

Stefan Waidele
Ensisheimer Straße 2
79395 Neuenburg am Rhein
Stefan@Waidele.info

AKAD University
Immatrikulationsnummer: 102 81 71

Modul WIN03 — Innovative Themen der Wirtschaftsinformatik
Assignment

WEB 3.0

DATEN SIND DAS ÖL DES 21. JAHRHUNDERTS

Betreuer: Prof. J. Anton Illig

Entwurf vom 8. November 2014



AKAD University

Inhaltsverzeichnis

Abbildungsverzeichnis	ii
Abkürzungsverzeichnis	iii
1 Einleitung	1
1.1 Begründung der Problemstellung	1
1.2 Ziele dieser Arbeit	1
1.3 Abgrenzung	2
2 Grundlagen	3
2.1 Betrachtung von Erdöl	3
2.2 Definition: Daten, Information, Wissen	3
2.3 Definition: Web 1.0, Web 2.0	4
2.4 Web 3.0 = Web 2.0 + Semantik = Semantisches Web	5
2.5 Die Bedeutung der Daten	6
3 Hauptteil	8
3.1 Auswirkungen des Semantischen Webs auf die Gesellschaft	8
3.2 Auswirkungen des Semantischen Webs auf die Wirtschaft	10
3.3 Vergleich der Auswirkungen mit denen des Öls	10
4 Fazit & Ausblick	11
4.1 Fazit	11
4.2 Ausblick	11
Literatur- und Quellenverzeichnis	iv

Abbildungsverzeichnis

1	Der Stack des semantischen Web	7
2	Internetnutzung in Deutschland 2014	8

Abkürzungsverzeichnis

HTML	Hypertext Markup Language
IoT	Internet of Things
OLAP	Online Analytical Processing
WWW	World Wide Web

1 Einleitung

1.1 Begründung der Problemstellung

Technologien rund um das Schlagwort *Big Data* sind laut Gartner die großen Triebfedern in der Informationstechnologie¹. Durch Entwicklungen wie *Internet of Things* und *Ubiquitous Computing* werden die in naher Zukunft die generierte Datenmenge als auch die Anzahl der verarbeitenden Instanzen in den nächsten Jahren wohl deutlich zunehmen.

Die steigende Menge der Daten macht eine systematische Aufbereitung der anfallenden Daten hin zu repräsentiertem Wissen möglich und notwendig. Das semantische Web, oder auch *Web 3.0*, verspricht Struktur in die Datenmenge zu bringen. Die Erwartungen, aber auch die Befürchtungen gegenüber den entsprechenden technischen Fortschritten sind immens.

1.2 Ziele dieser Arbeit

Ziel dieser Arbeit ist es, die momentanen Entwicklungen der Datenbeständen hin zum semantischen Web und deren Auswirkungen auf Wirtschaft und Gesellschaft mit den durch die Industrielle Nutzung des Öls im 20. Jahrhundert zu vergleichen.

Hierzu werden zunächst im Kapitel 2 *Grundlagen* die für diese Arbeit relevanten Begriffe und Konzepte definiert, bevor im Kapitel 3 *Hauptteil* die Auswirkungen der Technologien auf Wirtschaft und Gesellschaft beschrieben werden.

To do: An die tatsächliche Arbeit anpassen

¹vgl. [Betsy Burton et al., 2014]

1.3 Abgrenzung

Der Augenmerk dieser Arbeit liegt auf den behandelten Konzepten und Technologien der Informationstechnologie. Die Entsprechungen in der Ölindustrie werden nicht in der gleichen Tiefe erörtert und belegt werden, sondern beschränken sich auf generelle Aussagen und Ansichten.

To do: An die tatsächliche Arbeit anpassen

2 Grundlagen

2.1 Betrachtung von Erdöl

Bei Erdöl handelt es sich um den wichtigsten Energieträger des 20. Jahrhunderts bis hin zur Gegenwart. Durch Erdöl wurden viele technologische Entwicklungen begünstigt oder gar erst möglich gemacht. Durch systematische Aufbereitung in Refinerien sind neben den offensichtlichen Anwendungen wie der Nutzung als Schmier- und Kraftstoff für Mobilität und Maschinen oder zur Wärmegewinnung sind viele weitere Anwendungen möglich, ohne die die moderne Gesellschaft nur schwer vorstellbar wäre².

