit* das Entwickeln und Umsetzen nachhaltiger, verantwortungsvoller, sowie vertrauensstiftender Strategien, die ein Instrumentarium liefern sollen, die Herausforderungen der sich wandelnden Informationsgesellschaft bewältigen zu können. Der Begriff *Archivierung* wiederum ist breiter zu verstehen als ein bloßes Speichern. Es beinhaltet

²¹⁵GAMS, gams.uni-graz.at, 05.02.2017

²¹⁶Vgl. KEITEL: Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, S.1-3.

die dauerhafte Verfügbarkeit, Nutzung und Interpretierbarkeit der erschlossenen digitalen Ressourcen.²¹⁷ Für KEITEL und SCHOGER ist ein digitales Langzeitarchiv eine

*"Organisation (bestehend aus Personen und technischen Systemen), die die Verantwortung für den Langzeiterhalt und die Langzeitverfügbarkeit von Information in digitaler Form sowie die Bereitstellung für eine bestimmte Zielgruppe übernommen haben."*²¹⁸

Um Information für bestimmte Zielgruppen verfügbar machen zu können, benötigt es IR- Funktionalitäten, um einer Nutzerin, die gesuchten Informationsressourcen zugänglich machen zu können. Dieses Anliegen und die Speicherung im Allgemeinen werden im Referenzmodell zur Umsetzung der Langzeitarchivierung, dem **Open Archival Information System** (OAIS), beschrieben. BRÜBACH definiert das OAIS als eine

*'Organisation, in der Menschen und Systeme mit der Aufgabenstellung zusammenwirken, digitale Informationen dauerhaft über einen langen Zeitraum zu erhalten und einer definierten Nutzerschaft verfügbar zu machen.'*²¹⁹

Darunter ist keine fertige Spezifikation zu verstehen, sondern ein theoretisches Grundgerüst, das auf unterschiedlichste Art und Weise in der Praxis umgesetzt werden kann und somit offen und skalierbar ist. Die Offenheit des Systems ist darin begründet, dass technologische Entwicklungen nach einer gewissen Zeit obsolet werden. Das bedeutet, dass Anwendungsprogramme, Datenmodelle, Speichermedien etc. nach einer bestimmten Zeit durch neue Technologien ersetzt werden müssen. Um diesen Anforderungen entsprechen zu können werden keine bestimmten Datenformate oder -typen definiert, sondern eine Architektur, in der Abläufe beschrieben werden, die mit einer entsprechenden Technologie umgesetzt wird. Das OAIS beschreibt den technischen, administrativen und institutionellen Rahmen, der für die Langzeitarchivierung notwendig ist und erstreckt sich dabei vom Ingest-Prozess, über die physikalische Speicherung und Retrieval Funktionalitäten, bis zur Ausgabe von Inhalten für unterschiedliche Nutzerinnengruppen.²²⁰

Die Einbettung der GAMS-Infrastruktur im universitären und wissenschaftlichen Rahmen des Zentrums für Informationsmodellierung an der Karl-Franzens Universität Graz, erfüllt laut Definition die im OAIS Referenzmodell gesetzten Paradigmen.²²¹

²¹⁷Vgl. LIEGMANN, Hans/NEUROTH, Heike: Einführung. In NEUROTH, Heike et al. (Hrsg.): nestor-Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Göttingen, 2010 (URL: http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_390.pdf).

²¹⁸KEITEL: Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, S.5.

²¹⁹Vgl. CCSDS: Reference Model for an Open Archival Information System (OAIS), Recommended Practice. CCSDS 650.0-M-2 (Magenta Book) Issue 2, 2012 – Technischer Bericht.

²²⁰Vgl. BRÜBACH: Das Referenzmodell OAIS.

²²¹Vgl. BIRKNER et al.: Guideline zur Langzeitarchivierung, S.41-54.

Digitales Objekt

GIARETTA definiert ein digitales Objekt als *'object of a set of bit sequences'*.²²² Dieser Definition schließt sich auch *NESTOR* an und hebt hervor, dass somit alles, was von einem Computer gespeichert und verarbeitet werden kann, ein digitales Objekt darstellt. Darunter ist keine einzelne Datei zu verstehen, sondern ein Paket miteinander in Verbindung stehender Dateien, die im Ganzen betrachtet ein digitales Objekt ergeben. Diese Mehrgliedrigkeit ergibt sich aus den drei unterschiedlichen Ebenen, die das Objekt beschreiben: das **physische Objekt**, als Manifestation eines Bitstroms auf einem Datenträger, das **logische Objekt** als Vorgabe, wie der Bitstrom verarbeitet werden muss und das **konzeptuelle Objekt** als Repräsentation des Inhaltes in unterschiedlichen Formen und mit unterschiedlichen Funktionalitäten.²²³ Dieser 'Paketgedanke' wird auch im OAIS verfolgt. Informationsressourcen, in Verbindung mit ihren Metadaten, werden als Übergabeinformationspaket (SIP) ins Archiv ingestiert, als Archivinformationspaket (AIP) im digitalen Archiv verwaltet und als Auslieferungsinformationspaket (DIP) an eine Nutzerin übermittelt. Das gleiche digitale Objekt kann unterschiedliche Formen mit unterschiedlichen Funktionalitäten und Metadaten annehmen.²²⁴ In Kapitel 5.3.4 wird dieses Prinzip am praktischen Beispiel vertieft.

Ein weiteres Grundprinzip in Verbindung mit digitalen Objekten ist ein **handle**. Dabei handelt es sich um einen eindeutigen globalen Identifikator, der es ermöglicht, digitale Objekte zu referenzieren und zu zitieren.²²⁵

5.3.2 Metadaten und Metadatenstandards

Oft werden Metadaten mit 'Daten über Daten' gleichgesetzt. Doch diese Definition ist zu unscharf. Generell dienen Metadaten der Beschreibung von Informationsressourcen, um einen gewissen Mehrwert zu erhalten. Ein Objekt in einem Museum wird mit einer Objektbezeichnung, einer Inventarnummer oder einer Beschreibung inventarisiert. Der Mehrwert dieser Metadaten liegt darin, dass mit diesen Daten ein Objekt verwaltet, wiedergefunden und auf andere bestimmte Art und Weise verwertet werden kann. In diesem Zusammenhang unterscheidet man folgende Typen von Metadaten.²²⁶

- **Deskriptive Metadaten** sind Daten, die eine Informationsressource beschreiben. Die Objektbezeichnung, ein Datum oder eine Person, die mit diesem Objekt in Verbindung steht (z.B. Hersteller oder Autor) sind Metadaten, die dazu verwendet werden ein Objekt zu identifiziert. Solche Metadaten haben eine große Bedeutung für IR-Prozesse, da sie in kontrollierten Vokabularien, Findbüchern oder Katalogen Verwendung finden.

²²²GIARETTA, David: Advanced Digital Preservation. Berlin and Heidelberg, 2011, S.37.

²²³Vgl. FUNK, Stefan E.: Digitale Objekte und Formate. In NEUROTH, Heike et al. (Hrsg.): nestor-Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Göttingen, 2010 (URL: http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_445.pdf).

²²⁴Vgl. KEITEL: Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, S.15-17.

²²⁵Vgl. BAEZA-YATES et al.: Modern information retrieval. The concept and technology behind search, S.714-715.

²²⁶Vgl. BIRKNER et al.: Guideline zur Langzeitarchivierung, S.51-52.

- **Administrative Metadaten** dienen der Verwaltung von Informationsressourcen. Sie klären beispielsweise rechtliche Fragen, die mit einem Objekt verbunden sind. Für ein Objekt in einer digitalen Sammlung ist es wichtig festzulegen mit welchen Rechten es veröffentlicht wurde.
- **Strukturelle Metadaten** beschreiben die Verbindung von Objekten zu anderen Informationsressourcen, Ordnungsstrukturen oder Metadaten. Beispielsweise kann die Zusammengehörigkeit alle digitalisierten Seiten eines Buches mit diesem Metadatatyp beschrieben werden.
- **Technische Metadaten** beinhalten Angaben zu Hashwerten (Prüfsumme), Datenformaten oder Datengröße.²²⁷

Weiter können Metadaten in intrinsische, Objekt beschreibende (Klassifikation, Schlagwort) und extrinsische Metadaten, ein Objekt einordnend (Erscheinungsjahr, -form), unterschieden werden.

Die Grundüberlegung Standards zu verwenden beruht auf dem Problem der **Interoperabilität**. Dieser Begriff beschreibt die Problemstellung von Informationskorpora, die nicht über gemeinsame Technologien, Standards und Methoden verfügen, um einen kompatiblen Austausch zwischen Beständen und Institutionen zu ermöglichen. Dazu ist in der Regel die Einhaltung gemeinsamer Standards unabdingbar.²²⁸

Unterschiedliche Standards verfolgen unterschiedliche Ziele. An dieser Stelle sollen, aus der großen Menge existierenden Standards, nur drei Beispiel angeführt werden, die essentiell Bestandteil der Umsetzung im Studienprojekt dargestellt haben.

Der **Dublin Core (DC)**²²⁹ ist ein Beispiele für einen häufig verwendeten, basalen Metadatenstandard im Sinne deskriptiver Metadaten, zur Beschreibung jedes denkbaren digitaler Objektes mittels eines Sets an Kernelementen.²³⁰ Der DC-Metadatensatz der Karteikarte *o:km.87*, der automatisch aus einem Mapping des TEI/XML-Ausgangsdatensatzes generiert wird, umfasst Informationen wie: Titel (*dc:title*), Sprache (*dc:language*), Rechte (*dc:rights*) und Beziehungen (*dc:relation*):²³¹ Folgender Code-Ausschnitt zeigt das in XML serialisierte Ergebnis des DC-Mappings an:

²²⁷Vgl. KEITEL: Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, S.107-108.

²²⁸Interoperabilität, <https://www.w3.org/blog/2008/05/open-standards-interoperability>, 12.04.2017.

²²⁹Dublin Core Metadata Element Set, <http://dublincore.org>, 05.02.2017.

²³⁰Vgl. KARGL, Carina: Der Geschichte ein Gesicht geben - Die Modellierung und Aufbereitung der 'Stamm- und Wappenbücher' der Stadt Regensburg. Graz, 2012, S.29-34.

²³¹DC Metadatensatz einer Karteikart, <http://gams.uni-graz.at/o:km.87/DC>, 12.02.2016.

```

1 <oai_dc:dc>
2   <dc:title>Karteikarte KM-KK.87</dc:title>
3   <dc:publisher>Universitätsmuseen, Karl-Franzens-Universität Graz</dc:publisher>
4   <dc:identifizier>o:km.87</dc:identifizier>
5   <dc:source>87</dc:source>
6   <dc:language>deu</dc:language>
7   <dc:relation>Hans Gross Kriminalmuseum, Virtuelle Sammlung</dc:relation>
8   <dc:relation>http://gams.uni-graz.at/km</dc:relation>
9   <dc:rights>Creative Commons BY-NC-SA 4.0</dc:rights>
10 </oai_dc:dc>

```

Die **Text Encoding Initiative (TEI)** bietet einen De-facto-Standard für Geisteswissenschaftlerinnen zur Repräsentation von Text, der in digitaler Form existiert, an.²³² Im Zuge des Studienprojektes wurden Karteikarten des Kriminalmuseums nach diesem Standard ausgezeichnet. Im Anhang, unter Punkt 7.1.1, findet sich ein Beispiel-Datensatz eines TEI/XML, der Karteikarte *o:km.87*, sowie im nächsten Kapitel ein TEI/XML-Ausschnitt. Die *Guidlines* der TEI umfassen einen modular verwendbaren Rahmen, in dem Elemente und Strukturen nach einer definierten Logik verwendet werden können, um Texte so zu kodieren, dass ein semantischer Mehrwert entsteht.²³³

Musealien wurden mit **Lightweight Information Describing Objects (LIDO)** ausgezeichnet. Dabei handelt es sich um einen Standard, der Objekte im musealischen Sinne beschreibt und diese Informationen austauschbar macht.²³⁴ Im Zuge des Studienprojektes wurde dieser Bereich von Mitarbeiterinnen des ZIM umgesetzt. Unter der URL http://gams.uni-graz.at/o:km.8408/LIDO_SOURCE findet sich ein beispielhafter LIDO-Datensatz zur einem Objekt im Sammlungsbereich der Wilderer.

5.3.3 X-Technologien

Darunter versteht man alle formalen Sprachen zur Verarbeitung von Daten, die mittels **Extensible Markup Language (XML)**²³⁵ modelliert wurden. XML ist ein vereinfachtes Derivat der Metasprache SGML mit dem Zweck, Daten hierarchisch zu strukturieren und sie plattform- und implementationsunabhängig austauschbar zu machen. In XML kodierte Daten unterscheiden sich von Daten in relationalen oder objektorientierten Datenbanken, da Daten und Datenschema direkt zusammen kodiert werden. Die Daten beschreiben ihre eigene logische und physikalische Struktur.²³⁶ Im Zuge des Studienprojektes definiert XML eine Syntax, um verwendete Metadatenstandards (DC, TEI, LIDO) zu serialisieren. Auch die semantische Modellierung in RDF erfolgt in einer XML-Serialisierung. Folgender Ausschnitt zeigt einen Ausschnitt des TEI/XML

²³²TEI, <http://www.tei-c.org/index.xml>, 05.02.2017.

²³³Vgl. KARGL: Der Geschichte ein Gesicht geben - Die Modellierung und Aufbereitung der 'Stamm- und Wapenbücher' der Stadt Regensburg, S.36-38.

²³⁴LIDO, www.lido-schema.org, 05.02.2017.

²³⁵Extensible Markup Language, <https://www.w3.org/XML/>, 27.12.2016.

²³⁶Vgl. LINCKELS/MEINEL: E-librarian service: user-friendly semantic search in digital libraries, S.24-30.

der Karteikarte o:km.87 (siehe Anhang,7.1.1), um den Aufbau eines XML-Dokuments beispielhaft zu beschreiben:

```
1 <div type="suspectDiv">
2   <p><hi>Name des Täters:</hi>
3     <persName role="suspect" xml:id="S0007">
4       <forename> Franz</forename><surname> H.</surname>
5     </persName>
6   </p>
7   <p><hi>Alter:</hi>
8     geb.<date corresp="#S0007" type="birthdate" when="1900-12-01">
9       1/12.1900</date>
10  </p>
11  <p><hi>Beruf:</hi>
12    <term ana="Besitzerssohn" corresp="#S0007" type="profession">
13      Besitzerssohn</term>
14  </p>
15  <p><hi>Vorstrafen:</hi>
16    <term corresp="#S0007" ref="#unbescholten" type="law">
17      unbescholten</term>
18  </p>
19 </div>
```

Im `<div>` Element mit dem Attribut `type='suspectDiv'` wird der Teil der Karteikarte ausgezeichnet, der den Täter eines Deliktes angibt. Dieses `<div>` verfügt über vier `<p>`. Jeder dieser Paragraphen führt weitere Informationen zum Täter an. Der Name des Täters wurde in einem `<persName>` erfasst und mit dem Attribut `role` als Täter gekennzeichnet und mit `xml:id` eindeutig identifiziert. Die nun folgenden `corresp` Attribute verbinden die Information über Alter, Beruf und Vorstrafen mit der Person. Auf diese Weise sind Informationen zu einer Person in einer Karteikarte erschlossen, standardisiert und weiter verwertbar.

Die funktionale Programmiersprache **Extensible Stylesheet Language Transformations (XSLT)** ermöglichen es XML Dokumente in andere XML-Dokumente zu transformieren.²³⁷ Beispielsweise wird das zuvor angeführte TEI/XML-Beispiel mit einer XSL-Transformation in ein HTML-Dokument transformiert, das mittels Browser angezeigt werden kann.

Ein Hilfsmittel, um in der internen, hierarchische Struktur eines XML-Dokuments bestimmte Bereiche, wie beispielsweise Elemente oder Attribute, anzusprechen, ist die XML-Querysprache **XML Path Language XPath**.²³⁸

²³⁷XSLT, <https://www.w3.org/TR/xslt>, 05.02.2016.

²³⁸Vgl. BAEZA-YATES et al.: Modern information retrieval. The concept and technology behind search, S.211-216.

5.3.4 Funktionsweise der GAMS-Infrastruktur und Umsetzung des Projekts

Ausgehend von einem beispielhaften Datenstrom soll in diesem Kapitel die Funktionsweise dieser Infrastruktur am konkreten Beispiel der digitalen Sammlung beschrieben werden. Die **Flexible Extensible Digital Object and Repository Architecture (FEDORA)**²³⁹ bildet das Herzstück der GAMS-Infrastruktur. FEDORA definiert sich als plattformunabhängige, modulare und open-source Systeminfrastruktur, mit deren Hilfe digitale Objekte und ihre Metadaten mit einer eindeutigen Kennung - eines persistenten Identifikators (PID) - in einem Repository verwaltet werden können. Dies umfasst den Zugriff auf Datenströme, deren Repräsentationsmöglichkeiten, Dissemination und Methoden.²⁴⁰ Somit sind die Grundvoraussetzungen für eine OAIS-konforme Architektur mit FEDORA gegeben.

Die URL `http://gams.uni-graz.at/o:km.87` verweist auf ein digitales Objekt in FEDORA und verfügt über den PID `o:km.87`. Dabei handelt es sich um eine Karteikarte des Kriminalmuseums, ein Blatt Papier, das die Aufnahme einer Musealie in den Bestand dokumentiert und als eigenständiges museales Objekt verstanden werden kann. Dieses digitale Objekt besteht aus mehreren **Datenströmen**, von denen hier eine Auswahl angeführt wird. Zum einen sind darin zwei Bilder abgelegt, die Vorder- und Rückseite der Karteikarte, ein XML-File, das mittels TEI-Standard ausgezeichnet wurde, ein Metadatensatz im DC-Format und ein RDF/XML. Weiters umfasst das digitale Objekt mehrere ihm zugeordnete Methoden. Diese Methoden verarbeiten Datenströme und erzeugen dynamisch neue Inhalte. Sie verbinden digitale Objekte mit anderen oder beschreiben das Verhalten von Objekten zu anderen. Alle diese Datenströme und Methoden gemeinsam ergeben das digitale Objekt `o:km.87`, das in Abbildung 16 ausschnittsweise und modellhaft skizziert ist. Dabei sind alle Ebenen eines digitalen Objekts, wie in Kapitel 5.3.1 definiert, enthalten.

²³⁹FEDORA, <http://fedorarepository.org/about>, 05.02.2016.

²⁴⁰Vgl. STEINER, Elisabeth: Kulturelles Erbe virtuell repräsentiert: Überlegungen zur Konzeption eines Digitalen Archivs der Stamm- und Wappenbücher der Stadt Regensburg. Graz, 2012, S.43-44.

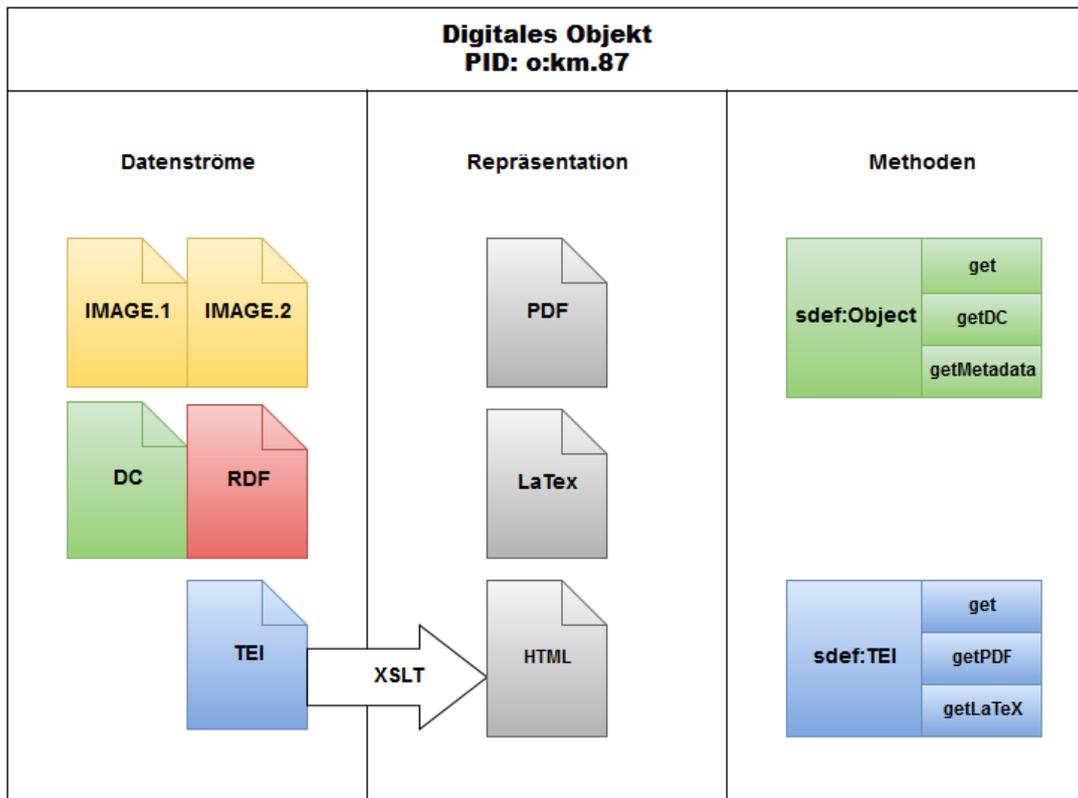


Abbildung 16: Digitale Objekt einer Karteikarte, bestehend aus Datenströmen, ihrer Repräsentation und Methoden.

Die Repräsentationsformen werden *on-the-fly* aus den Daten, also aus dem TEI, mittels XSL-Transformation erzeugt und beispielsweise in HTML-Code umgewandelt. Dieses HTML kann mit der Methode `/methods/sdef:TEI/get` als URL

<http://gams.uni-graz.at/archive/objects/o:km.87/methods/sdef:TEI/get?>

im Browser aufgerufen werden. Dieser Vorgang wird *Single Source Publishing* genannt und unterstreicht den Aufbau von FEDORA auf Basis von X-Technologien. Aus einer einzelnen Quellen können unterschiedliche Darstellungen erzeugt werden.²⁴¹ Die **Methoden** umfassen Getter-Methoden, um Repräsentationsformen, Metadaten oder andere Funktionalitäten aufzurufen. Dabei steht *sdef:* für *service definition objects* und beschreibt Kontrollobjekte, die über Serviceoperationen verfügen.²⁴² Diese können auf das digitale Objekt mit *sdef:Object* selbst angewandt werden, um beispielsweise mit *getDC* die Metadaten aufzurufen, oder auf das TEI, um mit *get* die HTML oder eben mit *getPDF* eine PDF-Version des TEI aufrufen zu können. Die Methoden erzeugen aber nicht die Repräsentationen. Zu diesem Zweck wird eine XSL-Transformation aus einem anderen, eigenständigen Datenstrom referenziert. Zusammengefasst wird die Gene-

²⁴¹Vgl. STIGLER, Johannes Hubert: Neue Wege in der Digitalen Edition. Jenseits von Hypertext und Nicht-Linearität. In Beihefte zu editio 30. Graz, 2009, S.203-212.

²⁴²Vgl. STEINER: Kulturelles Erbe virtuell repräsentiert: Überlegungen zur Konzeption eines Digitalen Archivs der Stamm- und Wappenbücher der Stadt Regensburg, S.45-47.

rierung und Verwaltung einer bestimmten Primärquelle in einem sogenannten *Content-Model*. Ein Content-Model beschreibt alle vorhandenen Methoden für bestimmte Datenströme, wie zum Beispiel für alle TEI- oder alle LIDO-Datenströme. Die XSL-Transformationen müssen nicht händisch jedem Objekt zugeordnet werden, sondern können innerhalb der Content-Model für alle Objekte des gleichen Typs definiert werden. Ein weiteres Content-Model *cirilo:context* dient der Strukturierung und Ordnung von Objekten. Beispielsweise wurden alle Karteikarten dem übergeordneten Konzept 'context:karteikarten' zugeordnet.²⁴³ Damit ist man in der Lage die Wilderer-Objekte im Museum mit der Zuordnung 'context:wilderer' von anderen Objekten im Museum oder den Karteikarten zu separieren. Unter folgender URL sind die unterschiedlichen Sammlungsbereiche, wie etwas Chemikalien, Objekte oder Wilderer, aufrufbar:

```
http://gams.uni-graz.at/archive/objects/context:km/methods/sdef:Context/get?
mode=collection
```

Die Auswahl nach Sammlungsbereichen schafft eine erste und einfache Form des Retrievals und Discoveries. An Hand dieser URL sieht man die Anwendung von Methoden auf das eigenständige digitale Objekt *context:km*, dessen Context-Zuordnung gleichzeitig auch das ganze Projekt beschreibt. Der Parameter *mode* mit dem Wert *collection* liefert den Parameter, der dazu dient unterschiedliche Webseiten ansteuern zu können.

Erzeugung der HTML-Repräsentationen

Es existieren mehrere XSL-Stylesheets, die dynamisch aus den unterschiedlichen Datenströmen eine gemeinsame HTML-Repräsentationen generieren:

- Das 'Gerüst' jeder Seite, bestehend aus Header, Footer und Navigationsleiste, erfüllt auch den Zweck globale Variablen zu definieren, sowie CSS²⁴⁴, und JavaScript Files einzubinden. Dieses Stylesheet wird in alle anderen inkludiert.
- Die Auswahl der Bestandsbereiche des Museums. Für jeden Bestandsbereich wurde ein eigenes Context-Objekt angelegt, wie *context:km.wilderer* mit der darin enthaltenen Teilmenge *context:km.karteikarten*. Für jede dieser Bereiche kann eine eigenständige Transformation wirken.
- Die Detailansicht der LIDO-Objekte. Ausgehend von Objekten, deren Datenausgangslage ein LIDO-File ist, erzeugt diese Transformation eine HTML-Darstellung der Metadaten, sowie der vorhanden Scans in einem Viewer. Dieses Stylesheet wirkt im *cirilo:km.LIDO* generisch auf alle LIDO-Objekte.
- Die Detailansicht der TEI-Objekte zeigt den transkribierte und semantisch angereicherte

²⁴³GAMS Dokumentation, <http://gams.uni-graz.at/docs>, 05.02.2016.

²⁴⁴Cascading Style Sheets, https://de.wikipedia.org/wiki/Cascading_Style_Sheets, 04.03.2017.

Text der Karteikarten, so wie die dynamische Aufschlüsselung aller auftretenden Paragraphen-Angaben als Tooltip.

- Die Möglichkeit des Entdeckens und des RD, dem sich das folgende Kapitel widmet.

Das Zusammenführen von XSL-Transformationen, digitalen Objekten und Zuordnungen zu Context-Objekten gilt im Grunde für alle Projekte, die in dieser Umgebung umgesetzt werden und ist ein essentieller Teilaspekt der Langzeitarchivierung, die auch nachhaltige und standardisierte, sowie nachvollziehbare Workflows bedingt. Das Ergebnis des letzten Stylesheets und die Ausgangsdaten dazu werden ausführlicher im folgenden Kapitel beschrieben.

5.4 Technische Umsetzung der semantischen Suche und Discovery Möglichkeiten

Dieses Kapitel behandelt die technische Implementierung der Information Retrieval bzw. Resource Discovery Funktionalitäten der digitalen Sammlung und erstreckt sich von der Aufbereitung der digital edierten Karteikarten als RDF-Graphen, über ihre Abfrage mittels SPARQL bis zur Ausgabe der Suchtreffer in HTML, um der Nutzerin Möglichkeiten anzubieten, sich im Bestand explorativ verhalten und stöbern zu können. Die Umsetzung basiert auf semantischen Technologien, die in folgender Einführung theoretisch erörtert werden.

5.4.1 Einführung: Semantische Technologien

Im **Semantic Web**, so ist die Idee, sollten nicht nur Menschen miteinander kommunizieren, sondern auch Maschinen mit Menschen und Maschinen mit anderen Maschinen. BURNERS-LEE beschreibt das Semantic Web als

*"[...] the Web of data with meaning in the sense that a computer program can learn enough about what the data means to process it."*²⁴⁵

Dazu sind folgende drei Prinzipien notwendig: Alles, sowohl digital wie analog, ist eine *resource*. Jede *resource* verfügt über eine eindeutige Kennung in Form eines Uniform Resource Identifiers (URI), der zum Beispiel eine ISBN für ein analoges Buch oder eine URL für einen Webinhalt darstellt. Für jede *resource* gibt es Datensätze, die in einer formalen Sprache - beispielsweise XML - etwas über die *resource* aussagen. Das **Resource Description Framework (RDF)**²⁴⁶ ist ein Werkzeug um Daten in Form von Graphen zu formalisieren. Graphen bestehen aus Knoten, den genannten *resources*, und aus Kanten, den Beziehungen der Knoten miteinander. Die Aussagen, die in RDF ausgedrückt werden können, haben die Form Subjekt-Prädikat-Objekt. RDF stellt eine grundlegende Möglichkeit der Formalisierung von Daten in Form von Graphen dar.²⁴⁷ Ein Buch (Subjekt) behandelt (Prädikat) den Schriftsteller William Shakespeare (Objekt). Folgender Quellcode in XML kodierten RDF beschreibt diese Beziehung und Abbildung 17 stellt sie dar.

²⁴⁵BURNERS-LEE, T./FISCHETTI, M.: Weaving the Web: The Past, Present and Future of the World Wide Web by Its Inventor. London, 2000, S.260.

²⁴⁶Resource Description Framework, <https://www.w3.org/RDF/>, 20.12.2016.

²⁴⁷Vgl. VOGELER, Georg et al.: The Content of Accounts and Registers in their Digital Edition. XML/TEI, Spreadsheets, and Semantic Web Technologies. In SARNOWSKY, Jürgen (Hrsg.): Konzeptionelle Überlegungen zur Edition von Rechnungen und Amtsbüchern des späten Mittelalters. Göttingen, 2016, S.23.

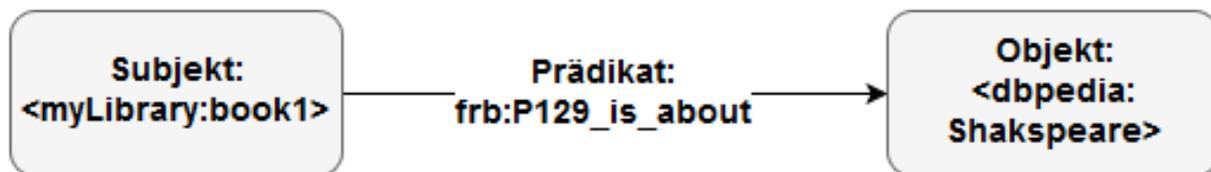


Abbildung 17: Beispiel für RDF

```

1 <?xml version="1.0"?>
2 <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
3     xmlns:frbr="http://purl.org/vocab/frbr/core#"
4     xmlns:dbpedia="http://dbpedia.org/resource/">
5 <rdf:Description rdf:about="http://www.myLibrary/book1">
6   <frbr:P129_is_about rdf:resource="http://dbpedia.org/resource/William_Shakespeare"/>
7 </rdf:Description>
8 </rdf:RDF>

```

Subjekt, Prädikat und Objekt werden mittels URI referenziert.²⁴⁸ Dabei dient jeweils die Angabe *myLibrary:*, *frbr:* und *dbpedia:* als Namespace, also als Kurzschreibweise für die URI's. Der Namespaces *frbr:* verweist auf eine Form der Beziehung, die in einer Ontologie definiert wurde. Der Term 'is_about' ist im FRBR-Model so definiert, dass ein Objekt (ein Buch) von einer anderen Entität (William Shakespear) handelt. Der Namespace im Zusammenhang mit William Shakespear verweist auf eine externe Datenquelle, die DBpedia. Das Ziel von DBpedia ist, es strukturierte Informationen aus Wikipedia zu extrahieren und diese anderen Webanwednungen zur Verfügung zu stellen. Daraus entsteht sogenannte **Linked Data**.²⁴⁹

Die **SPARQL Protocol And RDF Query Language (SPARQL)**²⁵⁰ ist eine graphenbasierte Abfragesprache für RDF Datensätze.²⁵¹ Eine kurze Erklärung der Syntax: Mit einem vorangestellten „?“ werden Variablen gekennzeichnet, die als Ergebnis der Anfrage zurückgegeben werden. Aus Gründen der Leserlichkeit werden Präfixe definiert. Für projektspezifische Relationen wurde das Präfix *km:* eingeführt.²⁵² Weiters gibt es mehrere Möglichkeiten den Graph mittels sogenannter *Graph Patterns*, wie beispielsweise UNION (Verbindung von Teilgraphen), zu erschließen und die Abfragen zu erweitern.²⁵³ Im Anhang unter Kapitel 7.1.4 findet man ein Source-Code-Beispiel einer SPARQL-Abfrage aller RDF-Datenströme von Karteikarten. In Kapitel 5.4.2 wird weiter ausgeführt, wie die RDF-Daten der digital edierten Karteikarten abgefragt werden, um sie einer Nutzerin für das RD zugänglich zu machen.

Simple Knowledge Organization System (SKOS)²⁵⁴ ermöglicht es Vokabularien in Form von

²⁴⁸Uniform Resource Identifiers, https://de.wikipedia.org/wiki/Uniform_Resource_Identifier, 04-03.2017.

²⁴⁹Vgl. JORDANOUS: Enhancing information retrieval and resource discovery from data using the Semantic Web, S.105-106.

²⁵⁰SPARQL, <https://www.w3.org/TR/rdf-sparql-query>, 05.02.2016.

²⁵¹Vgl. LINCKELS/MEINEL: E-librarian service: user-friendly semantic search in digital libraries, S.30-42.

²⁵²Eine Modellierung nach einer vorhanden bzw. standardisierten, wie etwa CIDOC-CRM, hätte sich im Nachhinein für diesen Zweck angeboten.

²⁵³Vgl. DUCHARME, Bob: Learning SPARQL. Sebastopol, 2011, S.1-15.

²⁵⁴SKOS, <https://www.w3.org/2004/02/skos/>, 05.02.2017.

Taxonomien zu definieren. Abstrakte Konzepte (*skos:concept*) können in eine hierarchische Struktur mit untergeordneten und übergeordneten Konzepten, zusätzlichen Verweisen, oder unterschiedlichen Bezeichnungen eingebettet werden. Jedes Konzept kann mit einer URI referenziert werden.²⁵⁵ Im Zuge des Studienprojektes wurde an einem Objektthesaurus mitgearbeitet, in dem jedes Objekt und seine Unter- bzw. Oberbegriffe in SKOS modelliert wurden. Folgender Ausschnitt aus einem RDF einer Karteikarte zeigt die SKOS-Konzepte eines Gewehrs, das in einer Karteikarte angeführt wird.

```

1 <skos:Concept rdf:about="http://gams.uni-graz.at/skos/scheme/o:oth/#18050211">
2   <skos:broader rdf:resource="http://gams.uni-graz.at/skos/scheme/o:oth/#18050210"/>
3   <skos:prefLabel xml:lang="deu">Gewehr</skos:prefLabel>
4   <skos:altLabel xml:lang="en">rifle</skos:prefLabel>
5 </skos:Concept>

```

Das Attribut *rdf:about* definiert den eindeutigen URI, der das *skos:concept*, also die abstrakte Klasse des Objekttyps Gewehr, darstellt. Die Elemente *skos:prefLabel* und *skos:altLabel* führen unterschiedliche Schreibweisen bzw. Benennungen in anderen Sprachen an, die über das Attribut ausgedrückt werden. Die hierarchische Struktur wird mit *skos:narrower* (untergeordnet) und *skos:broader* (übergeordnet) erzeugt, in dem das dortige Attribut mit einem URI auf ein weiteres Konzept verweist. Dieses Konzept verfügt über keine weiteren Unterkonzepte, ist aber ein Kind-Konzept des Konzepts Feuerwaffen mit der ID *18050210*.