Die Wertschöpfungskette rund um Erdöl gliedert sich in die Phasen des Finden bzw. Förderns, des Sammelns bzw. Aufbereitens und in die anschließende Nutzung in diversen Endprodukten, vom Treibstoff bis hin zu diversen Kunststoffen. Dabei gilt, dass Erdöl zwar in großen Mengen³, aber nicht unbegrenzt zur Verfügung steht.

Die Auswirkungen des Erdöls auf Technologie, Wirtschaft, Gesellschaft und Politik des 20. Jahrhunderts sind enorm. Erdöl ermöglichte großen Reichtum von Unternehmen und Staaten, aber verursachte auch Krisen und Kriege.

2.2 Definition: Daten, Information, Wissen

In dieser Arbeit sollen die folgenden Definitionen gelten: Ein *Datum* ist eine formalisierte Sachverhaltsaussage, ohne inhärente Bedeutung (z.B. „23°C“). Durch Interpretation im Kontext kann daraus eine *Information* werden (z.B. „Die Außentemperatur beträgt 23°C“)⁴. Durch Vernetzung mehrerer Informationen mit-

²z.B. Kunststoffe und Lacke, welche zum großen Teil auf Erdöl basieren, oder auch die Reichweite, die Fahrzeuge wegen der hohen Energiedichte von Benzin bzw. Diesel erzielen.

³Ob diese Mengen, speziell im Vergleich zum weltweiten Verbrauch, tatsächlich immer noch groß sind, wird z.T. bezweifelt

⁴vgl. [Fuchs-Kittowski, 2000], Seite 40

einander, aber auch durch Erfahrung kann *informatives Wissen* entstehen (z.B. „Das Wetter ist schön“)⁵.

In weiteren Verfeinerungsschritt entsteht dann *handlungsorientiertes Wissen*, (z.B. „Ich benötige beim Nachmittagsspaziergang keinen Pullover“) das dann zu einer konkreten Entscheidung führen kann (z.B. „Ich lasse den Pullover zu Hause.“)⁶.

2.3 Definition: Web 1.0, Web 2.0

Unter *Web 1.0* versteht man das World Wide Web (WWW) wie es ursprünglich entwickelt wurde: Eine Menge von statischen Daten, die miteinander auf willkürliche Weise verknüpft werden konnten. Die Auszeichnungssprache Hypertext Markup Language (HTML) ermöglicht es Autoren, bestimmte Abschnitte zu kennzeichnen. Schon hier gibt es unterschiedliche Informationsgehalte der Auszeichnungen: Während `...` lediglich aussagt, dass der ausgezeichnete Abschnitt in fetter Schriftart angezeigt werden soll, ist eine mit `<h1>...</h1>` ausgezeichnete Überschrift tatsächlich als solche zu erkennen. Auch wenn die dadurch gewonnene Information für ein automatisch erstelltes Inhaltsverzeichnis schon nützlich sein kann, wird hier keine Aussage bzgl. des eigentlichen Inhalts getroffen. Somit sind die Dokumente des Web 1.0 dem Bereich der *Daten* zuzuordnen. Die darin enthaltenen Informationen bzw. das darin enthaltene Wissen ist erst zugänglich, wenn die Daten durch Menschen gelesen und ausgewertet werden⁷.