Ein weiterer Baustein des Semantic Webs sind **Ontologien**.²⁵⁶ LINCKELS und MEINEL definieren eine Ontologie als ein Datenmodell zur Darstellung eines Sets miteinander vernetzter Konzepte innerhalb einer (Fach-)Domäne.²⁵⁷ WELLER spricht von einer formalen und schematischen Darstellung einer Wissensdomäne auf Basis definierter Regeln und Vokabulars. Mittels einer Ontologie lassen sich komplexe Sachverhalte so darstellen, dass auch Maschinen, die durch die Ontologie definierte und standardisierte, Struktur interpretieren können. Der Mehrwert kann vor allem in der Möglichkeit automatisierter Schlussfolgerungen auf Basis der Ontologie bestehen. Dazu werden sogenannte Ontologiesprachen, wie etwa die **Web Ontology Language (OWL)** verwendet.²⁵⁸

Eine modellhafte Skizze des Semantic Web und der sehr knapp gefassten Bestandteile lässt sich mit dem **Semantic Web Stack**, der in Abbildung 18 visualisiert ist, zusammenfassen. Zur Verwirklichung des Semantic Webs müssen alle Teilbereiche umsetzbar sein, wobei dies für alle Bereiche, die über Ontologien hinausgehen, noch nicht möglich ist.²⁵⁹

²⁵⁵Vgl. VOGELER et al.: The Content of Accounts and Registers in their Digital Edition. XML/TEI, Spreadsheets, and Semantic Web Technologies, S.32-33.

²⁵⁶Ontology, <https://www.w3.org/standards/semanticweb/ontology>, 05.02.2016.

²⁵⁷Vgl. LINCKELS/MEINEL: E-librarian service: user-friendly semantic search in digital libraries, S.18-24.

²⁵⁸Vgl. WELLER, Katrin: Ontologien. In KUHLEN, Rainer/SEMAR, Wolfgang/STRAUCH, Dietmar (Hrsg.): Grundlagen der praktischen Information und Dokumentation: Handbuch zur Einführung in die Informationswissenschaft und -praxis. 6. Auflage. Berlin and Bosten, 2013, S.207-218.

²⁵⁹Vgl. LINCKELS/MEINEL: E-librarian service: user-friendly semantic search in digital libraries, S.15-24.

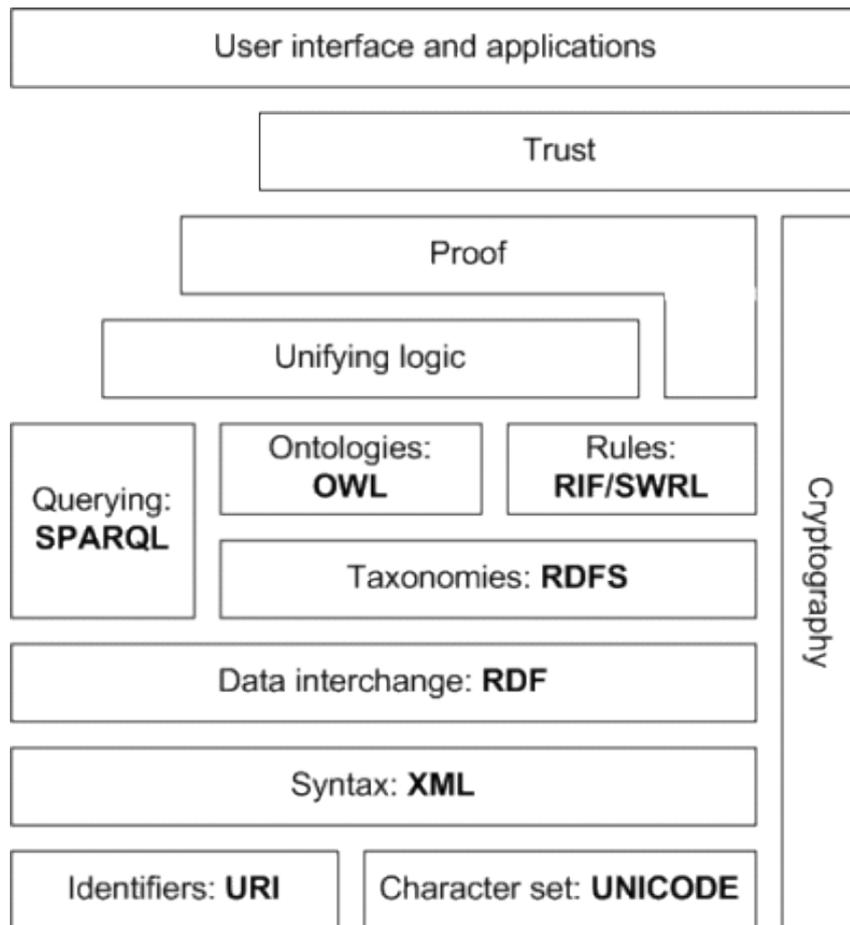


Abbildung 18: Semantic Web Stack

VOGELER unterstreicht den besonderen Mehrwert semantischer Technologien für geisteswissenschaftliche Fragestellungen. In seinen Ausführungen zur digitalen Edition von Rechnungsbüchern, sieht er in den semantischen Technologien eine Möglichkeit die textuelle, bildliche und vor allem die inhaltliche Ebene von Quellenmaterial zu erschließen und anderen zur Verfügung zu stellen. Dieser Gedanke lässt sich auf andere Quellenmaterialien und geisteswissenschaftliche Projekte ummünzen.²⁶⁰ Im Studienprojekt wurden Karteikarten des Hans Gross Kriminalmuseums auf einer textuellen Ebene (Transkription + TEI), auf einer bildlichen Ebene (Scan und Viewer) und auf einer inhaltlichen Ebene (RDF) erschlossen. Im Kapitel 5.4.2 wird der Prozess der Erschließung und semantischen Modellierung in RDF und in Kapitel 5.4.3 die Abfrage der Daten mittels SPARQL ausführlich beschrieben.

5.4.2 Aufbereitung der Karteikarten: TEI/XML nach RDF/XML

Sämtliche Informationen, die in den Karteikarten liegen, sind nach dem TEI-Standard in XML kodiert und erschlossen. Das TEI/XML der Karteikarte *o:km.87*²⁶¹ ist im Anhang angeführt und

²⁶⁰Vgl. VOGELER et al.: The Content of Accounts and Registers in their Digital Edition. XML/TEI, Spreadsheets, and Semantic Web Technologies, S.41.

²⁶¹Aufrufbar unter gams.uni-graz.at/o:km.87

dient exemplarisch der weiteren Ausführungen (siehe 7.1.1).

Beim Ingest wird ein zusätzliches XSLT-Stylesheet (siehe 7.1.2) aktiv, das aus jeder digital edierten Karteikarte einen, in RDF/XML kodierten, Graphen erzeugt. Dabei werden die wichtigsten Informationen, die für die Retrieval-Prozesse Verwendung finden, herausgefiltert und strukturiert. Der Source-Code, des auf diese Weise generierten RDF/XML, befindet sich ebenfalls im Anhang (siehe 7.1.3). Folgende Abbildung stellt den Graphen des Objekts *o:km.87* dar.

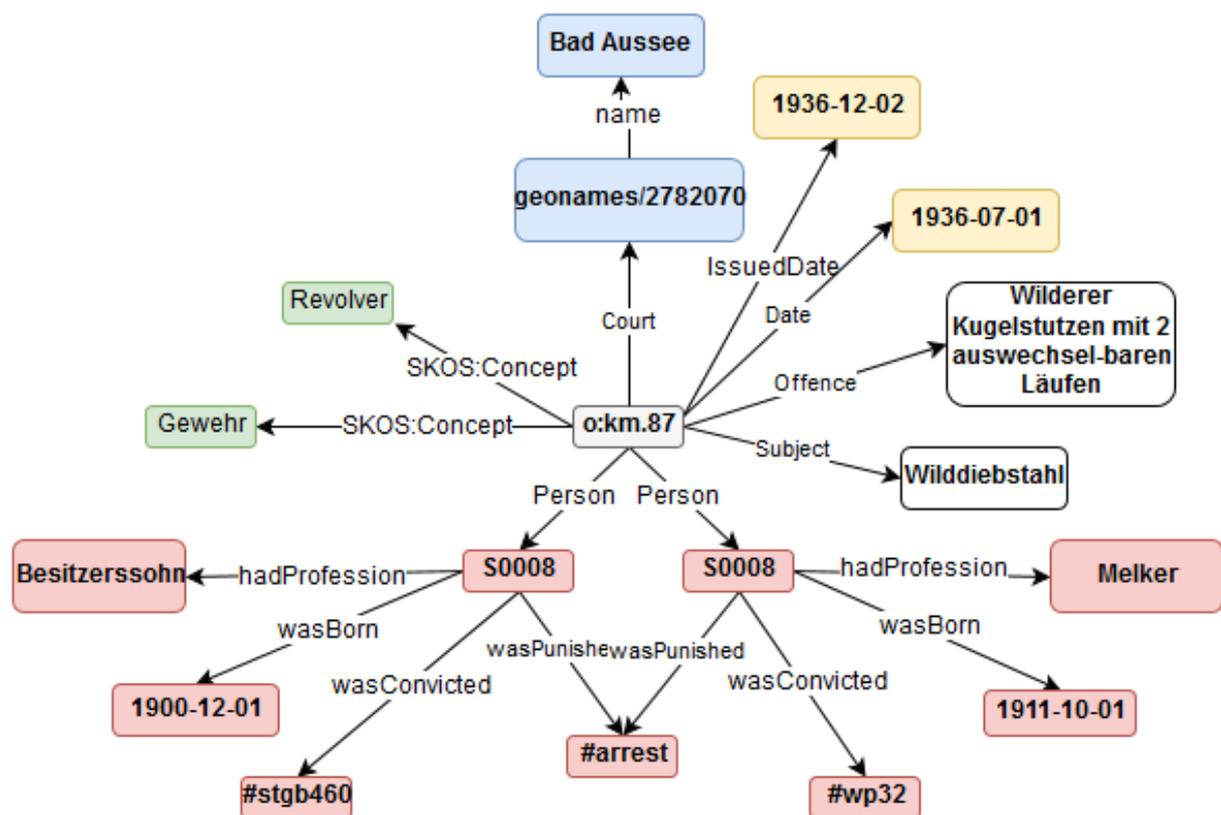


Abbildung 19: RDF-Graph einer Karteikarte.

Grün markiert sind zwei Referenzen, die auf den **SKOS-Objektthesaurus** verweisen. In der dargestellten Karteikarte wurden ein 'Revolver' und ein 'Gewehr' als Objekte in das Kriminalmuseum aufgenommen. Für beide wurde ein eigenständiges Konzept im Objektthesaurus definiert und wird im RDF referenziert. In Kapitel 5.4.1 wird ein Beispiel des SKOS-Thesaurus angeführt. Dies ermöglicht es theoretisch die hierarchische Struktur der Objekt- bzw. Materialklassen nachvollziehen zu können.

Blau sind **Orte**, wobei zwischen zuständigem Gerichtsort und anderen Orten unterschieden wird. Da Informationen über Orte und auch über Datumsangaben variabel und stark kontextbezogen sind, ist eine generische Verarbeitung - nach derzeitigen Auszeichnungsgrad der Karteikarten - nur bedingt möglich. Jedoch sind es Informationen, die für eine Suche oder andere Darstellungsformen interessant sein könnten. Die Orte werden mittels GeoNames-Referenz²⁶²

²⁶²GeoNames, <http://www.geonames.org>, 30.12.2016.

angesprochen. Die GeoNames-Referenzen sind nicht nur geeignete Normdatensätze um Daten zu normalisieren, sondern werden beim Ingest in Koordinaten aufgelöst, die wiederum in einem Geo-Viewer dargestellt werden können und es ermöglichen die Karteikarte zu verorten. Im Kriminalfall dieser Karteikarte war das zuständige Gericht in **Bad Aussee**.²⁶³

Gelb sind sämtliche **Datumsangaben**, bis auf das Geburtsdatum einer Person, da dies als personenspezifische Angaben mit der jeweiligen Person verknüpft ist. Wenn vorhanden wurde ein Ausstellungsdatum (*IssuedDate*) der Karteikarte als eigenständige Entität ausgezeichnet. Die Karteikarte lässt sich auf den **02.12.1936** datieren. Daneben existieren noch weitere Datumsangaben, für die das gleiche gilt, wie für nicht weiter spezifizierte Orte. Diese Datumsangaben sind zu stark kontextabhängig, als dass man sie generisch ohne größeren Aufwand behandeln könnte. Aus Gründen des Projektplanes wurde im Studienprojekt davon abgesehen. In diesem Fall handelt es sich bei **1936-07-01** um das Datum der Tat.

Rot sind die in der Karteikarten angeführten **Personen**. Sie verfügen über eine eindeutige interne ID. Im Fall dieser Karteikarte sind es die Personen **S0007** und **S0008**. Aus rechtlichen Gründen dürfen die Originalnamen der Personen nicht angegeben werden. RDF-Einträge zu Personen unterteilen sich in weitere Bestandteile, bestehend aus:

- **Alter:** hadAge
- **Beruf:** hadProfession
- **Vorstrafen:** hadPoliceRecord
- **Urteil:** wasConvicted
- **Strafe:** wasPunished

Bei Urteilen und Strafen verweist eine Referenz auf ein eigenständiges, statisches TEI-XML, in dem die Paragraphen aufgeschlüsselt werden.²⁶⁴ Altersangaben werden aus der Differenz zwischen Ausstellungsdatum und Geburtsdatum berechnet. Auch existieren Karteikarten, in denen das Alter direkt angegeben wurde und nicht das Geburtsdatum. Für Berufe wurde eine weitere Einordnung in Berufsgruppen vorgenommen, die sich am Beispiel von SCHINDLER orientiert, um in einer Auswertung der Karteikarten über einen Vergleichsdatensatz zu verfügen. SCHINDLER hat Gerichtsakten des Gollinger Pfleggericht ausgewertet, einem Bestand, der mit dem des Kriminalmuseums vergleichbar ist.²⁶⁵

Für jede Karteikarte existiert ein solcher RDF-Datenstrom, der aber in seiner Informationsfülle stark variieren kann. Beispielsweise gibt es Karteikarten, deren einziger Inhalt die Bezeichnung der Tatwaffe bzw. des Beweisstückes ist. Andere wiederum beschreiben ausführlich mehrere Täter und Tathergänge.

²⁶³GeoNames-ID, <http://www.geonames.org/2782070>, 30.12.2016.

²⁶⁴Paragraphen, http://gams.uni-graz.at/o:km.paragraphen/TEI_SOURCE, 30.12.2016.

²⁶⁵Vgl. SCHINDLER, Norbert: Wilderer im Zeitalter der Französischen Revolution: ein Kapitel alpiner Sozialgeschichte. München, 2001, S.41-83.

5.4.3 Abfrage der Karteikarten: SPARQL

Beim Ingest jeder Wilderer-Karteikarte wird die im vorangegangenen Kapitel beschriebene XSL-Transformation wirksam. Die dadurch generierten RDF/XML werden als RDF-Tripel via Webservice im Java Framework **OpenRDF Sesame**²⁶⁶ gesendet und angelegt. In der GAMS-Infrastruktur ist es möglich ein Objekt vom Typ *cirilo:Query* zu verwenden. Diese Objekt ermöglicht es Abfragen für in Sesame angelegten RDF-Tripel zu formulieren, zu verwalten und auszugeben. Die Abfragen werden in SPARQL 1.1 formuliert, laufen über alle RDF/XML bzw. Tripel im Tripelstore, und verfügen über drei unterschiedliche GETTER-Methoden, die ein XML, ein HTML oder ein Format für Tabellenkalkulationen zurückgeben können.²⁶⁷

Dadurch wird die Abfrage mittels SPARQL und die Darstellung komplexerer semantischer Strukturen, beispielsweise von Ontologien, ermöglicht. Ein weitere Triplestore Datenbank ist **Blazegraph**²⁶⁸. Dieser ermöglicht es mittels Volltext-Index aller ingestierten digitalen Objekte eine Volltextsuche umzusetzen.²⁶⁹ Blazegraph, als graphenbasierte Datenbank, verfügt über eine Schnittstelle, um Abfragen in SPARQL formulieren zu können.²⁷⁰

Die SPARQL-Query (siehe 7.1.4) ist im Fall der Karteikarten simpel. Auf der HTML-Repräsentation des RD, müssen alle Informationen auf einer Seite zusammen kommen. Aus diesem Grund umfasst das Ergebnis der SPARQL alle Objekt-Subjekt-Beziehungen aller RDF-Karteikarten. Auch komplexere Abfragen sind auf diese Art und Weise möglich, wie z.B. eine Abfrage aller männlichen und vorbestraften Arbeiter, die gegen das Waffenpatent verstoßen haben. Folgender Ausschnitt veranschaulicht das Ergebnis für die RD-Webseite:²⁷¹

```
1 <result>
2   <indexcard uri="http://gams.uni-graz.at/o:km.87"/>
3   <relation uri="http://gams.uni-graz.at/km/#Person"/>
4   <object uri="http://gams.uni-graz.at:8980/bigdata/S0007"/>
5   <profession>Besitzerssohn</profession>
6 </result>
7 <result>
8   <indexcard uri="http://gams.uni-graz.at/o:km.87"/>
9   <relation uri="http://gams.uni-graz.at/km/#Person"/>
10  <object uri="http://gams.uni-graz.at:8980/bigdata/S0008"/>
11  <profession>Melker</profession>
12 </result>
```

²⁶⁶Sesame, <http://rdf4j.org>, 05.02.2017.

²⁶⁷How to semantically enrich your data, <http://gams.uni-graz.at/docs#semantic>, 13.02.2017.

²⁶⁸Blazegraph, <https://www.blazegraph.com/>, 31.12.2016.

²⁶⁹Volltextsuche mit blazegraph, <https://wiki.blazegraph.com/wiki/index.php/FullTextSearch>, 31.12.2016.

²⁷⁰How to semantically enrich your data, GAMS Dokumentation, <http://gams.uni-graz.at/docs>, 13.02.2017.

²⁷¹XML Query-Ergebnis, <http://gams.uni-graz.at/archive/objects/query:km.wildererkk/methods/sdef:Query/get?mode=query>, 05.02.2017.

Das Ergebnis der SPARQL ist in XML kodiert. Das Element *indexcard* zeigt an, dass sich die folgende Relation auf die Karteikarte *o:km.87* bezieht und ermöglicht es für jede ausgegebene Information die URL zur Detailansicht des Objekts auf der RD-Webseite zu referenzieren. Weiters umfasst das Ergebnis die Aussage, dass die Personen *S0007* und *S0008* den Beruf *Besitzerssohn* und den Beruf *Melker* ausübten. Vergleicht man das mit dem TEI²⁷² und mit der Darstellung des Graphen²⁷³, so ist das Ergebnis der SPARQL Query konsistent.

5.4.4 Resource Discovery der Karteikarten: SPARQL nach HTML

Mit Zuweisung einer XSL-Transformation (siehe 7.1.5) auf das Query-Objekt, wird aus dem so erzeugten XML ein beliebiges HTML transformiert, das in die Webpräsenz der digitalen Sammlung integriert wird. Auf Basis der so generierten Daten werden verschiedene Visualisierungs- und Discovery-Szenarien umgesetzt, die dem interessierten 'Laien', aber auch einem wissenschaftlichem Publikum, einen Zugang zu dieser digitalen Quellensammlung ermöglicht und sie beim 'digitalen Stöbern' durch einen Teilbereich eines Sammlungsbestandes unterstützen. Das Ergebnis ist eine Webseite, die es erlaubt, im Bestand der Karteikarten nach vier unterschiedlichen Dimensionen zu stöbern. Diese vier Dimensionen, die im Folgenden weiter ausgeführt werden, sind: **Personen, Delikte, Raum und Zeit**.

Berufsgruppen und Altersverteilung²⁷⁴

Ein erster Überblick über den Bestand der Karteikarten zu Wildereidelikten erfolgt mit einer soziographischen Analyse. Angelehnt an und vergleichend mit SCHINDLER²⁷⁵, gibt eine statistische Auswertung der Berufs- und Altersstruktur der Täter (und einer Täterin) in Wildereidelikten. Die Nutzerin hat so die Möglichkeit einen besseren Überblick darüber zu bekommen aus welchem sozialen Umfeld Täter von Wildereidelikten kamen und erhält auch einen ersten Eindruck darüber, welche Informationen in Karteikarten enthalten sind. Die Auswertung zur Berufsstruktur wird in einem Balkendiagramm dargestellt, auf dem der linke Balken die Daten von SCHINDLER darstellt und der rechte Balken, die Daten aus den Karteikarten. Die Berufsgruppen wurden in Anlehnung an SCHINDLER definiert und bei der Datenerfassung in TEI normalisiert. Die Auswertung erfolgt in der XSL-Transformation generisch und 'on-the-fly' und kann, bei gleicher Datenausgangssituation, theoretisch auf alle Karteikarten des Kriminalmuseums umgemünzt werden. Die gleiche Modellierung und Verarbeitung gilt für die rechts dargestellten Auswertung zur Altersstruktur als Flächendiagramm. Sichtbar wird zum einen die These SCHINDLER's, dass junge Söhne von Bauern bzw. hofbesitzende Bauern im Alter von 20-24 den Kern der Wilderer gebildet haben und das im Kriminalmuseums-Bestand, der ca. 100

²⁷²siehe Source-Code 7.1.1.

²⁷³siehe Abbildung 19.

²⁷⁴siehe <http://gams.uni-graz.at/archive/objects/query:km.auswertung/methods/sdef:Query/get?#auswertung>

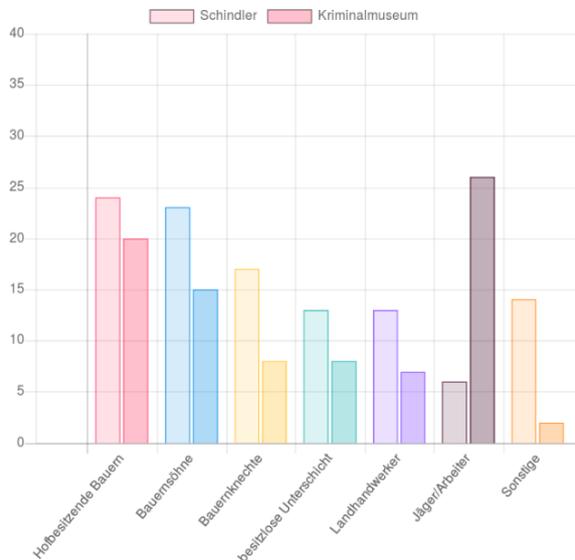
²⁷⁵Vgl. SCHINDLER: Wilderer im Zeitalter der Französischen Revolution: ein Kapitel alpiner Sozialgeschichte, S.41-83.

Jahre später entstanden ist, die Gruppe der 'Arbeiter' hinzugekommen ist.

Berufsgruppen und Altersverteilung

Hier finden Sie eine statistische Auswertung der Karteikarten auf der rechten Seite im Vergleich zu den von Norbert SCHINDLER erschlossenen Daten. [SCHINDLER Norbert, Wilderer im Zeitalter der Französischen Revolution: Ein Kapitel alpiner Sozialgeschichte, München 2001]

Berufsgruppen im Pfliegergericht Golling nach SCHINDLER



Altersverteilung im Pfliegergericht Golling nach SCHINDLER

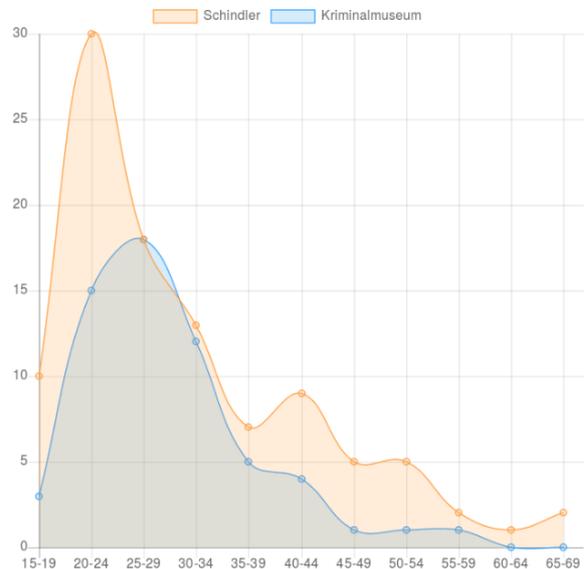


Abbildung 20: Berufsgruppen und Altersverteilung, Screenshot des Resource Discovery.

Zur Implementierung wurde die Open Source JavaScript Bibliothek **Chart.js**²⁷⁶ verwendet. Es verfolgt den Zugang der Informationsvisualisierung (siehe 4.5) großer Datenmengen in Form von statistischen Auswertungen und der Darstellung in Diagramm-Form. Die Bibliothek ist auf Grund der ausführlichen Dokumentation und Code-Beispiele einfach zu verwenden und kann flexibel an die Bedürfnisse angepasst werden. Der JavaScript-Code wird direkt in das XSL-Stylesheet eingefügt.

Mit dieser statistischen Auswertung wird eine Perspektive auf die Gesamtheit dieses Teilbestandes gelegt. Die Information welche Personen, welchen Berufsgruppen zugehörig sind und in welches Alter sie haben, kann bereits einen Überblick liefern, ob der Quellenbestand für eine Nutzerin relevant ist.

Nach Delikten stöbern²⁷⁷

Der Zugang nach Delikten zu stöbern fokussiert sich nun auf die Einzelobjekt, im Gegensatz zum vorher beschriebenen Zugang zur Gesamtheit der Objekte. Hier hat die Nutzerin die Möglichkeit durch alle 'Vorderseiten' der Karteikarten zu browsen. Auf jeder Vorderseite einer Karteikarte werden die Punkte *Gegenstand* und *Delikt* angeführt, die im RDF als *km:Subject* und *km:Offence* erschlossen wurden. Mit Klick auf den orange hervorgehobenen *Gegenstand* kommt man zur Detailansicht der jeweiligen Karteikarte.

²⁷⁶Chart.js, <http://www.chartjs.org/>, 13.02.2017.

²⁷⁷siehe <http://gams.uni-graz.at/archive/objects/query:km.auswertung/methods/sdef:Query/get?#liste>

Liste der Delikte



Gegenstand: Einläufiges Gewehr	Gegenstand: Gewehr (Jagdgewehr)	Gegenstand: 2 Wilderermasken
Delikt: Versuchter Wildddiebstahl § 8. 460 Stg. u § 36 WP.	Delikt: Wildddiebstahl §§ 460, 464 Stg. u § 36 WP.	Delikt: Wildddiebstahl; § 36 WP.
Täter: Landwirt;	Täter: Hilfsarbeiter; Knecht; Besitzersohn; Bes. Sohn; Schuhmacher;	Täter: Besitzersohn;
Gegenstand: 2 St. kurze Zündschnüre (zu St. A. 3479)	Gegenstand: Lancaster-Gewehr	Gegenstand: Buchsflinte
Delikt: Brandlegung	Delikt: Wildddiebstahl	Delikt: Wildddiebstahl u. Tragen einer verbotenen Waffe.
Täter:	Täter:	Täter:

Abbildung 21: Liste der Delikte, Screenshot des Resource Discovery.

Dies erlaubt eine Ansicht des Sammlbereiches *context:km.karteikarten*, der über den digitalen Objekten zugeordneten Metadaten hinausgeht und zentrale Inhalte für das Discovery sichtbar macht.

Nach Orten stöbern²⁷⁸

Alle Karteikarten sind im Datenstrom *context:km.karteikarten* zusammengefasst. In diesem Datenstrom existiert ein KML-Datenstrom²⁷⁹. Unter **Keyhole Markup Language (KML)**²⁸⁰ versteht man eine in XML-Syntax kodierte Aufzeichnungssprache zur Beschreibung von Geodaten, die beispielsweise von Google Earth verwendet wird. Der KML-Datenstrom ist das Ergebnis aller aufgelösten GeoNames-Referenzen, der dem *context:km.karteikarten* zugeordneten TEI's. Dariah stellt einen Geo-Browser zur Verfügung, der 'Geistes- und Kulturwissenschaften für die Analyse von Raum-Zeit-Relationen von Daten zur Verfügung' steht.²⁸¹ Als Webtool kann der Geo-Browser in die Webseite eingebettet werden und kann den KML-Datenstrom darstellen. Das Ergebnis wird in Abbildung 22 dargestellt. Zu erkennen sind orange Punkte, die jeweils ein Gericht symbolisieren, das für das jeweilige Delikt zuständig war. Bei Klick auf einen orangenen Punkt öffnet sich ein Fenster mit einem Thumbnail der ersten Seite der Karteikarte, das mit einem weiteren Klick auf die Detailansicht verlinkt. Weiters verfügt der Geo-Browser über eine Zeitleiste, mit der die zeitliche Dimension erfasst werden kann.

²⁷⁸siehe <http://gams.uni-graz.at/archive/objects/query:km.auswertung/methods/sdef:Query/get?#orte>

²⁷⁹siehe <http://gams.uni-graz.at/archive/objects/context:km.karteikarten/datastreams/KML/content>

²⁸⁰Keyhole Markup Language, <https://developers.google.com/kml/documentation/kmlreference?csw=1>, 31.12.2016

²⁸¹Dariah Geo-Browser, <https://de.dariah.eu/geobrowser>, 31.12.2016.

Orte der zuständigen Gerichte

Per Klick auf einen orangen Kreis werden Ihnen alle Karteikarten angezeigt, die von diesen Gericht bearbeitet wurden.

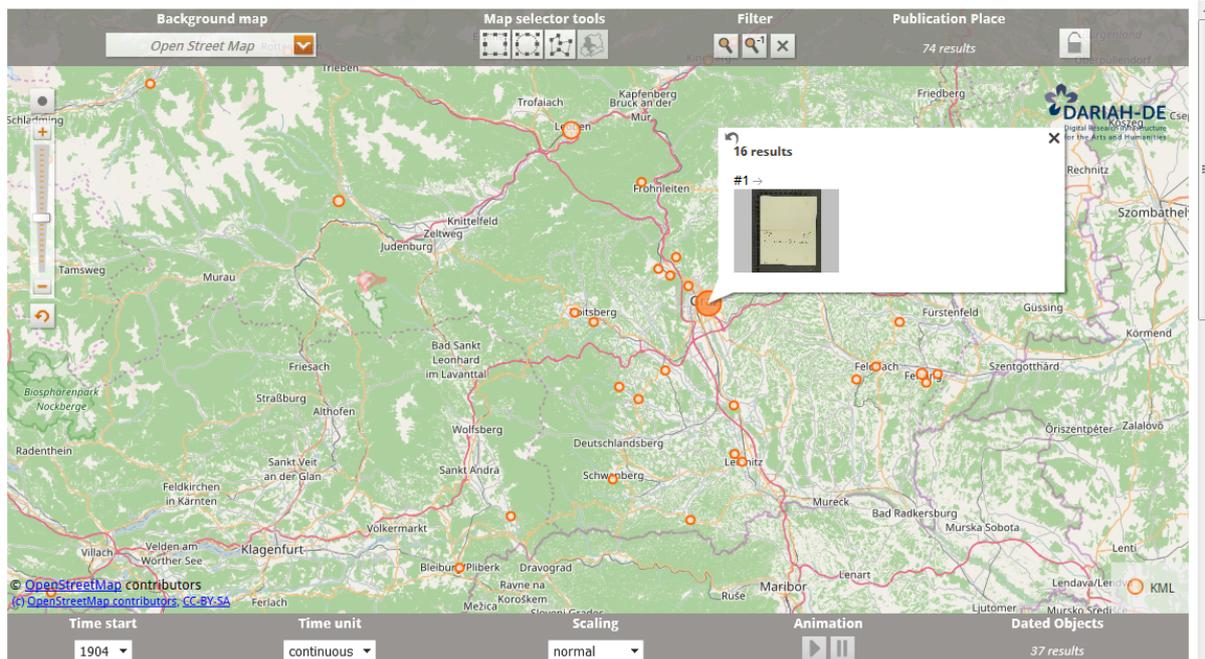


Abbildung 22: Gerichtsorte im eingebetteten Dariah Geo-Browser, Screenshot des Resource Discovery.

Nach Datum stöbern²⁸²

Eine weitere Möglichkeit die zeitliche Dimension des Teilbestands zu erfassen bietet der letzte Bereich der RD-Seite. Dieser ermöglicht einen Überblick über alle Datumsangaben, die mit der Ausstellung einer Karteikarte verbunden sind. Dargestellt auf auf einem Zeitstrahl repräsentiert jeder Punkt eine Karteikarte. Für jedes *km:IssuedDate*, das in einem RDF einer Karteikarte vorkommt, wird ein Punkt auf dem Zeitstrahl hinzugefügt. Dabei wird jeder Eintrag mit *Gegenstand* und *Delikt* versehen, sowie mit einer Verlinkung zur Karteikarte. Abbildung 23 vermittelt einen Eindruck des Ergebnisses.

Nach Ausstellungsdatum

Mit Klick auf ein Delikt kommen Sie zur Detailsicht der Karteikarte. Sie können sich nach links und rechts mit Ziehen der Maus bewegen.

[Zurück zur ganzen Ansicht](#)



Abbildung 23: Zeitstrahl der Datumsangaben der Ausstellung, Screenshot des Resource Discovery.

²⁸²siehe <http://gams.uni-graz.at/archive/objects/query:km.auswertung/methods/sdef:Query/get?#datum>

Umgesetzt wird der Zeitstrahl mit der JavaScript Bibliothek *vis.js*.²⁸³ Der dazu nötige JavaScript Source Code wurde im XSLT-Stylesheet eingefügt.²⁸⁴ Durch das Scrollen im Zeitstrahl kann man von einem Überblick über Jahre bis auf Monate und Tage hinein scrollen.

Bei jeder 'Discovery-Dimension' handelt sich um eine andere Repräsentationen von Dokumenten bzw. digitaler Objekte aus einem zuvor definierten Teilbestand. Die funktionelle Ebene dieser Repräsentationen impliziert weder ein Ranking noch die Relevanz der Dokumente hinsichtlich eines Informationsbedürfnisses eines kognitiven Agenten, sondern dient rein der Nutzerentscheidung, ob dargestellte Dokumente relevant oder nicht relevant sind. Es ist nur bedingt das Ziel von Discovery-Funktionalitäten Relevanz abzubilden. Zum einen weil es kein ausformuliertes Informationsbedürfnis gibt, zum anderen da dies bewusst dem Nutzer und seinem explorativen Wesen vorbehalten bleiben soll. Es kann gesagt werden, dass der Recall-Wert groß ist, alle Dokumente des Teilbestands dargestellt werden und der Precision-Wert, als Wert der Vollständigkeit über alle relevanten Dokumente, niedrig oder nicht vorhanden ist.

Die Indexierung der Dokumente wird in diesem Fall des RD aus den TEI/XML-Auszeichnung extrahiert. Die Information ob es sich bei einer Karteikarte um ein Wilderei-Delikt handelt, ob eine Darstellung der Dokumente auf Ebene der Entitäten Person, Delikt, Raum oder Zeit erfolgt, hängt davon ab, ob diese Aspekte in der Karteikarte ausgezeichnet und gleichzeitig für das RD kontrolliert indexiert wurden. Wurde beispielsweise ein Objekt (z.B. ein Gewehr) in einem Delikt ausgezeichnet und über ein Attribut in XML eine Referenz zum SKOS-Objektthesaurus hergestellt so sieht man die kontrollierte Indexierung dieser Dokumente auf Basis einer ausdrucksstarken Wissensbasis, wie einem Thesaurus.

²⁸³ JavaScript-Bibliothek, visjs.org, 31.12.2016.

²⁸⁴ siehe Source-Code 7.1.5

6 Fazit und Ausblick

*"Verachte nur Vernunft und Wissenschaft
des Menschen aller höchste Kraft.
Lass nur in Blend und Zauberwerken
dich von dem Lügengeist bestärken."
- GOETHE, Faust I*

Wie der Beginn dieser Arbeit mit Faust's frustrierender Erkenntnis, dass er nichts wissen kann, beginnt, so endet diese Arbeit mit den Worten seines Gegenspielers. Mephistos sieht in der Vernunft- und Erkenntnisfähigkeit des Menschen, in Form der Wissenschaft, dessen bedeutendste Fertigkeit und nimmt den 'postfaktischen Lügengeist' Jahrhunderte vorweg.