Durch die Technologien des *Web 2.0* werden die Daten i.d.R. in Datenbanken vorgehalten und die Webseiten erst bei Abruf generiert. Durch die Popularität von Werkzeugen wie Blogs und Wikis sind deutlich mehr Menschen an der Erstellung der Inhalte beteiligt. Weitere Daten werden durch Techniken rund um das *Internet of Things (IoT)* und *Ubiquitous Computing* automatisch erfasst. Diese

⁵vgl. [Nikodemus, 2005], Seite 106

⁶vgl. [Taylor, 1982], Seite 342

⁷vgl. [Alkhatib, 2012], Seite xvi

Daten mit sogenannten *Meta-Daten* angereichert. Hierdurch wird in maschinenlesbarer Form angegeben, welche Informationen die Dokumente enthalten. Neben vom Autor selbst zugeordneten *Tags* und *Kategorien* kommen auch automatisch generierte Meta-Daten hinzu. Beispiele hierfür sind etwa das Veröffentlichungsdatum, Beziehungen zu anderen Dokumenten⁸ oder Geoinformationen (Wo wurde das Dokument erstellt). Inzwischen werden auch die Stimmung des Autors erfragt (z.B. bei Runtastic-Aktivitäten) oder ermittelt (z.B. bei Facebook-Einträgen). Die durch Daten und Meta-Daten erzielte Informationsstufe ist deutlich über der von *Web 1.0*, unterliegt aber deutlichen Schwankungen je nach Dienst bzw. Nutzereingaben.

2.4 Web 3.0 = Web 2.0 + Semantik = Semantisches Web

„*Semantik*, auch *Bedeutungslehre*, nennt man die Theorie oder Wissenschaft von der Bedeutung der Zeichen. Zeichen können in diesem Fall Wörter, Phrasen oder Symbole sein. Die Semantik beschäftigt sich typischerweise mit den Beziehungen zwischen den Zeichen und den Bedeutungen dieser Zeichen.“⁹

Im *Web 3.0* werden die Daten bzw. Informationen des Web 2.0 durch Beifügung von Bedeutung zu Information bzw. informativem Wissen veredelt¹⁰. Hierdurch soll es möglich werden, die schnell steigenden Datenmengen sinnvoll zu nutzen¹¹. Die Bezeichnung *semantisches Web* ermöglicht eine Abgrenzung gegenüber anderen Interpretationen des Buzzwords *Web 3.0*, wie sie z.T. im Marketing¹² oder in der Politikwissenschaft¹³ zu finden sind.

⁸Realisiert durch sog. Backtracks – Wer verlinkt auf dieses Dokument?

⁹citewp:semantik

¹⁰vgl. [Markoff, 2006]

¹¹vgl. [Tolksdorf, 2007]

¹²z.B. „Web 3.0 marketing is the convergence of new technologies and rapidly changing consumer buying trends.“ in [Tasner, 2010], Abschnitt „What is Web 3.0 Marketing?“

¹³z.B. „Is this Web 3.0? Not a tech-upgrade, a smarter algorithm, slicker fibre optic or better Bluetooth beam. Instead, Web 3.0 as in an outcome, the demonstrated consequences of being able to access information?“ in [Pont, 2012], Abschnitt „Web 3.0: Regime Change“

2.5 Die Bedeutung der Daten

Den vorhandenen oder neu gesammelten Daten kann auf verschiedene Arten Bedeutung beigelegt werden. Sogenannte *Meta-Daten* beschreiben Dateien durch Schlagworte oder Zuordnung von Werten zu definierten Schlüsselworten. Die Meta-Daten können in separaten Dateien gespeichert werden¹⁴. Ebenfalls möglich ist die Speicherung direkt in der entsprechenden Datei¹⁵. Diese Daten werden i.d.R. automatisch¹⁶ oder halbautomatisch¹⁷ den Daten zugeordnet. Manche Metadaten können auch bei Abruf direkt ermittelt werden¹⁸.

In textbasierten Dateiformaten wie XML werden hingegen einzelne Elemente durch Markup-Tags maschinenlesbar mit Bedeutung versehen werden. So werden Zeichenketten als Namen, Adressen, Telefonnummer etc. identifizier- und auswertbar. Die Meta-Daten sind hier Teil des Datenstroms.

Durch Techniken wie Gesichtserkennung oder Textanalyse können Meta-Daten auch automatisch aus den Daten generiert werden. Hierbei bestimmt die Menge der auszuwertenden Daten und die Qualität der bereits vorhandenen Metadaten die Qualität der Ergebnisse. Die Ergebnisse werden attraktiver, je mehr die Entwicklung von *Big Data* zu *Big Information* und gar *Big Knowledge*¹⁹ geht²⁰.

Durch Regelwerke, hinterlegte Logik, Verknüpfungen und standardisiertem Datenaustausch können aus den Informationen auch Wissen erzeugt werden: Person A ist auf einem Bild zusammen mit Person B zu sehen (Gesichtserkennung). Die Geoinformation und Uhrzeit (EXIF-Tags) zeigen, dass das Foto auf einer Veranstaltung aufgenommen wurde, das zu diesem Zeitpunkt an diesem Ort stattfand (Semantisches Markup der Veranstaltung). Daraus kann auf politische Ge-

¹⁴z.B. in den .INFO-Dateien in der Benutzeroberfläche des Comodore Amiga „Workbench“, oder in der Bilderdatenbank der Applikation „Adobe Lightroom“

¹⁵z.B. EXIF-Daten in Bilddateien oder Informationen zum Werk in MP3-Dateien

¹⁶z.B. durch die Kamera beim fotografieren.

¹⁷Einmalig manuell, dann Bereitstellung für andere Nutzer per CDDb oder andere Dienste.

¹⁸z.B. Dateigröße

¹⁹vgl. [Graves, 2014]

²⁰vgl. [Schmidt, 2009]

sinnung, Musikgeschmack und/oder Trinkfreudigkeit von Person A geschlossen werden.

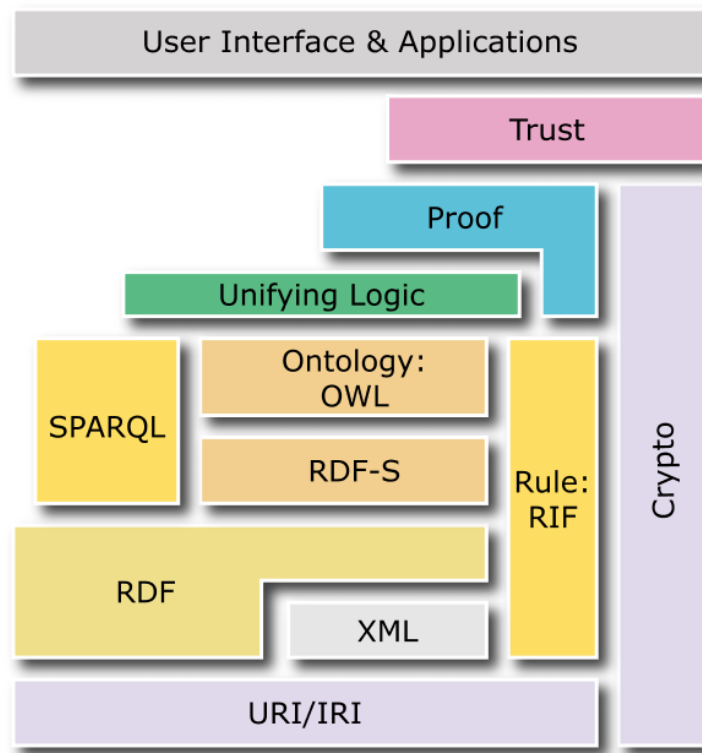


Abbildung 1: Der Stack des semantischen Web²¹

²¹[Bratt, 2007], Seite 24

3 Hauptteil

3.1 Auswirkungen des Semantischen Webs auf die Gesellschaft

Schon heute drehen sich die meisten der regelmäßig ausgeführten Aktivitäten im Internet um *Informationen* oder gar *Wissen*, nicht um *Daten*. So sind 56% der im Digitalindex 2014 befragten Deutschen überzeugt, im Internet die aktuellsten Informationen zu finden, 60% sucht benötigte Informationen zuerst im Netz²². Die Erwartungen bzgl. Aktualität und Informationsgehalt an das WWW sind also sehr hoch.