Die Erkenntnisfähigkeit des Menschen ist Voraussetzung dafür, dass er sich im ständigen Abgleich und in ständiger Interpretation der Wirklichkeit befindet. Das ganze Leben eines jeden Menschen ist ein ständiges Verarbeiten von Daten und Informationen auf Basis der ihm zur Verfügung stehenden Konzepte und Modelle - seinem Wissen. Ein entscheidender Aspekt dabei sind Kommunikationsprozesse, die es jedem Individuum ermöglichen im Austausch mit anderen zu stehen. Kommunikation prägt und ermöglicht Erkenntnis. Um diese Kommunikationsprozesse effizient zu gestalten und sie im kollektiv nutzbar machen zu können, bedarf es gemeinsamer Regeln und Werkzeuge. Eines der wohl wichtigsten Werkzeuge ist die Wissenschaft. Sie definiert Kriterien, um auf rationaler Ebene über 'Wahrheiten' diskutieren zu können und auf diese Weise Probleme, die das Individuum und das Kollektiv in der Auseinandersetzung mit der Wirklichkeit konfrontieren, zu lösen.

Die Informationsverarbeitung ist schon längst im digitalen Zeitalter angekommen und im gesellschaftlichen, ökonomischen und wissenschaftlichen Miteinander nicht mehr wegzudenken. Information war in der Geschichte der Menschheit immer schon ein wichtiges Gut, hat aber mit der Digitalisierung der Welt noch mehr an Bedeutung gewonnen. Je mehr Datensätze oder Informationsressourcen erzeugt und verteilt werden, desto wichtiger sind geeignete Werkzeuge, um Information wiederzufinden. Information Retrieval Systeme ermöglichen es Informationsressourcen zu verwalten, zugänglich zu machen, zu verteilen und letztendlich angekommene Information zu neuem Wissen zu generieren. In der schier Masse an Dokumenten im Web benötigt es erweiterte Formen des IR, um die Rückgewinnung von Information und Wissen zu ermöglichen. Etwas das 'rückgewonnen' wird, muss nicht noch einmal erzeugt werden.

Das bereits angeführte Zitat *'I don't know what I'm looking for, I'll know it when I find it.'*²⁸⁵ bringt eine Problemstellung auf den Punkt, die für 'klassische' IR-Methoden nur bedingt lösbar ist. In der analogen Welt kann ein kognitiver Agent im Bestand einer Bibliothek stöbern und Informationsressourcen entdecken und einen Bibliothekar persönlich Fragen aller Art zum Bestand stellen. Das Resource Discovery - das digitale Stöbern - im digitalen Raum, stellt Retrieval-Systeme vor neue Herausforderungen. Folgende Punkte, die für die Konzeption und

²⁸⁵JOEMON: Personalisation Techniques in Information Retrieval, S.22.

Umsetzung von RD-Systemen, die im Kontext geisteswissenschaftlicher Daten und Quellen stehen, lassen sich zusammenfassen:

- Interaktive Zugänge seitens der Nutzerinnen ermöglichen es Dokumente, Sammlungsbe-
reiche und ihrer Metadaten besser zu erfassen und entsprechen gegenwärtigen Anforde-
rungen an Informationssysteme. Ein solcher Zugang stellt ein grundlegendes Paradigma
für RD dar.
- Die Möglichkeiten eigenständig Dokumente und Inhalte zu organisieren und zu struktu-
rieren kann ein weiteres Kriterium für erfolgreiches RD sein. Sowohl in der eigenständigen
Organisation von Dokumenten wie in einem 'Datenkorb', als auch die Möglichkeit
Abfragen zu steuern in Form einer facettierten Suche oder in der Form indirekt oder direkt
Feedback zu geben.
- Gerade geisteswissenschaftliche Inhalte können sich sehr stark voneinander unterschei-
den. Unterschiedliche Formen der Informationsvisualisierung, angepasst an die Daten,
kann dabei helfen Dokumente und Sammlungen besser zu verstehen. Aus einem bes-
seren Verständnis heraus können effektiver Anfragen formuliert werden. Es macht Sinn
Briefkorrespondenz als Netzwerk, historische Rechnungsbücher als dynamische digitale
Edition und Karteikarten in ihrer zeitlichen und räumlichen Dimension in Verbindung zu
einem musealischen Objekt zu visualisieren.
- Die semantischen Technologien verfügen über eine hohe Ausdruckskraft und eignen sich
deswegen zur Umsetzung von RD. Nicht das Stöbern nach Text, sondern nach Bedeutung
kann dabei Helfen, Retrieval-Prozesse effektiver zu gestalten.

Im Zuge des Studienprojektes wurden nicht zugängliche geisteswissenschaftliche Ressourcen erschlossen und der Gesellschaft in einem digitalen Repositorium, der GAMS-Infrastruktur, zugänglich gemacht. Diese digital edierten, Karteikarten zu Wildereidelikten wurden standardisiert und normalisiert (TEI) und in einem weiteren Schritt semantisch in RDF formalisiert. Semantic Web Technologien geben geisteswissenschaftlicher Forschung ein Werkzeug in die Hand, das es erlaubt geisteswissenschaftliche Daten, die oft komplexer als naturwissenschaftliche Daten sind, effektiv zu verarbeiten und darzustellen. Die Vorteile digitale Arbeitstechniken in den Geisteswissenschaften, die am konkreten Beispiel der Karteikarten exerziert wurden, umfassen:

- Die (halb)automatisierte Verarbeitung großer Datenmengen ermöglicht es Aussagen über
größere Zeitbereiche, Räume und Quellenmengen tätigen zu können.
- Die Ungebundenheit an Raum und Zeit und die Offenlegung der standardisierten Daten
im Web, liefern transparente und nachvollziehbare Ergebnisse. Jeder Arbeitsschritt, vom
Scan des Originals, bis zum RDF-Datenmodell ist zugänglich und kann geprüft werden.

- Geisteswissenschaftliches Quellenmaterial bedarf oft unterschiedlicher Darstellungsweisen. Die Karteikarten können sowohl in ihrem räumlichen Kontext (Orte von Wilderdelikten), in ihrer statistischen Auswertung, in ihrer Gesamtheit oder in ihrem zeitlichen Kontext interessant erscheinen. Digitale Arbeitstechniken ermöglichen es, aus der gleichen Datenlage, der Forschungsfrage entsprechende, Darstellungsformen im Sinne der Informationsvisualisierung zu generieren.

Digitale Arbeitstechniken und vor allem semantische Technologien werden in den nächsten Jahren verstärkt in - manchmal leider immer noch verstaubter und konservativer - geisteswissenschaftliche Forschung eingesetzt werden. Nicht nur weil sie ein adäquates Werkzeug darstellen, sondern auch weil neue Technologien sich ihren Platz in allen Lebensbereichen verschaffen werden.

Am Ende dieser Arbeit, die hoffentlich zeigt welches Potenzial in Digitalisierung und digitalen Methoden und vor allem dem Information Retrieval und seiner explorativen Umsetzung - dem Resource Discovery - steckt, soll ein Appell stehen, offen und neugierig, aber auch stets kritisch, mit neuen Technologien und ihre Anwendungsgebiete für geisteswissenschaftliche Fragestellungen zu sein.

7 Anhang

7.1 Source Code

Alle im Anhang befindlichen Source-Codes sind auf essentielle Bestandteile reduziert. Header, Namespaces oder einfache HTML-Strukturen wurden aus Gründen der Übersichtlichkeit entfernt, jedoch wurden weitere Kommentare gesetzt. Jeder Source-Code kann online aufgerufen werden.

7.1.1 TEI/XML der Karteikarte o:km.87

Das folgende XML beschreibt eine nach dem TEI-Standard semantisch ausgezeichnete Karteikarte aus dem Kriminalmuseum. Ziel des XML ist es, relevante Information in der Karteikarte zu formalisieren und zu standardisieren. Interessante Entitäten wie Personen, Orte, Gegenstände oder Paragraphen wurden mit entsprechenden TEI-Elementen und Attributen und damit verbunden mit Referenzen, wie etwa Normdaten (z.B. GeoNames) formalisiert. Jede logische Einheit (Gegenstand, Delikt, etc) wird mit einem `<div @type>` strukturiert.

http://gams.uni-graz.at/o:km.87/TEI_SOURCE

```
1 <body>
2 <pb facs="#IMAGE.1"/>
3   <div type="metaheader">
4     <div type="runningID">
5       <p><hi>Z.</hi> 2/37</p>
6     </div>
7     <div type="origLoc">
8       <p><hi>Kasten:</hi> VI</p>
9       <p><hi>Fach:</hi></p>
10    </div>
11  </div>
12  <div type="contentDiv">
13
14    <!-- === Gegenstand: Objekt des Deliktes. Referenz auf Objektthesaurus (oth).
15    term @type="object" um Objekte zu erfassen und term @type="law"
16    für Paragraphen etc.=== -->
17    <div type="subject">
18      <p><hi>Gegenstand:</hi> Wilderer<term ana="oth:18050211" corresp="#OTH.1"
19        type="object"> Kugelstutzen</term> mit 2 auswechsel-<lb/>baren Läufen</p>
20    </div>
21
22    <!-- === Delikt: Referenz auf statische Auflösung der Paragraphen === -->
23    <div type="offence">
```

```

24     <p><hi>Delikt:</hi><term ref="#stgb460" type="law"> Wildddiebstahl</term><lb/></p>
25 </div>
26 </div>
27
28 <pb facs="#IMAGE.2"/>
29 <!-- === Gericht === -->
30 <div type="courtDiv">
31   <p>
32     Bezirks <hi>gericht</hi>
33     <placeName key="#GID.1" ref="http://www.geonames.org/2782070" xml:id="SID.1">
34     Bad Aussee</placeName>
35   </p>
36   <p><hi>G.-Z.</hi> 21410/36</p>
37 </div>
38
39 <!-- === Modellierung personenbezogener Daten: Name, Alter, Beruf, Berufsgruppe,
40 Strafen. @type spezifiziert Elemente weiter: suspect, victim etc. === -->
41 <div type="suspectDiv">
42   <p><hi>Name des Täters:</hi>
43     1.)<persName role="suspect" xml:id="S0007">
44     <forename> Franz</forename><surname> H.</surname></persName>
45     2.)<persName role="suspect" xml:id="S0008">
46     <forename> Emmerich</forename><surname> S.</surname></persName>
47   </p>
48   <p><hi>Alter:</hi>
49     geb.<date corresp="#S0007" type="birthdate" when="1900-12-01">
50     1/12.1900</date> geb. <date corresp="#S0008" type="birthdate"
51     when="1911-10-01"> 7/10.1911</date>
52   </p>
53   <p><hi>Beruf:</hi>
54     <term ana="Besitzerssohn" corresp="#S0007" type="profession">
55     Besitzerssohn</term>
56     <term ana="Landhandwerker" corresp="#S0008" type="profession">
57     Melker</term>
58   </p>
59   <p><hi>Vorstrafen:</hi>
60     <term corresp="#S0007" ref="#unbescholten" type="law">
61     unbescholten</term>
62     <term corresp="#S0008" ref="#unbescholten" type="law">
63     unbescholten</term>
64   </p>
65 </div>
66 <div type="sentenceDiv">
67 <p><hi>Urteil</hi>
68   <add place="above">v.<date type="issuedate" when="1936-12-02">
69   2/12. 1936</date></add>
70   <hi> nach:</hi> 1.)

```

```

71
72 <!-- === Modellierung der Paragraphen und Referenzierung auf statisches TEI
73 zur Auflösung der Paragraphen. Auf Webseite mittels Tooltipp === -->
74 <term corresp="#S0007" ref="#stgb460" type="law"> 460</term> Stg.
75 <term corresp="#S0007" ref="#wp2" type="law"> 2</term>,
76 <term corresp="#S0007" ref="#wp8" type="law"> 8</term>,
77 <term corresp="#S0007" ref="#wp32" type="law"> 32</term>,
78 <term corresp="#S0007" ref="#wp36" type="law"> 36</term> WP 2.)
79 <term corresp="#S0008" ref="#wp2" type="law"> 2</term>,
80 <term corresp="#S0008" ref="#wp8" type="law"> 8</term>,
81 <term corresp="#S0008" ref="#wp32" type="law"> 32</term>,
82 <term corresp="#S0008" ref="#wp36" type="law"> 36</term> WP.<lb/>
83 </p>
84
85 <!-- === term @ref zeigt auf Eintrag im statischen Paragraphen-TEI:
86 http://gams.uni-graz.at/o:km.paragraphen/TEI_SOURCE === -->
87 <p><hi>Strafe:</hi>
88 1. <term corresp="#S0007" ref="arrest" type="punish"> 14 Tage Arrest</term>
89 auf 3 Jahre bedingt Vorhafteten eingerechnet, bei <persName ref="#S0008">
90 2.</persName> daher reduziert. 2.<term corresp="#S0008" ref="arrest" type="punish">
91 12 Stunden</term> <lb/> gemäß <term corresp="#S0007" ref="wp32" type="punish">
92 32</term> <term corresp="#S0008" ref="wp32" type="punish"> WP.
93 Verfall der Waffen und Munition ausgesprochen</term>.
94 </p>
95 </div>
96
97 <!-- === Beschreibungstext des Deliktes. Auftreten der Personen
98 über @corresp und @id referenziert. === -->
99 <div type="crimeDesc">
100 <p><hi>Tatbestand und kriminologisch Relevantes:</hi>
101 <term corresp="#S0007" type="suspect"> Täter 1)</term> hat im
102 <date notAfter="1936-07-31" notBefore="1936-07-01"> Juli 1936</date> im
103 Revier der Bundesforstverwaltung<lb/> Hinterberg unter der Rötelsteinwand
104 gewildert und einen <lb/> Gamsbock i. W. v. 35 S - erlegt u. gestohlen.
105 Bei dieser Gelegenheit<lb/> hat er ohne behördliche Bewilligung
106 zum Tragen verbotener<lb/>Waffen diesen abnehmbaren
107 <term ana="oth:18050211" corresp="#OTH.1" type="object"> Kugelstutzen</term>
108 mit auswechselbaren<lb/>
109 Läufen getragen <term corresp="#S0008" type="suspect"> Täter 2)</term> hat
110 ohne behördliche Bewilligung<lb/> zum Tragen verbotener Waffen sich im
111 Besitz eines Trommel-<lb/>
112 <term ana="oth:18050222" corresp="#OTH.2" type="object">revolvers</term>
113 unter 18cm befunden.<lb/>
114 </p>
115 </div>
116 </body>

```

7.1.2 XSL-Stylesheet zur Transformation der TEI/XML-Karteikarten nach RDF/XML

Diese XSL-Transformation erzeugt beim Ingest von TEI/XML-Dokumenten in die GAMS ein RDF/XML, das die wichtigsten Entitäten des Dokuments (Personen, Orte etc.) weiter formalisiert und mittels URI referenzierbar macht. Diese RDF/XML werden im *tripelstore* angelegt und können anschließend über SPARQL abgefragt werden.

<http://gams.uni-graz.at/archive/objects/cirilo:km/datastreams/TORDF/content>

```
1 <xsl:template match="/">
2   <xsl:variable name="pid" select="//t:publicationStmt/t:idno[@type = 'PID']"/>
3   <rdf:RDF>
4
5   <!-- === Einbindung der SKOS-Referenzen über GETTER-Methode des Datenstroms
6   des GAMS-Objektthesaurus für die in der Karteikarte angeführten Objekte === -->
7   <xsl:for-each select="//t:textClass/t:keywords/t:term">
8     <xsl:variable name="SKOS_URI" select="@ref"/>
9     <xsl:variable name="from" select="(:', ' ', '#')"/>
10    <xsl:variable name="to" select="('%3A', '%2F', '%23')"/>
11    <xsl:variable name="ConvertedSKOS_URI"
12    select="func:replace-multi($SKOS_URI, $from, $to)"/>
13    <xsl:variable name="SKOSConcept">
14      <xsl:copy-of select="document(concat(
15        'http://gams.uni-graz.at/archive/objects/o:oth/methods/sdef:SKOS
16        /getConceptByURI?uri=', $ConvertedSKOS_URI))/*:Concept" />
17    </xsl:variable>
18    <xsl:copy-of select="$SKOSConcept" copy-namespaces="no"/>
19  </xsl:for-each>
20
21  <!-- === Orte: Unterscheidung von eindeutigen Orten der Gerichte (km:Court)
22  und weiteren unspezifischen Orten (km:Location),
23  Verwendung des GeoNames Namespaces (gn:) === -->
24  <xsl:for-each select="//t:body//t:placeName">
25    <xsl:choose>
26      <xsl:when test="ancestor::t:div[@type='courtDiv']">
27        <km:Court rdf:about="{@ref}">
28          <gn:name xml:lang="de">
29            <xsl:value-of select="normalize-space(.)"/>
30          </gn:name>
31        </km:Court>
32      </xsl:when>
33      <xsl:otherwise>
34        <km:Location rdf:about="{@ref}">
35          <gn:name xml:lang="de">
```

```

36         <xsl:value-of select="normalize-space(.)"/>
37     </gn:name>
38 </km:Location>
39 </xsl:otherwise>
40 </xsl:choose>
41 </xsl:for-each>
42
43 <!-- === Datum: Unterscheidung von Ausstellungsdatum der Karteikarte (km:IssueDate)
44 und von anderen weiteren unspezifischen Datumsangaben (km:Date) === -->
45 <xsl:for-each select="//t:body//t:date[not(@type='birthdate')]">
46 <xsl:choose>
47     <xsl:when test="@type='issuedate'">
48         <km:IssueDate rdf:about="{@when}"/>
49     </xsl:when>
50     <xsl:otherwise>
51         <km:Date rdf:about="{@when | @notBefore}"/>
52     </xsl:otherwise>
53 </xsl:choose>
54 </xsl:for-each>
55
56 <!-- === Personen: Verarbeitung der personenbezogenen Daten (Alter, Beruf,
57 Vorstrafen, Urteil etc.) === -->
58 <xsl:for-each select="//t:div[@type='suspectDiv']//t:persName">
59 <xsl:variable name="PersonID" select="@xml:id"/>
60 <km:Person rdf:about="{\$PersonID}">
61     <xsl:variable name="BirthdayofPerson" select=
62         "normalize-space(//t:date[@corresp = concat('#',\$PersonID)]/@when)"/>
63     <xsl:if test="\$BirthdayofPerson">
64         <km:wasBorn rdf:resource="{\$BirthdayofPerson}"/>
65     </xsl:if>
66
67 <!-- === Berufe: Geteilte Modellierung von Berufsgruppen (km:hadProfessionGroup)
68 und Beruf (km:hadProfession) im TEI für die Auswertung nach Berufsgruppen. === -->
69 <km:hadProfessionGroup>
70     <xsl:value-of select="//t:term[@corresp =
71         concat('#', \$PersonID)][@type='profession']/@ana"/>
72 </km:hadProfessionGroup>
73 <km:hadProfession>
74     <xsl:value-of select="normalize-space(//t:term[@corresp=
75         concat('#', \$PersonID)][@type='profession'])"/>
76 </km:hadProfession>
77
78 <!-- === Vorstrafen === -->
79 <xsl:for-each select="//t:div[@type='suspectDiv']//t:term[@corresp=
80     concat('#', \$PersonID)][@type='law']/@ref">
81     <km:hadPoliceRecord rdf:resource="{normalize-space(.)}"/>
82 </xsl:for-each>