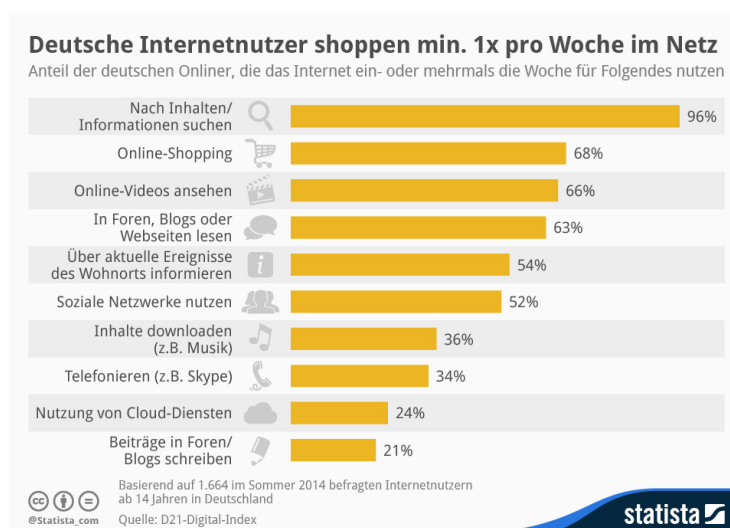


Abbildung 2: Internetnutzung in Deutschland 2014²³

Unter den Top 10 der regelmäßig durchgeführten Tätigkeiten der Befragten im Web finden sich die informationsorientierten Tätigkeiten „nach Inhalten/Informationen suchen“ auf Platz eins, „über aktuelle Ereignisse des Wohnorts informieren“ auf Platz 6 und die Erfassung eigener Daten auf Platz 10²⁴. Weitere Tätigkeiten wie „Soziale Netzwerke nutzen“ (Platz 7) bzw. „Online-Videos anse-

²²vgl. [Initiative D21 et al. (Hrgb.), 2014a], Seite 6

²³[Initiative D21 et al. (Hrgb.), 2014a], Seite 37

²⁴vgl. [Initiative D21 et al. (Hrgb.), 2014a], Seite 37

hen“ (Platz 3) sowie „Online-Shopping“ (Platz 2) nutzen Internetangebote, die per Design sehr gut mit Taxonomien, Meta-Daten und Verknüpfungen ausgestattet sind.

In den TV-Werbespots von Google und Apple werden Verbraucher gezielt auf entsprechend Formulierte Anfragemöglichkeiten hingewiesen²⁵. Somit werden die Möglichkeiten wie Routenplanung, Musiktiterkennung, Wettervorhersage und andere, die schon heute durch einzelne Anwendungen zur Verfügung gestellt werden in einer Oberfläche Gebündelt und in das Bewusstsein der Verbraucher gebracht.

Neben dem offensichtlichen Nutzen des WWW bringt die Entwicklung des Webs auch negative Auswirkungen auf die Gesellschaft. Dies äußert sich mit der Besorgnis von 60% der Nutzer über die im Internet möglicherweise verfügbaren persönlichen Daten²⁶. Bei der Nutzung von Webdiensten der öffentlichen Verwaltung haben 65% der Befragten Angst vor Datendiebstahl, und 73% der Deutschen haben ein starkes Interesse daran, wie Behörden mit den Daten der Bürger umgeht²⁷. Sowohl fiktive²⁸ also auch reale²⁹ Unrechtsstaaten basieren auf der intensiven Erfassung und Auswertung von personenbezogener Daten der Bürger, was entsprechende Befürchtungen nährt und erklärt.

Man kann davon ausgehen, dass mit weiter wachsenden Datenmengen, aber auch durch entsprechenden Wachstum an generierten und erfassten Informationen und Wissen sowohl die positiven Erwartungen als auch die Befürchtungen und Ängste in der Bevölkerung zunehmen werden.

²⁵z.B. „Ok Google, wie lange brauche ich nach Spandau?“

²⁶vgl. [Initiative D21 et al. (Hrgb.), 2014a], Seite 6

²⁷vgl. [Initiative D21 et al. (Hrgb.), 2014b], Seite 9 bzw. Seite 34

²⁸vgl. [Orwell, 1949]

²⁹vgl. [Wikipedia, 2014]

3.2 Auswirkungen des Semantischen Webs auf die Wirtschaft

Schon heute haben Unternehmen erkannt, dass es nicht ausreicht, immer mehr Daten anzuhäufen. Der Schritt von *Big Data* zu *Big Information* bringt Benutzerfreundlichkeit und echten Mehrwert für Unternehmen³⁰. In großen Unternehmen wird unter den Schlagwort *Business Intelligence* bzw. *Online Analytical Processing (OLAP)* mit unterschiedlichen Technologien aus den im *Data Warehouse* gespeicherten Daten Informationen und informatives Wissen zu extrahieren, auf dem dann Entscheidungen und Handlungen basieren können.

Schon in der Vergangenheit haben Fortschritte in der Daten- und Informationsbearbeitung erhebliche Auswirkungen auf die Wirtschaft gehabt. So war die Entwicklung der Telegraphie ein wichtiger Schritt für die Meteorologie, da hiermit zeitnahe Datensammlung und Auswertung möglich wurde. Als Resultat waren Wettervorhersagen für die Schifffahrt möglich. Mit steigender Datenmenge und immer noch wachsender Verarbeitungsgeschwindigkeit werden im Zusammenspiel mit dem Fortschritten des semantischen Webs auch in anderen Bereichen immer treffendere Analysen und Vorhersagen möglich werden. Hier sind also auch erhebliche Auswirkungen auf die Wirtschaft zu erwarten.

Die wie auf Seite 9 beschriebenen, weit verbreiteten und zum Teil großen Befürchtungen stellen Herausforderungen an die Unternehmen dar. Neue Technologien und Anwendungen ebendieser sollten so entworfen und kommuniziert werden, dass Verbraucher ihnen vertrauen können.

3.3 Vergleich der Auswirkungen mit denen des Öls

³⁰vgl. [Hubbart, 2012]

4 Fazit & Ausblick

4.1 Fazit

In der Analogie zu Erdöl lassen sich die *Daten* wie in der These beschrieben mit Rohöl vergleichen. Die Techniken rund um *Big Data* gemeinsam mit den datenbankbezogenen Schichten des *semantischen Webs* entsprechen somit den Ölspeichern, welche den Rohstoff bereit halten. Die Weiterverarbeitung erfolgt dann in den logischen Schichten des *Web 3.0*, welche somit am ehesten mit den Raffinerien vergleichbar sind, die aus den klebrigen schwarzen Rohöl der Daten die verschiedenen Informations- und Wissensprodukte erzeugt. Ob diese dann mit Schnierstoffen, schwerem Schiffsdiesel, hocheffizientem Kerosin oder Kunststoffen vergleichbar sind hängt sowohl von den Ausgangsdaten, aber auch wie beim Öl von der Zielsetzung und Anforderungen ab.

Datenschutz ist in diesem Bild vergleichbar mit Umweltschutz, der auf allen Ebenen dafür kämpft, dass die negativen Auswirkungen der neuen Technologie gemindert bzw. eliminiert werden. Wo beim Öl Strände, Seevögel, die Atmosphäre und das Klima geschützt werden, sind es im Umfeld der Informationstechnologie Bürger-, Grund- und Persönlichkeitsrechte, die vor übermäßigem und falschem Einsatz der Technologie geschützt werden sollen.

4.2 Ausblick

Nach dieser Einordnung des *Web 3.0* bietet sich für weitere Arbeiten die genauere Untersuchung der einzelnen Aspekte an, die hier zum großen Teil nur genannt werden konnten. Speziell die automatische Informationserkennung durch Mustererkennung in Texten, Grafiken und Videos und die daraus entstehenden Möglichkeiten sind eine nähere Betrachtung wert. Ebenso die höheren Ebenen des *Web 3.0 Stacks*, in denen es ebenfalls darum geht, Zusammenhänge rechnergestützt zu erkennen. Durch die Automatisierung der Informationsgewinnung

wird diese skalierbar, so wie es die Datenspeicherung heute schon ist. Dies wird die in dieser Arbeit beschriebenen Entwicklungen nochmals deutlich beschleunigen und in ihrer Tragweite vergrößern. Daher empfiehlt sich auch eine gründliche Untersuchung der gesellschaftlichen und politischen Aspekte des Semantischen Webs.