```

```

83
84 <!-- === Urteile === -->
85 <xsl:for-each select="//t:div[@type='sentenceDiv']//t:term[@corresp=
86   concat('#', $PersonID)][@type='law']">
87   <km:wasConvicted rdf:resource="{normalize-space(@ref)}"/>
88 </xsl:for-each>
89
90 <!-- === Strafen === -->
91 <xsl:for-each select="//t:div[@type='sentenceDiv']//t:term[@corresp=
92   concat('#', $PersonID)][@type='punish']">
93   <km:wasPunished rdf:resource="{normalize-space(@ref)}"/>
94 </xsl:for-each>
95
96 <!-- === Alter: Wird aus Ausstellungsdatum und Geburtsdatum berechnet, ist in
97 der Form: '23 Jahre' bereits vorhanden oder wird auf Basis anderer
98 Datumsangaben und Geburtsdatum berechnet. === -->
99 <km:hadAge>
100 <xsl:choose>
101   <xsl:when test="//t:term[@type='age']">
102     <xsl:value-of select="//t:term[@type='age'][@corresp=
103       concat('#', $PersonID)]"/></xsl:value-of>
104   </xsl:when>
105   <xsl:otherwise>
106     <xsl:choose>
107       <xsl:when test="//t:date[@type='issuedate']">
108         <xsl:variable name="issuedDateYear" select="number(
109           substring-before(//t:date[@type='issuedate']/@when, '-'))"/>
110         <xsl:variable name="wasBornYear" select="number(
111           substring-before($BirthdayofPerson, '-'))"/>
112         <xsl:value-of select="$issuedDateYear - $wasBornYear"/>
113       </xsl:when>
114       <xsl:otherwise>
115         <xsl:variable name="issuedDateYear" select="number(1940)"/>
116         <xsl:variable name="wasBornYear" select="number(
117           substring-before($BirthdayofPerson, '-'))"/>
118         <xsl:choose>
119           <xsl:when test="$wasBornYear">
120             <xsl:value-of select="$issuedDateYear - $wasBornYear"/>
121           </xsl:when>
122           <xsl:otherwise>
123             <xsl:text>NaN</xsl:text>
124           </xsl:otherwise>
125         </xsl:choose>
126       </xsl:otherwise>
127     </xsl:choose>
128   </xsl:otherwise>
129 </xsl:choose>

```

```

130     </km:hadAge>
131 </km:Person>
132 </xsl:for-each>
133
134 <!-- === Logisches Objekt: Bringt alle Entitäten der Karteikarte zusammen
135 und verbindet sie mit dem PID -->
136 <rdf:Description rdf:about="{
137     concat('http://gams.uni-graz.at/', //t:publicationStmt/t:idno)}">
138     <km:Subject>
139         <xsl:value-of select="normalize-space(
140             substring-after(//t:div[@type='subject']/t:p, 'Gegenstand:'))"/>
141     </km:Subject>
142     <km:Offence>
143         <xsl:value-of select="normalize-space(
144             substring-after(//t:div[@type='offence']/t:p, 'Delikt:'))"/>
145     </km:Offence>
146     <xsl:for-each select="//t:textClass/t:keywords/t:term">
147         <skos:Concept rdf:resource="{@ref}"/>
148     </xsl:for-each>
149     <xsl:for-each select="//t:body//t:placeName">
150         <xsl:choose>
151             <xsl:when test="ancestor::t:div[@type='courtDiv']">
152                 <km:Court rdf:resource="{@ref}"/>
153             </xsl:when>
154             <xsl:otherwise>
155                 <km:Location rdf:resource="{@ref}"/>
156             </xsl:otherwise>
157         </xsl:choose>
158     </xsl:for-each>
159     <xsl:for-each select="//t:body//t:date[not(@type='birthdate')]">
160         <xsl:choose>
161             <xsl:when test="@type='issuedate'">
162                 <km:IssueDate rdf:resource="{@when}"/>
163             </xsl:when>
164             <xsl:otherwise>
165                 <km>Date rdf:resource="{@when | @notBefore}"/>
166             </xsl:otherwise>
167         </xsl:choose>
168     </xsl:for-each>
169     <xsl:for-each select="//t:div[@type='suspectDiv']//t:persName">
170         <xsl:variable name="PersonID" select="@xml:id"/>
171         <km:Person rdf:resource="{PersonID}"/>
172     </xsl:for-each>
173 </rdf:Description>
174 </rdf:RDF>
175 </xsl:template>

```

7.1.3 RDF/XML der Karteikarte o:km.87

Das Ergebnis der Transformation TORDF.xsl ist ein RDF/XML einer Karteikarte. Aus Gründen der Übersicht wird es vereinfacht und ohne SKOS-Referenz dargestellt. Ein Beispiel für die SKOS-Referenz findet sich in Kapitel 5.4.1. Weiters zeigt Abbildung 19 die Visualisierung dieses Graphen, der in RDF modelliert ist.

<http://gams.uni-graz.at/o:km.87/RDF>

```
1 <rdf:RDF>
2
3 <!-- === SKOS-Konzepte (Objektthesaurus) der vorkommenden Objekte === -->
4 <skos:Concept xmlns:skos="http://www.w3.org/2004/02/skos/core#"
5   rdf:about="http://gams.uni-graz.at/skos/scheme/o:oth/#18050211">
6   <skos:broader rdf:resource="http://gams.uni-graz.at/skos/scheme/o:oth/#18050210"/>
7   <skos:exactMatch rdf:resource="http://vocab.getty.edu/aat/300037181"/>
8   <skos:inScheme rdf:resource="http://gams.uni-graz.at/skos/scheme/o:oth/">
9   <skos:prefLabel xml:lang="deu">Gewehr</skos:prefLabel>
10  <skos:prefLabel xml:lang="eng">rifle</skos:prefLabel>
11 </skos:Concept>
12 <skos:Concept xmlns:skos="http://www.w3.org/2004/02/skos/core#"
13   rdf:about="http://gams.uni-graz.at/skos/scheme/o:oth/#18050222">
14   <skos:broader rdf:resource="http://gams.uni-graz.at/skos/scheme/o:oth/#18050220"/>
15   <skos:exactMatch rdf:resource="http://vocab.getty.edu/aat/300037173"/>
16   <skos:inScheme rdf:resource="http://gams.uni-graz.at/skos/scheme/o:oth/">
17   <skos:prefLabel xml:lang="deu">Revolver</skos:prefLabel>
18   <skos:prefLabel xml:lang="eng">revolver</skos:prefLabel>
19 </skos:Concept>
20
21 <!-- === Orte === -->
22 <km:Court rdf:about="http://www.geonames.org/2782070">
23   <gn:name xml:lang="de">Bad Aussee</gn:name>
24 </km:Court>
25
26 <!-- === Datumsangaben === -->
27 <km:IssueDate rdf:about="1936-12-02"/>
28
29 <!-- === Weiter Datumsangaben === -->
30 <km>Date rdf:about="1936-07-01"/>
31
32 <!-- === Personen: Bestehend aus Geburtsdatum, Berufsgruppe, Beruf, Vorstrafen
33 Verurteilung, Strafen (jeweils Verweise auf das statische Paragraphen-TEI mit bsp.
34 '#wp2') und Alter === -->
35 <km:Person rdf:about="S0007">
```

```

36 <km:wasBorn rdf:resource="1900-12-01"/>
37 <km:hadProfessionGroup>Besitzerssohn</km:hadProfessionGroup>
38 <km:hadProfession>Besitzerssohn</km:hadProfession>
39 <km:hadPoliceRecord rdf:resource="#unbescholten"/>
40 <km:wasConvicted rdf:resource="#stgb460"/>
41 <km:wasConvicted rdf:resource="#wp2"/>
42 <km:wasConvicted rdf:resource="#wp8"/>
43 <km:wasConvicted rdf:resource="#wp32"/>
44 <km:wasConvicted rdf:resource="#wp36"/>
45 <km:wasPunished rdf:resource="#arrest"/>
46 <km:wasPunished rdf:resource="#wp32"/>
47 <km:hadAge>36</km:hadAge>
48 </km:Person>
49 <km:Person rdf:about="S0008">
50 <km:wasBorn rdf:resource="1911-10-01"/>
51 <km:hadProfessionGroup>Landhandwerker</km:hadProfessionGroup>
52 <km:hadProfession>Melker</km:hadProfession>
53 <km:hadPoliceRecord rdf:resource="#unbescholten"/>
54 <km:wasConvicted rdf:resource="#wp2"/>
55 <km:wasConvicted rdf:resource="#wp8"/>
56 <km:wasConvicted rdf:resource="#wp32"/>
57 <km:wasConvicted rdf:resource="#wp36"/>
58 <km:wasPunished rdf:resource="#arrest"/>
59 <km:wasPunished rdf:resource="#wp32"/>
60 <km:hadAge>25</km:hadAge>
61 </km:Person>
62
63 <!-- === Logisches Objekt: Zusammenführung der im RDF vorkommenden Entitäten,
64 sowie Text des Gegenstandes (km:Subject) und des Deliktes (km:Offence) === -->
65 <rdf:Description rdf:about="http://gams.uni-graz.at/o:km.87">
66 <km:Subject>Wilderer Kugelstutzen mit 2 auswechsel-baren Läufen</km:Subject>
67 <km:Offence>Wilddiebstahl</km:Offence>
68 <skos:Concept rdf:resource="http://gams.uni-graz.at/skos/scheme/o:oth/#18050211"/>
69 <skos:Concept rdf:resource="http://gams.uni-graz.at/skos/scheme/o:oth/#18050222"/>
70 <km:Court rdf:resource="http://www.geonames.org/2782070"/>
71 <km:IssueDate rdf:resource="1936-12-02"/>
72 <km>Date rdf:resource="1936-07-01"/>
73 <km:Person rdf:resource="S0007"/>
74 <km:Person rdf:resource="S0008"/>
75 </rdf:Description>
76 </rdf:RDF>

```

7.1.4 SPARQL zur Abfrage alle Karteikarten

Diese SPARQL Query läuft über alle RDF-Tripel (alle Aussagen im RDF werden als Tripel in der Datenbank gespeichert), die in der GAMS-Infrastruktur abgelegt wurden. Ausgewählt werden alle RDF-Tripel, die vom Typ Person sind und eine Verbindung zu einer Karteikarte aufweisen (PID). Von diesen Personen werden alle Informationen (Beruf, Alter, Strafe etc.) abgefragt, wenn sie vorhanden sind. Die Datenrückgabe erfolgt in XML über eine GETTER-Methode des Query-Objekts

(<http://gams.uni-graz.at/archive/objects/query:km.wildererkk/methods/sdef:Query/getXML>).

<http://gams.uni-graz.at/archive/objects/query:km.wildererkk/datastreams/QUERY/content>

```
1 PREFIX km:<http://gams.uni-graz.at/km/#>
2 PREFIX oa:<http://www.w3.org/ns/oa#>
3 PREFIX g2o:<http://gams.uni-graz.at/ontology/#>
4 PREFIX skos:<http://gams.uni-graz.at/skos/scheme/o:oth/#>
5 PREFIX tei:<http://www.tei-c.org/ns/1.0>
6
7 SELECT DISTINCT
8   ?indexcard ?relation ?object ?age ?professiongroup
9   ?profession ?born ?PoliceRecord ?wasConvicted ?punish
10 {
11   ?indexcard km:Person ?person;
12   ?relation ?object.
13   OPTIONAL
14   {
15     {?object km:hadAge ?age. }UNION
16     {?object km:hadProfessionGroup ?professiongroup.} UNION
17     {?object km:hadProfession ?profession.} UNION
18     {?object km:wasBorn ?born.} UNION
19     {?object km:hadPoliceRecord ?PoliceRecord.} UNION
20     {?object km:wasConvicted ?wasConvicted.} UNION
21     {?object km:wasPunished ?punish.} UNION
22     {?object km:Subject ?subject.} UNION
23     {?object km:Offence ?offence.}
24   }
25 }
```

7.1.5 XSL-Stylesheet zur Transformation: SPARQL Ergebnis zu HTML Repräsentation des Resource Discovery

Diese Transformation erzeugt aus dem Ergebnis der SPARQL, das als XML zurückgeben wird, eine HTML-Seite. Die RDF-Daten werden in diesem Stylesheet verarbeitet und angezeigt, um die statistische Auswertung, das Browsen durch die Delikte, die Darstellung auf der Karte und auf dem Zeitstrahl umsetzen zu können. Für alle diese Bereiche werden vorhanden und offene Templates verwendet, wie <http://visjs.org/> oder <http://www.chartjs.org/>.

http://gams.uni-graz.at/km/km-query_wildererck.xsl

```
1 <xsl:stylesheet>
2   <!-- [...] -->
3   <xsl:include href="km-static.xsl"/>
4
5   <xsl:variable name="bigdataURL"
6     select="string('http://gams.uni-graz.at:8980/bigdata/')"/>
7   <xsl:variable name="bigdataSPAQLURL"
8     select="string('http://gams.uni-graz.at:8980/bigdata/sparql#')"/>
9
10  <!-- === Einbindung des statischen TEI-Paragraphen === -->
11  <xsl:variable name="paragraphenTEI">
12    <xsl:copy-of select=
13      "document('http://gams.uni-graz.at/o:km.paragraphen/TEI_SOURCE')"/>
14  </xsl:variable>
15
16  <xsl:template name="content">
17
18  <!-- === Einbindung von verwendeten Templates === -->
19  <script src="/km/js/Chart.js"><xsl:text> </xsl:text></script>
20  <script src="/km/js/vis.min.js"><xsl:text> </xsl:text></script>
21  <link href="/km/css/vis.min.css" rel="stylesheet" type="text/css"/>
22
23  <!-- === Definition der Altersbereiche als Variabel. === -->
24  <xsl:variable name="AgePath" select="
25    //s:result[s:relation/@uri = 'http://gams.uni-graz.at/km/#Person']/s:age"/>
26  <xsl:variable name="Age15-19" select="$AgePath[number() > 14][number() &lt; 20]"/>
27  <xsl:variable name="Age20-24" select="$AgePath[number() > 19][number() &lt; 25]"/>
28  <xsl:variable name="Age25-29" select="$AgePath[number() > 24][number() &lt; 30]"/>
29  <xsl:variable name="Age30-34" select="$AgePath[number() > 29][number() &lt; 35]"/>
30  <xsl:variable name="Age35-39" select="$AgePath[number() > 34][number() &lt; 40]"/>
31  <xsl:variable name="Age40-44" select="$AgePath[number() > 39][number() &lt; 45]"/>
32  <xsl:variable name="Age45-49" select="$AgePath[number() > 44][number() &lt; 50]"/>
33  <xsl:variable name="Age50-54" select="$AgePath[number() > 49][number() &lt; 55]"/>
```

```

34 <xsl:variable name="Age55-59" select="$AgePath[number() > 54][number() &lt; 60]"/>
35 <xsl:variable name="Age60-64" select="$AgePath[number() > 59][number() &lt; 65]"/>
36 <xsl:variable name="Age65-69" select="$AgePath[number() > 64][number() &lt; 70]"/>
37 <xsl:variable name="Age70-74" select="$AgePath[number() > 69][number() &lt; 75]"/>
38
39 <!-- === Definition der Berufsgruppen als Variabel=== -->
40 <xsl:variable name="ProfessionGroupPath" select=
41 "//s:result[s:relation/@uri='http://gams.uni-graz.at/km/#Person']/s:professiongroup"/>
42 <xsl:variable name="HofbesitzendeBauern"
43   select="$ProfessionGroupPath[matches(.,'Besitzer')]" />
44 <xsl:variable name="Bauernsohn" select="$ProfessionGroupPath[matches(.,'s[o,ö]hn')]" />
45 <xsl:variable name="Bauernknechte" select="$ProfessionGroupPath[matches(.,'Knecht')]" />
46 <xsl:variable name="Besitzlose" select="$ProfessionGroupPath[matches(.,'[H,h]ilf')]" />
47 <xsl:variable name="Handwerker" select="$ProfessionGroupPath[matches(.,'[H,h]andw')]" />
48 <xsl:variable name="Arbeiter"
49   select="$ProfessionGroupPath[matches(.,'[A,a]rbeiter')]" />
50 <xsl:variable name="Sonstige" select="$ProfessionGroupPath[matches(.,'Sonstige')]" />
51
52 <section class="row">
53   <article class="col-sm-12">
54     <div class="panel">
55       <h3>Stöbern im Bestand der Wilderererkarteikarten</h3>
56       <!-- [...] -->
57       <div class="row">
58         <div class="col-sm-6">
59           <h5>Berufsgruppen im Pfliegericht Golling nach SCHINDLER</h5>
60           <canvas height="400" id="professionChartGOLLING" width="400">
61             <xsl:text> </xsl:text></canvas>
62
63 <!-- === Erzeugung der Tabelle für Berufsgruppen === -->
64 <script>
65   var ctx = document.getElementById('professionChartGOLLING');
66   var professionChartGOLLING = new Chart(ctx, {type: 'bar',
67   data: {
68     labels: ['Hofbesitzende Bauern','Bauernsöhne','Bauernknechte',
69     'besitzlose Unterschicht','Landhandwerker','Jäger/Arbeiter','Sonstige'],
70     datasets: [label: 'Schindler',
71       data:
72         [24,23,17,13,13,6,14],
73       backgroundColor: ['rgba(255, 99, 132, 0.2)',
74       'rgba(54, 162, 235, 0.2)',
75       'rgba(255, 206, 86, 0.2)',
76       'rgba(75, 192, 192, 0.2)',
77       'rgba(153, 102, 255, 0.2)',
78       'rgba(100, 55, 80, 0.2)',
79       'rgba(255, 159, 64, 0.2)'
80     ],

```

```

81     borderColor: [
82       'rgba(255,99,132,1)',
83       'rgba(54, 162, 235, 1)',
84       'rgba(255, 206, 86, 1)',
85       'rgba(75, 192, 192, 1)',
86       'rgba(153, 102, 255, 1)',
87       'rgba(110, 65, 90, 1)',
88       'rgba(255, 159, 64, 1)'
89     ],
90     borderWidth: 1
91   },{
92     label: 'Kriminalmuseum',
93
94     <!-- === Einführung der Berufsgruppen-Variablen === -->
95     data:
96     [<xsl:value-of select="count($HofbesitzendeBauern)" />,
97      <xsl:value-of select="count($Bauernsohn)" />,
98      <xsl:value-of select="count($Bauernknechte)" />,
99      <xsl:value-of select="count($Besitzlose)" />,
100     <xsl:value-of select="count($Handwerker)" />,
101     <xsl:value-of select="count($Arbeiter)" />,
102     <xsl:value-of select="count($Sonstige)" />],
103     backgroundColor: [
104       'rgba(255, 99, 132, 0.4)',
105       'rgba(54, 162, 235, 0.4)',
106       'rgba(255, 206, 86, 0.4)',
107       'rgba(75, 192, 192, 0.4)',
108       'rgba(153, 102, 255, 0.4)',
109       'rgba(100, 55, 80, 0.4)',
110       'rgba(255, 159, 64, 0.4)'
111     ],
112     borderColor: [
113       'rgba(255,99,132,1)',
114       'rgba(54, 162, 235, 1)',
115       'rgba(255, 206, 86, 1)',
116       'rgba(75, 192, 192, 1)',
117       'rgba(153, 102, 255, 1)',
118       'rgba(100, 55, 80, 1)',
119       'rgba(255, 159, 64, 1)'
120     ],
121     borderWidth: 1
122   } ] },options:{scales:{yAxes:[{ticks: {min: 0,max:40}}]}}});
123 </script>
124 </div>
125 <div class="col-sm-6">
126   <h5>Altersverteilung im Pfliegericht Golling nach SCHINDLER</h5>
127   <canvas height="400" id="myChartGOLLING" width="400"> </canvas>

```

```

128
129     <!-- === Erzeugung der Tabelle für Altersstruktur, Variablen für Alter werden hier
130 eingesetzt=== -->
131 <script>
132     var ctx = document.getElementById('myChartGOLLING');
133     var myChartGOLLING = new Chart(ctx, {type: 'line',
134     data:{
135     labels: [ '15-19','20-24', '25-29', '30-34', '35-39',
136     '40-44', '45-49','50-54','55-59', '60-64','65-69'],
137     datasets:[{label: 'Schindler',
138     data: [10,30,18,13,7,9,5,5,2,1,2, ],
139     backgroundColor: ['rgba(255, 159, 64, 0.2)'],
140     borderColor: ['rgba(255, 159, 64, 1)'],
141     borderWidth: 1
142     },{label: 'Kriminalmuseum',
143
144     <!-- === Einführung der Altersbereich-Variablen === -->
145     data:[
146     <xsl:value-of select="count($Age15-19)"/>,
147     <xsl:value-of select="count($Age20-24)"/>,
148     <xsl:value-of select="count($Age25-29)"/>,
149     <xsl:value-of select="count($Age30-34)"/>,
150     <xsl:value-of select="count($Age35-39)"/>,
151     <xsl:value-of select="count($Age40-44)"/>,
152     <xsl:value-of select="count($Age45-49)"/>,
153     <xsl:value-of select="count($Age50-54)"/>,
154     <xsl:value-of select="count($Age55-59)"/>,
155     <xsl:value-of select="count($Age60-64)"/>,
156     <xsl:value-of select="count($Age65-69)"/>
157     ],
158     backgroundColor: ['rgba(54, 162, 235, 0.2)'],
159     borderColor: ['rgba(54, 162, 255, 1)'],
160     borderWidth: 1 }],},
161     options: {scales: {yAxes: [{ticks: { beginAtZero:true}}]}]);
162 </script>
163 </div>
164 </div>
165 <hr/>
166
167 <!-- === Erzeugung des Browsings durch Karteikarten === -->
168 <h4 id="liste">Liste der Delikte</h4>
169 <div style=" height: 450px; overflow: scroll;">
170 <xsl:for-each-group group-by="s:indexcard/@uri" select="//s:results/s:result">
171 <xsl:sort select="current-grouping-key()"/>
172 <xsl:variable name="URI" select="current-grouping-key()"/>
173 <xsl:variable name="allProperties"
174     select="//s:results/s:result[s:indexcard/@uri = $URI]"/>

```

```

175 <div class="col-sm-4 well" style="height: 200px; padding-top: 0px;">
176 <h5>
177 <b>Gegenstand:</b><br/>
178 <a data-placement="bottom" data-toggle="tooltip" href="{URI}"
179 target="_blank" title="{concat('Zur Karteikarte ',
180 substring-after($URI,'http://gams.uni-graz.at/'))}">
181 <xsl:value-of select="$allProperties[s:relation/@uri
182 = 'http://gams.uni-graz.at/km/#Subject']"/>
183 </a>
184 </h5>
185 <p>
186 <b>Delikt:</b><br/>
187 <xsl:value-of
188 select="$allProperties[s:relation/@uri =
189 'http://gams.uni-graz.at/km/#Offence']"/><br/>
190 <b>Täter:</b><br/>
191 <xsl:for-each-group group-by="s:profession"
192 select="$allProperties[s:relation/@uri =
193 'http://gams.uni-graz.at/km/#Person']">
194 <xsl:if test="not(current-grouping-key() = '')">
195 <xsl:value-of select="current-grouping-key()"/>; </xsl:if>
196 </xsl:for-each-group>
197 </p>
198 </div>
199 </xsl:for-each-group>
200 </div>
201 <hr/>
202
203 <!-- ===Einbindung des GeoBrowsers + Zugriff auf KML-Datenstrom in GAMS === -->
204 <h4 id="orte">Orte der zuständigen Gerichte </h4>
205 <iframe id="geobrowser" src="http://geobrowser.de.dariah.eu/embed/?kml=
206 http://gams.uni-graz.at/archive/get/context:km.karteikarten/KML"> </iframe><hr/>
207 <div>
208
209 <!-- === Aufruf zur Erzeugung der Zeitleiste === -->
210 <h4 id="datum">Nach Ausstellungsdatum</h4>
211 <xsl:call-template name="CreateTimelineIssued"/>
212 </div>
213 </article>
214 </section>
215 </xsl:template>
216
217 <!-- === Erzeugung der Zeitleiste mittels Template === -->
218 <xsl:template name="CreateTimelineIssued">
219 <div id="visualizationIssued"><xsl:text> </xsl:text></div>
220 <script type="text/javascript">
221 var container = document.getElementById('visualizationIssued');

```

```

222     var items = new vis.DataSet([
223 <xsl:for-each select="//s:results/s:result[child::s:relation[@uri =
224   'http://gams.uni-graz.at/km/#IssueDate']]">
225   <xsl:text>{id:</xsl:text>
226   <xsl:value-of select="position()"/>
227   <xsl:text> , content: '</xsl:text>
228   <xsl:variable name="helpTimeline" select="s:indexcard/@uri"/>
229   <a target="_blank" href="{ $helpTimeline}">
230     <xsl:variable name="actualResults" select="//s:result[s:indexcard/@uri
231     = $helpTimeline]"/>
232     <xsl:value-of select="$actualResults/s:object"/>
233   </a>
234   <xsl:text>' , start: '</xsl:text>
235   <xsl:value-of select="substring-after(s:object/@uri,
236   'http://gams.uni-graz.at:8980/bigdata/')"/>
237   <xsl:text>'} ,</xsl:text>
238 </xsl:for-each>]);
239   var options = {};
240   var timeline = new vis.Timeline(container, items, options);
241   document.getElementById('fit').onclick = function()
242     timeline.fit();};
243 </script>
244 </xsl:template>
245 <xsl:template name="processedOutput">
246 <xsl:variable name="lawID" select="substring-after(., '#')"/>
247 <xsl:variable name="actual"
248   select="$paragraphenTEI//t:taxonomy//t:category[@xml:id = $lawID]"/>
249 <a data-placement="bottom" data-toggle="tooltip">
250   <xsl:attribute name="data-original-title" select="$actual//t:gloss"/>
251   <xsl:choose>
252     <xsl:when test="contains(., 'sparql#')">
253       <xsl:value-of select="substring-after(., $bigdataSPAURL)"/>
254     </xsl:when>
255     <xsl:otherwise>
256       <xsl:value-of select="substring-after(., $bigdataURL)"/>
257     </xsl:otherwise>
258   </xsl:choose>
259 </a>
260 <xsl:text>; </xsl:text>
261 </xsl:template>
262 </xsl:stylesheet>

```

7.2 Abkürzungsverzeichnis

CIDOC-CRM Comité international pour la documentation - Conceptual Reference Model

CSS Cascading Style Sheets

DC Dublin Core

Fedora Flexible Extensible Digital Object Repository Architecture

GAMS Geisteswissenschaftliches Asset Management System

HTML Hypertext Markup Language

HTTP Hypertext Transfer Protocol

IR Information Retrieval

KML Keyhole Markup Language

LIDO Lightweight Information Describing Objects

PID Persistent Identifier

RDF Resource Description Framework

RD Resource Discovery

SKOS Simple Knowledge Organization System

SPARQL SPARQL Protocol And RDF Query Language

TEI Text Encoding Initiative

URI Uniform Resource Identifiers

URL Uniform Resource Locator

URN Uniform Resource Name

W3C World Wide Web Consortium

XML Extensible Markup Language

XSL(-T) Extensible Stylesheet Language – Transformations

ZIM Zentrum für Informationsmodellierung in den Geisteswissenschaften

7.3 Onlineressourcen

Accessibility, <https://en.wikipedia.org/wiki/Accessibility>, 12.02.2017.

Bitstrom, <https://en.wikipedia.org/wiki/Bitstream>, 02.03.2017.

Blazegraph, <https://www.blazegraph.com>, 31.12.2016.

CSS, https://de.wikipedia.org/wiki/Cascading_Style_Sheets, 04.03.2017.

Dublin Core, <http://dublincore.org>, 11.12.2016.

Facettierte Suche, <http://gams.uni-graz.at/context:pardus>, 05.02.2017.

FEDORA, <http://fedorarepository.org/about>, 05.02.2016.

GAMS, gams.uni-graz.at, 05.02.2016.

GAMS Dokumentation, <http://gams.uni-graz.at/docs>, 05.02.2016.

Geo-Browser, <https://de.dariah.eu/geobrowser>, <https://de.dariah.eu/geobrowser>, 11.12.2016.

Google Scholar, https://de.wikipedia.org/wiki/Google_Scholar, 17.02.2017.

Gut, Wirtschaftswissenschaft,
[https://de.wikipedia.org/wiki/Gut_\(Wirtschaftswissenschaft\)](https://de.wikipedia.org/wiki/Gut_(Wirtschaftswissenschaft)), 20.09.2016.

Interoperabilität,
<https://www.w3.org/blog/2008/05/open-standards-interoperability>, 12.04.2017.

Infiniter Regress, https://de.wikipedia.org/wiki/Infiniter_Regress, 25.01.2017.

Informationsökonomik,
<http://wirtschaftslexikon.gabler.de/Archiv/54725/informationsoekonomik-v1.html>, 09.01.2017.

KML, <https://developers.google.com/kml/documentation>, 31.12.2016.

LIDO, www.lido-schema.org, 05.02.2017.

Lucene, <http://lucene.apache.org/core>, 02.07.2016.

Mulgara, <http://www.mulgara.org>, 31.12.2016.

Ontology, <https://www.w3.org/standards/semanticweb/ontology>, 05.02.2016.

Paragrafen, gams.uni-graz.at/o:km.paragrafen/TEI_SOURCE, 30.12.2016.

Radikaler Konstruktivismus,
https://de.wikipedia.org/wiki/Radikaler_Konstruktivismus, 04.11.2016.

Responsiveness, https://de.wikipedia.org/wiki/Responsive_Webdesign, 12.02.2017.

RDF, <https://www.w3.org/RDF>, 20.12.2016.

Sesame, <http://rdf4j.org/>, 05.02.2017.

SKOS, <https://www.w3.org/2004/02/skos>, 05.02.2017.

SPARQL, <https://www.w3.org/TR/rdf-sparql-query>, 05.02.2016.

STEPHAN Felix, Das Wort des Jahres ist falsch", <http://www.zeit.de/kultur/2016-12/postfaktisch-wort-des-jahres-post-truth-demokratie-jill-lepore>, 22.12.2016.

Social Web, https://de.onpage.org/wiki/Social_Web, 14.01.2017.

StoryMap.js, <https://storymap.knightlab.com>, 05.02.2016.

TEI, <http://www.tei-c.org/index.xml>, 05.02.2016.

Trello Web-basierte Projektmanagementsoftware, <https://trello.com>, 09.12.2016.

Usability, <https://en.wikipedia.org/wiki/Usability>, 12.02.2017.

URI, https://de.wikipedia.org/wiki/Uniform_Resource_Identifier, 04-03.2017.

Virtuelle Ausstellung als thematischer Discovery-Einsprungspunkt,
<http://gams.uni-graz.at/o:km.story>, 12.04.2017.

Volltextsuche mit blazegraph,
<https://wiki.blazegraph.com/wiki/index.php/FullTextSearch>, 31.12.2016.

W3C, <https://www.w3.org/>, 05.03.2017.

Webpräsenz Kriminalmuseum, <https://kriminalmuseum.uni-graz.at>, 02.07.2016.

XML, <https://www.w3.org/XML>, 27.12.2016.

XSL-T, <https://www.w3.org/TR/xslt>, 05.02.2016.

7.4 Projektspezifische Onlineressourcen

Digitale Sammlung des Kriminalmuseums,

<http://gams.uni-graz.at/context:km>, 16.12.2016.

Digitale Repräsentation einer Karteikarte,

<http://gams.uni-graz.at/o:km.87>, 16.12.2016.

DC Metadatensatz einer Karteikarte,

<http://gams.uni-graz.at/o:km.87/DC>, 12.02.2016.

Materialthesaurus, SKOS,

gams.uni-graz.at/archive/objects/o:mth/datastreams/ONTOLOGY/content, 02.07.2016.

Objektthesaurus, SKOS,

gams.uni-graz.at/archive/objects/o:oth/datastreams/ONTOLOGY/content, 02.07.2016.

RDF einer Karteikarte,

<http://gams.uni-graz.at/o:km.87/RDF>, 16.12.2016.

RD im Bestand der Wildererkarteikarten,

<http://gams.uni-graz.at/archive/objects/query:km.wildererkk/methods/sdef:Query/get?mode=query>, 16.12.2016.

TEI einer Karteikarte,

http://gams.uni-graz.at/o:km.87/TEI_SOURCE, 16.12.2016.

Trelloboard des Studienprojektes,

<https://trello.com/b/bgW70TtY/kriminalmuseum>, 05.02.2017.