Literatur– und Quellenverzeichnis

- [Alkhatib, 2012] Alkhatib, G. (2012). *Models for Capitalizing on Web Engineering Advancements*. IGI Global.
- [Betsy Burton et al., 2014] Betsy Burton et al. (2014). *Gartner’s Hype Cycle Special Report for 2014*. Gartner inc., Stamford, CT, USA.
- [Bratt, 2007] Bratt, S. (2007). Semantic web: Linked data on the web. [http://www.w3.org/2007/Talks/0130-sb-W3CTechSemWeb/#\(24\)](http://www.w3.org/2007/Talks/0130-sb-W3CTechSemWeb/#(24)), abgerufen am 07.11.2014.
- [Fuchs-Kittowski, 2000] Fuchs-Kittowski, K. (2000). *Wissens–Ko–Produktion — Organisationsinformatik*. Gesellschaft für Wissenschaftsforschung, 2. auflage edition.
- [Graves, 2014] Graves, T. (2014). Big-data, big-information, big-knowledge. <http://weblog.tetradian.com/2014/03/04/big-data-big-information-big-knowledge/>, abgerufen am 08.11.2014.
- [Hubbart, 2012] Hubbart, D. (2012). Enough about big data. let’s talk big information. <http://blog.opendns.com/2012/05/15/enough-about-big-data-lets-talk-big-information/>, abgerufen am 08.11.2014.
- [Initiative D21 et al. (Hrgb.), 2014a] Initiative D21 et al. (Hrgb.) (2014a). *Die Entwicklung der digitalen Gesellschaft in Deutschland*. Initiative D21 e.V., Berlin.
- [Initiative D21 et al. (Hrgb.), 2014b] Initiative D21 et al. (Hrgb.) (2014b). *eGovernment Monitor 2014*. Initiative D21 e.V.
- [Markoff, 2006] Markoff, J. (2006). Entrepreneurs see a web guided by common sense. <http://www.nytimes.com/2006/11/12/business/12web.html?pagewanted=all>, abgerufen am 05.11.2014.
- [Nikodemus, 2005] Nikodemus, P. (2005). *Wissensmanagement und Innovation*. Cuvillier Verlag.
- [Orwell, 1949] Orwell, G. (1949). *Nineteen Eighty-Four*. Secker & Warburg.
- [Pont, 2012] Pont, S. (2012). *The Better Mousetrap*. Kogan Page, web–edition edition.

- [Schmidt, 2009] Schmidt, H. (2009). Das internet der nächsten generation. <http://www.faz.net/aktuell/wirtschaft/netzwirtschaft/web-3-0-das-internet-der-naechsten-generation-1796817.html>, abgerufen am 08.11.2014.
- [Tasner, 2010] Tasner, M. (2010). *The Limiting Factors of Web 2.0 and How Web 3.0 Is Different*. FT Press, web edition.
- [Taylor, 1982] Taylor, R. S. (1982). Value-added processes in the information life cycle. *Journal of the American Society for Information Science*, pages 341–346.
- [Tolksdorf, 2007] Tolksdorf, R. (2007). Web 3.0 – die dimension der zukunft. <http://www.tagesspiegel.de/zeitung/web-3-0-die-dimension-der-zukunft/1028324.html>, abgerufen am 05.11.2014.
- [Wikipedia, 2014] Wikipedia (2014). Ministerium für staatssicherheit. http://de.wikipedia.org/wiki/Ministerium_f%C3%BCr_Staatssicherheit.

Eidesstattliche Erklärung

Ich versichere, dass ich das beiliegende Assignment selbstständig verfasst, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie alle wörtlich oder sinngemäß übernommenen Stellen in der Arbeit gekennzeichnet habe.

(Datum, Ort)

(Unterschrift)