XML Query-Ergebnis für RD,

<http://gams.uni-graz.at/archive/objects/query:km.wildererkk/methods/sdef:Query/get?mode=query>, **05.02.2016.**

Literatur

- BACHHIESL, Christian:** Räuber, Mörder, Sittenstrolche. 37 Fälle aus dem Kriminalmuseum der Karl-Franzens-Universität Graz. 2. Auflage. Graz, 2004
- BAEZA-YATES, Ricardo/RIBEIRO-NETO, Berthier et al.:** Modern information retrieval. The concept and technology behind search. 2. Auflage. 2011
- BALCERIS, Michael:** Medien-und Informationskompetenz: Modellierung und Messung von Informationskompetenz bei Schülern. 2011
- BALLOD, Matthias:** Informationen und Wissen im Griff: effektiv informieren und effizient kommunizieren. W. Bertelsmann Verlag, 2011
- BAUER, Reinhard:** Die digitale Bibliothek von Babel: über den Umgang mit Wissensressourcen im Web 2.0. Hülsbusch, 2010
- BAUERNHANSL, Thomas/TEN HOMPEL, Michael/VOGEL-HEUSER, Birgit:** Industrie 4.0 in Produktion, Automatisierung und Logistik: Anwendung· Technologien· Migration. Wiesbaden, 2014
- BECKER, Konrad/STALDER, Felix:** DEEP SEARCH: Politik des Suchens jenseits von Google. Bundeszentrale für politische Bildung, 2010
- BERNERS-LEE, T./FISCHETTI, M.:** Weaving the Web: The Past, Present and Future of the World Wide Web by Its Inventor. London, 2000
- BIRKNER, Michael et al.:** Guideline zur Langzeitarchivierung. 2016
- BRÜBACH, Nils:** Das Referenzmodell OAIS. In **Neuroth, Heike et al. (Hrsg.):** nestor-Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Göttingen, 2010 (URL: http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_438.pdf)
- BURKART, Margarete:** Thesaurus. In **Kuhlen, Rainer/Seeger, Thomas/Strauch, Dietmar (Hrsg.):** Grundlagen der praktischen Information und Dokumentation: Handbuch zur Einführung in die Informationswissenschaft und -praxis. 5. Auflage. München, 2004, S.141–154
- BÜTTCHER, Stefan/CLARKE, Charles LA/CORMACK, Gordon V:** Information retrieval: Implementing and evaluating search engines. Cambridge and London, 2016
- CARD, S.K./MACKINLAY, J.D./SHNEIDERMAN, B.:** Readings in Information Visualization: Using Vision to Think. San Francisco, 1999

- CARLIN, Sascha:** Schlagwortvergabe durch Nutzende (Tagging) als Hilfsmittel zur Suche im Web. Darmstadt, 2006, Dipl.-Arb.
- CCSDS:** Reference Model for an Open Archival Information System (OAIS), Recommended Practice. CCSDS 650.0-M-2 (Magenta Book) Issue 2, 2012 – Technischer Bericht
- DAVIS, John/DUKE, Alistair/KIRYAKOV, Atanas:** Semantic Search. In **Goker, Ayse/Davies, John (Hrsg.):** Information retrieval: searching in the 21st century. Chichester, 2009, S.179–209
- DE VIRGILIO, Roberto/GUERRA, Francesco/VELEGRAKIS, Yannis:** Semantic search over the web. Springer Science & Business Media, 2012
- DEERWESTER, Scott et al.:** Indexing by latent semantic analysis. Journal of the American society for information science, 41 1990, Nr. 6, 391
- DENGEL, Andreas:** Semantische Technologien: Grundlagen - Konzepte - Anwendungen. Heidelberg, 2012
- DÖRNER, Dietrich:** Vom Umgang mit Unbestimmtheit und Komplexität. Bern, 1983
- DUCHARME, Bob:** Learning SPARQL. Sebastopol, 2011
- EWERT, Gisela/UMSTÄTTER, Walther:** Lehrbuch der Bibliotheksverwaltung. Stuttgart, 1997
- FAVRE-BULLE, Bernard:** Information und Zusammenhang: Informationsfluß in Prozessen der Wahrnehmung, des Denkens und der Kommunikation. Springer, 2001
- FORTMÜLLER, Walter:** Informationsvisualisierung: Analyse und Evaluierung ausgewählter Applikationen zur interaktiven visuellen Exploration von Datenbeständen. Graz, 2011, MA-Arb.
- FUHR, Norbert:** Modelle im Information Retrieval. In **Kuhlen, Rainer/Semar, Wolfgang/Strauch, Dietmar (Hrsg.):** Grundlagen der praktischen Information und Dokumentation: Handbuch zur Einführung in die Informationswissenschaft und -praxis. 6. Auflage. Berlin and Boston, 2013, S.321–334
- FUNK, Stefan E.:** Digitale Objekte und Formate. In **Neuroth, Heike et al. (Hrsg.):** nestor-Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Göttingen, 2010 \langle URL: http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_445.pdf \rangle
- GAUS, Wilhelm:** Dokumentations- und Ordnungslehre: Theorie und Praxis des Information Retrieval. Springer-Verlag, 2006

- GESER, Guntram:** Resource Discovery - Position Paper: Putting the Users First. In **Geser, Guntram/Pereira, John (Hrsg.):** DigiCULT Thematic Issue 6 – Resource Discovery Technologies for the Heritage Sector. Salzburg and Glasgow, 2004, S.7–12
- GIARETTA, David:** Advanced Digital Preservation. Berlin and Heidelberg, 2011
- GOKER, Ayse/DAVIES, John:** Information retrieval: searching in the 21st century. Chichester, 2009
- GRAUPMANN, Jens:** Semantische Suche: intelligente Suchmaschinen durch innovative und zukunftsweisende Konzepte und Technologien. Saarbrücken, 2007
- GUMM, Heinz Peter/SOMMER, Manfred:** Einführung in die Informatik. 9. Auflage. Oldenbourg Verlag, 2011
- HEARST, Marti:** Search User Interfaces. Cambridge, 2009
- HEIDEGGER, Martin:** Sein und Zeit. 19. Auflage. Tübingen, 2006
- HEINRICH, Lutz J/RIEDL, René/STELZER, Dirk:** Informationsmanagement: Grundlagen, Aufgaben, Methoden. Oldenbourg, 2014
- HERMANS, Jan:** Ontologiebasiertes Information Retrieval für das Wissensmanagement. Band 39, Berlin, 2008
- HILDEBRAND, Michiel/OSSENBRUGGEN, Jacco Ronald van/HARDMAN, Lynda:** An analysis of search-based user interaction on the semantic web. 2007
- HORT, I:** Holger Nohr: Grundlagen der automatischen Indexierung: ein Lehrbuch. ABI TECHNIK, 24 2004, Nr. 1, 74–75
- JÄGER-DENGLER-HARLES, Ingeborg:** Informationsvisualisierung und Retrieval im Fokus der Informationspraxis. 2013, MA-Arb.
- JANNACH, Dietmar et al.:** Recommender systems: an introduction. Cambridge, 2010
- JANSEN, B.J./SPINK, A./PEDERSEN, J.O.:** A Temporal Comparison Of Altavista Web Searching. Journal Of The American Society For Information Science And Technology, 56(6), 2005, S.559–570
- JEHN, Mathias:** Metadatenstandards im Bereich der digitalen LZA. In **Neuroth, Heike et al. (Hrsg.):** nestor-Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Göttingen, 2010 (URL: http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_476.pdf)

- JOEMON, Jose:** Personalisation Techniques in Information Retrieval. In **Geser, Guntram/Pereira, John (Hrsg.):** DigiCULT Thematic Issue 6 – Resource Discovery Technologies for the Heritage Sector. Salzburg and Glasgow, 2004, S.22–29
- JORDANOUS, Anna:** Enhancing information retrieval and resource discovery from data using the Semantic Web. In Emerging Trends and Technologies in Libraries and Information Services (ETTLIS), 2015 4th International Symposium on. IEEE 2015, 105–110
- KÄKI, Mika:** Findex: search result categories help users when document ranking fails. In Proceedings of the SIGCHI conference on Human factors in computing systems. 2005, S.131–140
- KARGL, Carina:** Der Geschichte ein Gesicht geben - Die Modellierung und Aufbereitung der 'Stamm- und Wappenbücher' der Stadt Regensburg. Graz, 2012, MA-Arb.
- KARREBERG, Ulrich:** Signale-Prozesse-Systeme: eine multimediale und interaktive Einführung in die Signalverarbeitung. Springer-Verlag, 2016
- KEITEL, Christian; Schoger, Astrid (Hrsg.):** Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644. Berlin and Wien and Zürich, 2013
- KENDAL, Simon/CREEN, Malcolm:** An Introduction to Knowledge Engineering. London, 2007
- KOCHER, Gernot:** Das K.k kriminalistische Universitätsinstitut in Graz. In **Bachhiesl, Christian/Bachhiesl, Sonja Maria/Leitner, Johann (Hrsg.):** Kriminologische Entwicklungslinien. Eine interdisziplinäre Synopsis. Wien and Berlin, 2014, S.21–34
- KUHLEN, Rainer:** Erfolgreiches Scheitern — eine Götterdämmerung des Urheberrechts? In Schriften zur Informationswissenschaft Band 48 Herausgegeben vom Hochschulverband für Informationswissenschaft (HI) eV Konstanz.
- KUHLEN, Rainer:** Wissensökonomie und Wissensökologie zusammen denken. In **Helfrich, Silke (Hrsg.):** Commons: für eine neue Politik jenseits von Markt und Staat. Bielefeld, 2012, S.405–413
- KUHLEN, Rainer:** Information - Informationswissenschaft. In **Kuhlen, Rainer/Semar, Wolfgang/Strauch, Dietmar (Hrsg.):** Grundlagen der praktischen Information und Dokumentation: Handbuch zur Einführung in die Informationswissenschaft und -praxis. 6. Auflage. Berlin and Bosten, 2013, S.1–19
- KUSTER, Jürg et al.:** Handbuch Projektmanagement. Berlin and Heidelberg, 2011
- LEWANDOWSKI, Dirk:** Informationskompetenz und das Potenzial der Internetsuchmaschinen. In **Sühl-Strohmeier, W. (Hrsg.):** Handbuch Informationskompetenz. Berlin, 2012

- LEWANDOWSKI, Dirk:** Suchmaschinen verstehen. Berlin and Heidelberg, 2015
- LIEGMANN, Hans/NEUROTH, Heike:** Einführung. In **Neuroth, Heike et al. (Hrsg.):** nestor-Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Göttingen, 2010 (URL: http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_390.pdf)
- LINCKELS, Serge/MEINEL, Christoph:** E-librarian service: user-friendly semantic search in digital libraries. Springer Science & Business Media, 2011
- MANECKE, Hans-Jürgen:** Klassifikation, Klassieren. In **Kuhlen, Rainer/Seeger, Thomas/Strauch, Dietmar (Hrsg.):** Grundlagen der praktischen Information und Dokumentation: Handbuch zur Einführung in die Informationswissenschaft und -praxis. 5. Auflage. München, 2004, S.127–140
- MASON, Paul:** Postkapitalismus: Grundrisse einer kommenden Ökonomie. Berlin, 2016
- NIEDERMAIR, Klaus:** Gefährden Suchmaschinen und Discovery-Systeme die informationelle Autonomie? Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare, 67 2014, Nr. 1, 109–125
- NOHR, Holger:** Grundlagen der automatischen Indexierung: ein Lehrbuch. Berlin, 2003
- NORMAN, Donald A:** The design of everyday things: Revised and expanded edition. New York, 2013
- PILLER, Frank Thomas:** Einführung: Informationsrevolution und industrielle Produktion. In Mass Customization. Wiesbaden, 2000, S.1–9
- POETZSCH, Eleonore:** Information Retrieval: Einführung in Grundlagen und Methoden. Berlin-Brandenburg, 2005
- POPPER, Karl R:** Auf der Suche nach einer besseren Welt: Vorträge und Aufsätze aus dreißig Jahren. München, 1984
- REGINALD, Ferber:** Information Retrieval-Suchmodelle und Data-Mining-Verfahren für Textsammlungen und das Web. Heidelberg, 2003
- RÜGER, Stefan:** Multimedia Resource Discovery. In **Goker, Ayse/Davies, John (Hrsg.):** Information retrieval: searching in the 21st century. Chichester, 2009, S.39–62
- SALAMUN, Kurt:** Was ist Philosophie? Tübingen, 2009
- SANDKÜHLER, Hans Jörg et al.:** Europäische Enzyklopädie zu Philosophie und Wissenschaften. 1994

- SCHINDLER, Norbert:** Wilderer im Zeitalter der Französischen Revolution: ein Kapitel alpiner Sozialgeschichte. München, 2001
- SHANNON, C. E.:** A mathematical theory of communication. The Bell System Technical Journal, Vol. 27, 1948, S.379–423, 623–656
- SHNEIDERMAN, Ben:** Designing the user interface: strategies for effective human-computer interaction. Pearson Education India, 2010
- SPINK, Amanda/ZIMMER, Michael:** Web search: Multidisciplinary perspectives. Band 14, Springer Science & Business Media, 2008
- STECHOW, Arnim von:** Schritte zur Satzsemantik. 2004
- STEINER, Elisabeth:** Kulturelles Erbe virtuell repräsentiert: Überlegungen zur Konzeption eines Digitalen Archivs der Stamm- und Wappenbücher der Stadt Regensburg. Graz, 2012, MA-Arb.
- STELZER, Dirk:** Informationsbedarf. In **Mertens, P. (Hrsg.):** Lexikon der Wirtschaftsinformatik. 4. Auflage. Berlin, 2004, S.238–239
- STIGLER, Johannes Hubert:** Neue Wege in der Digitalen Edition. Jenseits von Hypertext und Nicht-Linearität. In Beihefte zu editio 30. Graz, 2009
- STOCK, Wolfgang G.:** Information retrieval: Informationen suchen und finden. Band 1, Oldenbourg Verlag, 2007
- STOCK, Wolfgang G./STOCK, Mechtild:** Wissensrepräsentation: Informationen auswerten und bereitstellen. Band 2, Oldenbourg Verlag, 2008
- SUTCLIFFE, Alistair/ENNIS, Mark:** Towards a cognitive theory of information retrieval. Interacting with computers, 10 1998, Nr. 3, S.321–351
- THALLER, Manfred:** What is a text within the Digital Humanities, or some of them, at least? dh2012 - Book of Abstracts, 2012 (URL: <http://www.dh2012.uni-hamburg.de/conference/programme/abstracts/beyond-embedded-markup>)
- TUFTE, Edward:** Envisioning Information. Cheshire, 2008
- VIPUL, Kashyap/BUSSLER, Christoph/MORAN, Matthew:** The Semantic Web-Semantics for Data and Services on the Web. Berlin and Heidelberg, 2008
- VOGELER, Georg et al.:** The Content of Accounts and Registers in their Digital Edition. XML/TEI, Spreadsheets, and Semantic Web Technologies. In **Sarnowsky, Jürgen (Hrsg.):** Konzeptionelle Überlegungen zur Edition von Rechnungen und Amtsbüchern des späten Mittelalters. Göttingen, 2016, S.13–41

- VON GLASERSFELD, Ernst:** Konstruktion der Wirklichkeit und der Begriff der Objektivität.
In **Von Foerster, Heinz (Hrsg.):** Einführung in den Konstruktivismus. München, 1992
- WAIDACHER, Friedrich:** Museologie-knapp gefasst. Wien and Köln and Weimar, 2005
- WELLER, Katrin:** Ontologien. In **Kuhlen, Rainer/Semar, Wolfgang/Strauch, Dietmar (Hrsg.):** Grundlagen der praktischen Information und Dokumentation: Handbuch zur Einführung in die Informationswissenschaft und -praxis. 6. Auflage. Berlin and Bosten, 2013, S.207–218
- WERSIG, Gernot:** Information-Komunikation-Dokumentation. Ein Beitrag zur Orientierung der Informations- und Dokumentationswissenschaft. Band 5, Pullach bei München, 1974
- WERSIG, Gernot:** Die Komplexität der Informationsgesellschaft. Konstanz, 1996
- WHITE, Ryen W/ROTH, Resa A:** Exploratory search: beyond the query-response paradigm (Synthesis lectures on information concepts, retrieval & services). Morgan and Claypool Publishers 3 2009
- WIESENMÜLLER, Heidrun/PFEFFER, Magnus:** Resource Discovery Systeme. In **Straub, Martina/Sühl-Strohmenger, Wilfried (Hrsg.):** Handbuch Informationskompetenz. 2. Auflage. Berlin, 2016, S.105–
- WITTEN, Ian H./BAINBRIDGE, David/NICHOLS, David M.:** How to Build a Digital Library, Second Edition. San Francisco, 2010
- WUCHTERL, Kurt:** Lehrbuch der Philosophie. Bern and Stuttgart, 1998
- YU, Liyang:** Introduction to the semantic web and semantic web services. New York, 2007

Abbildungsverzeichnis

1	Modell des Retrieval-Prozess, Quelle: Eigene Darstellung nach LINCKELS, Serge/MEINEL, Christoph: E-librarian service: user-friendly semantic search in digital libraries. Heidelberg, 2011, S.82.	2
2	Signal, Daten, Information und Wissen, Quelle: Eigene Darstellung nach FAVRE-BULLE: Information und Zusammenhang, S.38 und KENDAL / CREEN: An Introduction to Knowledge Engineering, S.7.	21
3	Modell der Informationskette für codierte Signalübertragung, Quelle: FAVRE-BULLE: Information und Zusammenhang, S.28.	23
4	Semiotisches Dreieck, https://de.wikipedia.org/wiki/Semiotisches_Dreieck , 17.12.2016	32
5	Ausdrucksstärke von Methoden der Wissensrepräsentation und Größe der Wissensdomäne, Quelle: Eigene Darstellung nach STOCK: Wissensrepräsentation: Information auswerten und bereitstellen, S.42.	37
6	Taxonomie der IR-Modelle, Quelle: Eigene Darstellung nach BAEZA-YATES et al.: Modern information retrieval. The concept and technology behind search, 2. Auflage, 2011, S.60.	39
7	Darstellung logische Operatoren für das Boolesche Modell, http://archiv.jura.uni-saarland.de/dissertationen , 12.04.2017.	40
8	Beispiel für das Vektormodel, Quelle: Eigene Darstellung	42
9	Darstellung Information Lookup vs. Exploratory Search, Quelle: WHITE, Ryen W/ROTH, Resa A:Exploratory search:beyond the query-response paradigm, S.14.	47
10	Screenshot einer Term Suggestion bei Google, Quelle: Screenshot, https://www.google.at	48
11	Screenshot einer Recommender Empfehlung bei Amazon, Quelle: Screenshot, https://www.amazon.de	49
12	Ausschnitt aus dem SKOS-Objektthesaurus des Kriminalmuseums, Quelle: Screenshot erstellt mit http://labs.sparna.fr/skos-play , gams.uni-graz.at/archive/objects/o:oth/datastreams/ONTOLOGY/content	50
13	Architektur einer semantischen Suchmaschine, Quelle: DENGL, Andreas: Semantische Technologien: Grundlagen - Konzepte - Anwendungen, Heidelberg 2012, S.243.	54
14	Zeitplan und Projektablauf, Quelle: Eigene Darstellung	62
15	Ausschnitt des Trello-Boards zur kollaborative Arbeit und Organisation einzelner, Work-Packes zugeordneten, Tasks, Quelle: Eigene Darstellung	62
16	Digitale Objekt einer Karteikarte, bestehend aus Datenströmen, ihrer Repräsentation und Methoden, Quelle: Eigene Darstellung	70

17	Beispiel für RDF, Quelle: Eigene Darstellung	74
18	Semantic Web Stack, https://en.wikipedia.org/wiki/Semantic_Web_Stack , 06.07.2016	76
19	RDF-Graph einer Karteikarte nach dem Ingest in die GAMS, Quelle: Eigene Darstellung	77
20	Berufsgruppen und Altersverteilung, Screenshot des Resource Discovery, Quel- le: Eigene Darstellung	81
21	Liste der Delikte, Screenshot des Resource Discovery, Quelle: Eigene Darstellung	82
22	Gerichtsorte im eingebetteten Dariah Geo-Browser, Screenshot des Resource Discovery, Quelle: Eigene Darstellung	83
23	Zeitstrahl der Datumsangaben der Ausstellung, Screenshot des Resource Dis- covery, Quelle: Eigene Darstellung	83